

**University of São Paulo
“Luiz de Queiroz” College of Agriculture**

**Carbon geomicrobiology, saturation deficit and sequestration potential of
Brazilian agricultural soils**

Heidy Soledad Rodríguez Albarracín

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

**Piracicaba
2024**

Heidy Soledad Rodríguez Albarracín
Forestry Engineer

**Carbon geomicrobiology, saturation deficit and sequestration potential of Brazilian
agricultural soils**

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
2024

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

Rodríguez Albarracín, Heidy Soledad

Carbon geomicrobiology, saturation deficit and sequestration potential of Brazilian agricultural soils / Heidy Soledad Rodríguez Albarracín. -- versão revisada de acordo com a Resolução CoPGr 6018 de 2011. - - Piracicaba, 2024.

192 p.

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

1. Sequestro de carbono 2. Mineralogia 3. Espectroscopia do solo 4. Atividade enzimática 5. Mapeamento digital do solo I. Título

CONTENTS

RESUMO	6
ABSTRACT	7
1. GENERAL INTRODUCTION	9
References	18
2. POTENTIAL OF SOIL MINERALS TO SEQUESTER SOIL ORGANIC CARBON ...	27
Abstract.....	27
2.1. Introduction	27
2.2. Methodology.....	31
2.2.1. Study area	31
2.2.2. Point modeling of soil carbon sequestration potential	32
2.2.3. Spatialization of carbon sequestration potential.....	36
2.3. Results	38
2.3.1. Point modeling of soil carbon sequestration potential	38
2.3.2. Carbon sequestration potential mapping	44
2.4. Discussion.....	50
2.5. Conclusions	56
References.....	58
3. SOIL ORGANIC CARBON SEQUESTRATION POTENTIAL EXPLAINED BY MINERALOGY AND MICROBIOLOGICAL ACTIVITY BY SPECTRAL TRANSFER FUNCTIONS.....	67
Abstract.....	67
3.1. Introduction	67
3.2. Material and Methods.....	70
3.2.1. Study area	70
3.2.2. Microbiological activity estimated by spectrotransfer functions	71
3.2.3. Predictive models for C sequestration potential	78
3.3. Results	79
3.3.1. Method 1.....	79
3.3.2. Method 2.....	84
3.3.3. Predictive models for C sequestration potential.....	87
3.4. Discussion.....	90
3.4.2. Methodology 2.....	91

3.5. Conclusions.....	99
References	100
Appendix	109
4. A PROGRESSIVE KNOWLEDGE STRATEGY TO DETECT MICROBIOLOGICAL ACTIVITY IN SOIL THROUGH VIS NIR-SWIR AND MID IR SPECTRAL CHARACTERIZATION AND MACHINE LEARNING TECHNIQUES	113
Abstract	113
4.1. Introduction.....	114
4.2. Methodology.....	118
4.2.1. Study area.....	118
4.2.2. Selection of sampling points	118
4.2.3. Laboratory analysis	119
4.2.4. Spectral data acquisition	119
4.2.5. Statistical analysis and spectral behavior.....	120
4.2.6. Mid-IR and XRF Spectroscopy.....	120
4.3. Results.....	121
4.3.1. Mid IR spectral interpretation	121
4.3.2. Spectral interpretation Vis NIR SWIR.....	130
4.3.3. Mid-IR and XRF spectromicroscopy	131
4.4. Discussion	139
4.4.1. Mid IR spectral interpretation	139
4.4.2. Spectral interpretation Vis NIR SWIR.....	143
4.4.3. Mid-IR and XRF spectromicroscopy	144
4.5. Conclusions.....	146
References	148
5. SPATIALIZATION OF SOIL BETA-GLUCOSIDASE, PHOSPHATASE AND UREASE ENZYME ACTIVITY INDICES FOR BRAZILIAN AGRICULTURAL AREAS: A PARADIGM SHIFT.....	157
Abstract	157
5.1. Introduction.....	157
5.2. Methodology.....	161
5.2.1. Study area.....	161
5.2.2. Soil observations	161
5.2.3. Enzyme activity indexes	162

5.2.4. Environmental covariates	164
5.2.5. Spatial prediction.....	166
5.2.6. Data interpretation	167
5.2.7. Spatial validation	167
5.3. Results and discussion.....	168
5.3.1. Index of enzyme activity	168
5.3.2. Environmental covariates	171
5.3.3. Spatial prediction.....	174
5.3.4. Spatial validation	179
5.4. Conclusions	180
References.....	181

RESUMO

Geomicrobiologia, déficit de saturação e potencial de sequestro de carbono dos solos agrícolas brasileiros

O serviço ecossistêmico de regulação climática provido pelo solo se dá por sua capacidade de sequestrar C, sendo que o carbono orgânico do solo (COS) é essencial para sua saúde. A capacidade do solo de reter COS depende dos minerais e de sua interação com a microbiota. O presente trabalho aborda no *capítulo 1* a análise do potencial de sequestro de COS da fração argila para solos da região de Piracicaba, estado de São Paulo, com base em uma equação do déficit potencial de saturação de C de partículas finas do solo, ajustada para solos agrícolas tropicais. Esse potencial foi ajustado por um modelo de regressão espacial. Na camada superficial, o potencial de sequestro é explicado principalmente pela abundância relativa de caulinita, hematita, goethita e gibbsita determinados via espectroscopia Vis-NIR-SWIR. Foi observada uma relação direta com a goethita e a gibbsita. Em uma profundidade de 80 a 100 cm, a caulinita e a hematita foram responsáveis pela maior variação no potencial de sequestro. A contribuição de cada mineral para o potencial de sequestro de COS também foi mapeada, verificando-se altas contribuições de goethita e gibbsita para as camadas profundas do solo. O *Capítulo 2* foi baseado no ajuste do modelo de potencial de sequestro de C com variáveis microbiológicas e mineralógicas. A modelagem e o mapeamento de diversas propriedades microbiológicas foram realizados por meio de funções de transferência espectral e mapeamento digital do solo (MDS), atingindo R^2 de 0,77 a 0,85. Todas essas propriedades foram detectadas usando bandas específicas, que alcançaram correlações de 0,64 a 0,98 com as análises laboratoriais. Os modelos autorregressivos obtiveram r de 0,61 a 0,7. As variáveis explicativas foram associadas à caulinita, hematita, goethita, gibbsita e abundância de fungos, actinomicetos, fungos micorrízicos vesiculares-arbusculares, atividade enzimática da beta-glucosidase, urease e fosfatase e matéria orgânica particulada (POM), sendo a abundância geral de fungos a variável microbiológica mais importante. O *Capítulo 3* baseou-se no desenvolvimento de uma estratégia para analisar a atividade microbiológica em microescala por meio da detecção espectroscópica de 35 amostras com análise do carbono da biomassa microbiana (MBC) e atividade enzimática de betaglucosidase, urease e fosfatase, e fracionamento da matéria orgânica do solo (MOS) em POM e MOS associada à fração mineral (MAOM). A fim de caracterizar os espectros Vis-NIR-SWIR e Mid-IR das diferentes frações em função das variáveis microbiológicas, foram selecionadas bandas específicas para cada variável. Por último, no capítulo 4, foi desenvolvida uma técnica para calcular e espacializar os índices de atividade das enzimas beta-glucosidase, fosfatase e urease para as áreas agrícolas do Brasil utilizando MDS e tendo como covariáveis uma Imagem Sintética do Solo (SYSI), variáveis associadas ao relevo, clima, biomas e mapas mineralógicos. Os mapas de atividade enzimática foram obtidos para a área agrícola do Brasil (3481362,60 km²), com um R^2 de validação variando de 0,68 a 0,35. Esses índices de atividade enzimática em escala de 30 m podem ser considerados uma contribuição importante para o monitoramento e o mapeamento da qualidade e da saúde dos solos brasileiros, pois são sensíveis ao uso e ao manejo da terra.

Palavras-chave: Sequestro de carbono, Mineralogia, Espectroscopia do solo, Atividade enzimática, Mapeamento digital do solo

ABSTRACT

Carbon geomicrobiology, saturation deficit and sequestration potential of Brazilian agricultural soils

The ecosystem service of climate regulation provided by soil is due to its capacity to sequester C, and soil organic carbon (SOC) is essential for its health. The capacity of the soil to retain OC depends on the minerals and their interaction with the microbiota. *Chapter 1* of this work analyzes the potential for COS sequestration in the clay fraction for soils in the Piracicaba region, in the state of São Paulo, based on an equation for the potential C saturation deficit of fine soil particles, adjusted for tropical agricultural soils. This potential was adjusted using a spatial regression model. In the surface layer, the sequestration potential is mainly explained by the relative abundance of kaolinite, hematite, goethite and gibbsite determined by Vis-NIR-SWIR spectroscopy. A direct relationship was observed with goethite and gibbsite. At a depth of 80 to 100 cm, kaolinite and hematite were responsible for the greatest variation in sequestration potential. The contribution of each mineral to COS sequestration potential was also mapped, with high contributions from goethite and gibbsite in the deep soil layers. *Chapter 2* was based on the adjustment of the C sequestration potential model with microbiological and mineralogical variables. The modeling and mapping of different microbiological properties was carried out using spectral transfer functions and digital soil mapping (DSM), achieving R^2 of 0.77 to 0.85. All these properties were detected using specific bands, which achieved correlations of 0.64 to 0.98 with the laboratory analyses. The autoregressive models obtained r from 0.61 to 0.7. The explanatory variables were associated with kaolinite, hematite, goethite, gibbsite and the abundance of fungi, actinomycetes, vesicular-arbuscular mycorrhizal fungi, enzymatic activity of beta-glucosidase, urease and phosphatase and particulate organic matter (POM), with the overall abundance of fungi being the most important microbiological variable. *Chapter 3* was based on the development of a strategy to analyze microbiological activity at the microscale through the spectroscopic detection of 35 samples with analysis of microbial biomass carbon (MBC) and enzymatic activity of beta-glucosidase, urease and phosphatase, and fractionation of soil organic matter (SOM) into POM and SOM associated with the mineral fraction (MAOM). In order to characterize the Vis-NIR-SWIR and Mid-IR spectra of the different fractions according to the microbiological variables, specific bands were selected for each variable. Finally, in *chapter 4*, a technique was developed to calculate and spatialize the activity indices of the enzymes betaglycosidase, phosphatase and urease for agricultural areas in Brazil using DSM and having as covariates a Synthetic Soil Image (SYSI), variables associated with relief, climate, biomes and mineralogical maps. The enzyme activity maps were obtained for the agricultural area of Brazil (3481362.60 km²), with a validation R^2 ranging from 0.68 to 0.35. These enzyme activity indices on a 30 m scale can be considered an important contribution to monitoring and mapping the quality and health of Brazilian soils, as they are sensitive to land use and management.

Keywords: Carbon sequestration, Mineralogy, Soil spectroscopy, Enzyme activity, Digital soil mapping

1. GENERAL INTRODUCTION

Soil is a fundamental natural resource that supports 95% of human food production (Sarkheil et al., 2020), and represents the difference between the survival and extinction of most terrestrial life (Doran and Zeiss, 2000). Therefore, maintaining or improving the health of our soils is important for achieving the Sustainable Development Goals (SDGs), ensuring the provision of ecosystem services, food security and sustaining life on Earth (Sarkheil et al., 2023).

Soil dynamics responds to the direct interaction between microorganisms, biomineralization and synergistic coevolution with plants (Gouda et al., 2018), which gives it a number of functions corresponding to (i) food and biomass production, (ii) storage, filtration and transformation of compounds, (iii) habitats for living beings and gene pools, (iv) physical and cultural environment, (v) source of raw materials, (vi) carbon reservoir and (vii) archive of geological and archaeological heritage (European Commission (COM 2006.231). These functions are related to the ecosystem services provided by soil associated with provisioning (food, fiber and timber production), regulation (climate, flood and water regulation), and cultural and supporting services (nutrient cycling, soil formation) (Silvero et al., 2023).

Within the climate regulation service, soil has the capacity to sequester carbon (C), which favors the reduction of carbon dioxide (CO₂) concentration in the atmosphere (Houghton 2003, Kimble et al. 2003). This is because global soils have the potential to absorb about 20% of anthropogenic C emissions (Yang et al., 2021). Therefore, C sequestration is a phenomenon that can help to partially mitigate climate change (Padarian et al., 2022), as well as greenhouse gas emissions (Minasny et al., 2017).

It is important to mention that the largest stock of organic carbon (OC) associated with terrestrial ecosystems is found in soil (Eswaran et al. 1993, Hounkpatin et al. 2018), comprising approximately two-thirds of global C (Scharlemann et al., 2014), estimated at 2500 Gt, which includes 1550 Gt of soil organic carbon (SOC) and 950 Gt of inorganic soil C (Hounkpatin et al., 2018). Changes in these soil C stocks constitute a change in atmospheric CO₂ concentration that may affect global climate change, as soil stocks contain 3.3 times more than atmospheric stocks (760 Gt) and 4.5 times more than biota stocks (560 Gt) (Hounkpatin et al., 2018). Analysis of the vertical and horizontal distribution of soil C interferes with the abundance of C-related greenhouse gases (Hobley et al. 2015, Xu et al. 2011). The importance of SOC not only influences climate change mitigation but is also critical for maintaining soil productivity and health in agricultural systems (Li et al., 2019).

SOC dynamics is closely related to the soil development process and thus to its formative factors (Hobley et al., 2015). There are active factors, such as climate and organisms, and passive factors, such as parent materials, relief, and time. To make predictions of the C cycle, it is necessary to understand the complete processes related to C sequestration and release (Marschner et al., 2008).

In relation to the climate-forming factor, the variables that most affect C storage are related to temperature and precipitation (Shi et al., 2020), which determine the intensity of weathering of parent material, net primary productivity (Lal 2004, Adhikari et al. 2014, Weil & Brady 2016, Lamichhane et al. 2019) and the rate of C decomposition by microorganisms (Lal, 2004). According to Zeraatpishe & Khormali (2012) OC decreases with decreasing precipitation and increasing temperature. In addition, precipitation plays a key role in biomass productivity that determines litter input to the soil (Chaplot et al., 2010), due to its influence on the volume, quality and quantity of mineralization (Zeraatpishe & Khormali, 2012).

In relation to relief, topographic indices, such as elevation, slope and landscape position, influence the action of climatic factors on soils (Weil and Brady, 2016, Lamichhane et al. 2019). Terrain parameters, associated with land surface shape, elevation, slope and curvature, are used as predictive covariates of SOC (Sothe et al. 2022, Guevara et al. 2020, Rentschler et al. 2019, Forkuor et al. 2017, Fissore et al. 2017, Hengl et al. 2017, Nyssen et al. 2008, Oueslati et al. 2013). Because these variables control soil water status, litter mineralization dynamics, erosion and deposition processes (Hengl et al., 2015). In addition, terrain elevation is related to temperature and thus responds to the rate of decomposition (Schindlbacher et al. 2010). The slope of the terrain influences the type and intensity of cultivation (Kobler et al., 2019).

Vegetation also interferes with SOC stocks because it favors the addition of plant biomass and its ease of decomposition (Bui et al., 2009), associated with the type of species (Mueller et al. 2015, Vesterdal et al. 2012). These biomass inputs are relevant in the most superficial horizons but gain importance with increasing depth due to increased subsurface OC stocks (Rasse et al. 2005, Lorenz et al. 2017) and their interaction with the mineralogy of the parent material (Angst et al., 2018), as rock composition influences OC conservation and litter production and thus the interaction of roots with the soil (Hassink 1997).

Soil cover is an indicator of organic activity (Lamichhane et al. 2019) and is therefore one of the most influential predictors of carbon stocks (Rial et al. 2017, Wiesmeier et al. 2011). For Akpa et al. (2016), SOC contents vary with land use change, for example, significant changes occur when moving from a natural or semi-natural cover to an agricultural type, due

to vegetation variation that affects the proportion of SOC contributed by biomass, similarly the amount of biomass also varies according to land use (Sothe et al. 2022). Dos Reis et al (2014), indicates that the type of land use will especially affect the amounts of soil organic matter (SOM) reaching the soil, and the change in use affects SOM associated with minerals (MAOM).

In general, management practices, as well as soil type and climate, intervene in C dynamics through their influence on physical, chemical and biological agents and thus on its availability to microorganisms and its loss through mineralization and erosion (Singh et al., 2018). Erosion and deposition processes, degradation and volatilization related to the activity of organisms, translocation of dissolved and particulate carbon caused by water runoff and infiltration, and stabilization within micro- and macroorganisms also influence SOC dynamics (Curtin et al., 2012).

In general, in relation to the dynamics of C, much progress has been made on the influence of climate and relief factors, and on organisms related to vegetation on the soil. Climatic and topographic variables such as mean annual precipitation and temperature, slope, among others, land use, soil physical characteristics associated with texture, parent material, among others, and microbial biomass have been used as covariates for SOC mapping (Albaladejo et al. 2013, Jobbágy & Jackson 2000, Ladd et al. 2013, Powers et al. 2011, Poeplau et al. 2011, Wiesmeier et al. 2012). The definition of variables that interfere with SOC varies in relation to regional conditions (Li et al., 2019), and these variables are related to climate, parent material and organisms in the area (Minasny et al., 2013), which combinations create unique local conditions (Carvalho et al., 2019). However, it is important to clarify that, on a global and continental scale, SOC is controlled by temperature and precipitation, increasing at higher precipitation and lower temperature (Hengl et al. 2015, Minasny et al. 2013). But one should not ignore the importance of the other factors, such as parent material, which is related to the amount of clay that influences SOC protection due to the interaction of SOC with the reactive surface of clays (Grüneberg et al. 2013, Mayer 1994), thus interfering with carbon stabilization (Wagai et al., 2008) associated with organo-mineral aggregation (von Lützow et al., 2008).

The ability of soil to sequester C depends especially on mineralogy and its interaction with microorganisms. Ingram and Fernandes (2001), Weil and Brady (2016) indicated the importance of mineralogy on SOC storage potential, especially soils in deeper layers (Grey and Bishop 2015, Wiesmeier et al. 2011). In contrast, land use presents a more significant influence in the surface layers (Adhikari et al., 2014, Hobley et al., 2015).

For Ashton et al. (2016), clay content alone is not related to the increase or decrease of C. The concentration is influenced by the mineralogy and geochemistry of the soil, so it depends on the presence of clay minerals and high specific surface area iron oxides and pH. Therefore, the clay fraction is key to C sequestration because it involves different minerals and varying amounts of pedogenic Fe and Al oxides (Yang et al. 2021, Kirsten et al. 2021, Prout et al. 2021). Minerals such as iron and aluminum oxides, specifically goethite and gibbsite, are recognized as critical C sorption surfaces (Kaiser and Zech 2000, Dos Reis et al. 2014).

Soil oxides, oxyhydroxides and hydroxides present electrostatic attractions and ionic bonds between the hydroxyl groups of the oxides and the carboxyl or hydroxyl groups of the SOM, with additional strong relationships between iron, manganese and humic substances (De Mastro et al., 2020). This adsorption of SOM by minerals favors SOC stabilization, reducing microbial mineralization (Kalbitz et al., 2005), due to the specific surface area of these minerals and their surface charge that favors these bonds and stabilizes SOM. Kirsten et al. (2021) indicated that clay minerals and oxyhydroxides (pedogenic metal oxides) are the most reactive and control SOC persistence. However, most studies did not explicitly evaluate the contribution of clay minerals to SOC. For example, Weismeyer et al. (2013) estimated the SOC sequestration potential of silt and clay particles in soils from Germany, without specifying the type of clay mineral. Ashton et al. (2016) analyzed SOC concentrations in different parent materials and clay mineralogies, evaluating total concentrations without determining the specific contribution of each mineral. Yang et al. (2021) evaluated the spatiotemporal dynamics of carbon adsorption and release in aggregates of a transparent smectite clay, also relating enzymatic decomposition, through 4D imaging on a microfluidic chip. It should be noted that this study was not performed on a specific soil. Kirsten et al. (2021) determined the contribution of kaolinite, gibbsite, goethite and hematite to C storage in soils under forests and agricultural lands, evaluating only the variation of C and clay mineral concentrations, but keeping the mineral types invariant.

Mapping the C sequestration potential of the clay minerals that make up the clay fraction is important for understanding their dynamics and soil management (Padarian et al., 2022), remembering that SOC contents and their stable forms vary in relation to the amount and type of mineral (Yang et al. 2021, Kirsten et al. 2021, Prout et al. 2021, Dos Reis et al. 2014).

In addition to mineralogy, microbiological activity is also important in predicting SOC sequestration potential. Microorganisms are responsible for 80 to 95% of C mineralization (Hassink, 1994), by the use of C through respiration (Follett et al., 2001). They contribute with the decomposition of SOM and C stabilization by participating in the formation of

microaggregates with extracellular activity and with their dead remains (Nicolas et al. 2019, De Mastro et al. 2020).

In general, the degradation of SOM, especially its degree of decomposition, is related to the intra- and extracellular enzymatic activity of fungi and bacteria (De Beeck et al., 2021). Enzymes act as regulators in litter decomposition thus influencing labile forms of SOM (Zhang et al 2020). Considering the microbial C-pump concept described by Adamczyk et al. (2019), stable forms of SOM result from microbiological activity. As microorganisms generate biomass through metabolic processing of plant residues, microbial residues are converted into stable forms by interacting with soil minerals.

Aluminum and iron oxides, as well as clay minerals, especially smectites, control C storage and release; however, their relationship with the presence of microbes and extracellular enzymes that degrade SOM is still unclear (Yang et al., 2021). According to Parink et al. (2014), bacterial adhesion to clay minerals and oxides is mediated by interactions with proteins, extracellular enzymes, and hydrogen bonds.

There is a need to improve the understanding of the activity and diversity of exoenzymes and their interactions with minerals in SOC modeling, as microbial and extracellular enzyme activity directly affects the efficacy of SOC mineral protection and promotes its release (Yang et al., 2021). Hart et al. (2020) proposed the use of spectroscopy-focused technologies in association with molecular cues to model the presence of fungi and bacteria in soil samples. Since microbiological activity presents a close relationship with the types and amounts of SOM, that is, it responds to the quantity and quality of OC (Rasche et al., 2013), Vis NIR SWIR and Mid IR reflectance spectroscopy analysis is key for the characterization of this activity, since it facilitates the differentiation of the functional groups of the different organic compounds in the soil (Ojeda et al. 2008, Viscarra Rossel and Hicks, 2015) and allows the interpretation of clay mineralogy and iron oxides that are related to soil microorganisms (Viscarra Rossel et al. 2022, Yang et al. 2021).

However, there are no specific vis-NIR-SWIR and Mid IR absorbances assigned to microbial communities, but, since the soil physicochemical environment conditions the abundance and function of microorganisms (Rasche et al., 2011), and the spectral response of the soil favors the prediction of fundamental components such as minerals, SOM and water content (Yang et al. 2021, Viscarra Rossel et al. 2022) that are required by fungal and bacterial communities for their growth and obtaining energy (Müller, 2015), it is possible to relate wavelengths to microbiological characteristics.

The major difficulty in accessing soil microbiological properties directly through spectral frequencies is due to the low content of microbial biomass and enzyme activity which limits the induction of changes in spectra (Rinnan and Rinnan, 2007). However, microbiological properties can be predicted due to the strong relationship with the compounds comprising the SOC which influence the spectral response of the soil (Cohen et al. 2005, Chodak 2011). As an example, extracellular enzymes produced by fungi for lignin and lignocellulose degradation release C into the soil solution (Nicolas et al., 2019), which allows interpreting spectral information as an indicator of soil microbiology (Rasche et al. 2013, Viscarra Rossel et al., 2022). Additionally, interaction with soil mineralogy can also be interpreted spectrally, Fe, for example, is related to microbial energy generation for Fe-reducing microorganisms (Weber et al., 2006), as well as for phototrophic bacteria (Hegler et al., 2008), and the spectral response of iron oxides is recognized in the 540, 640 and 900 nm bands, oxides associated with hematite in the 550 nm band, goethite around 440 to 470nm (Dematte et al. 2014). Fe oxides have also been recognized in the SWIR 1400 and 1900 nm bands (Dalmolin et al., 2005). However, some authors highlight the modification of these bands upon interaction with microbiological activity, for example, Fe oxides associated with fungal diversity in the bands 390, 410, 460 nm (Vis Nir) (Yang et al., 2022), CH-goethite bond ratio in 1725 nm (Sharma et al., 2021). Parikh et al. (2014) indicates that the interaction of phosphate groups presents in bacterial cell walls and goethite surfaces, which favors POFe bonds, are distinguished in the bands 1027, 1037 and 1045 cm^{-1} , in the Mid IR spectrum. Additionally, Rong et al. (2010) recognized the interaction of water and polymer bridges favoring bacteria-goethite bonds at the 1085 cm^{-1} band.

In general, as mentioned by Nath et al. (2021) the soil matrix provides and regulates the habitat of different microbial communities, therefore, reflectance spectroscopy analyses on soil samples contribute to the prediction of compounds used by microbes, and to the prediction of products of microbiological activity, since these analyses provide integrated measures of the mineral-organic composition of the soil (Viscarra Rossel et al., 2016). Rasche et al. (2013) points out some bands of the Mid IR spectra that favor the prediction of beta-glucosidase, xylosidase and urease, as well as bacterial abundance. In relation to Vis-Nir spectra some predictions associated with bacterial abundance (Viscarra et al. 2022., Zornoza et al. 2008), microbial biomass (Coûteaux et al. 2003, Zornoza et al. 2008, Chodak 2011), enzyme activity in soil (Cohen et al. 2005, Zornoza et al. 2008, Chodak 2011), respiration (Zornoza et al. 2008, Chodak 2011), fungal diversity and abundance (Yang et al., 2022).

NIR and Mid IR reflectance spectra respond to the concentration of compounds related to C-H, N-H, S-H, C=O and O-H chemical bonds, thus allowing to differentiate the organic composition of a soil sample (Zornoza et al. 2008, Parikh et al 2014, Viscarra Rossel et al. 2022). It is these compounds with which microbial properties (for example, soil microbial biomass or enzymatic activities) are closely related (Parikh et al., 2014). Because they are integrated with fungal and bacterial products. Ammann & Brandl (2011) indicate typical fingerprints for microorganisms by referring to the Mid IR spectrum bands associated with cellular carbohydrate and protein compounds. According to Jiang et al. (2004) hydroxyl, carboxyl, phosphoryl and amide groups are common among bacterial cell walls. In general the organic nature of bacteria and fungi makes peak locations similar between SOM and microbial samples. For this reason, many Vis NIR SWIR and Mid IR band assignments are common to the response of different soil organic compounds (Rasche et al. 2013, Parikh et al. 2014).

Spectral order and specific band identification is important in the prediction of soil properties, since characteristic peaks occur at different positions in the spectrum and may represent differences in information (Zhang et al., 2020). This spectral response information requires its mathematical extraction and subsequent correlation with soil properties (Reda et al., 2020), which is achieved by using machine learning techniques and multivariate regression analysis to extract the wave coverages related to the specific property to be predicted (Zornoza et al., 2008). For example, Rasche et al. (2013) developed a partial least squares regression (PLSR) analysis based on Mid IR reflectance spectroscopy to predict soil microbial biomass and enzyme activities.

Evaluating the interactions of organic compounds with mineral surfaces is fundamental to understand the stabilization of SOM (Parik et al., 2011). The charge of biomolecular functional groups on the surfaces of microorganisms and on minerals determines the adhesion process, being useful reflectance spectroscopy as it allows to evaluate such organo-mineral interactions facilitating the analysis of the binding mechanisms (Parik et al., 2014). Moreover, if this type of analysis is performed on the different fractions of the SOM, it favors the understanding of SOC dynamics (Six et al., 2004). The biogeochemical interaction between mineral particles and OM is a fundamental factor in the preservation of SOC and the study of the selective contribution of clay minerals and iron and aluminum oxides to this stabilization of organic compounds is favored by individual analyses of SOM fractions (Kirsten et al., 2021).

In general, traditional methods for analyzing soil microbiological activity are costly and time consuming, and given their importance in understanding C dynamics, soil quality and soil health, there is a need to develop techniques through simple and cost-effective methods (Rasche et al. 2013). Spectroscopy has come a long way in detecting specific bands in the Mid IR region.

In addition, enzyme activity has been recognized as a biological indicator that is closely related to soil physical and chemical properties, and allows predicting global microbial activity (Ma et al., 2021). Such activity influences C depletion and sequestration (Zhang et al., 2020), and is considered as one of the indicators with higher sensitivity to soil management practices than other variables (Adetunji et al., 2017).

Enzyme activity plays a vital role in agriculture and nutrient cycling (Balota and Chaves, 2010), especially hydrolases that are associated with carbon (betaglucosidase), nitrogen (urease) and phosphorus (phosphatase) cycles (Karaca et al., 2010), and are widely used as indicators of soil quality (Bandick and Dick, 1999).

Despite the importance of enzymatic activity of microorganisms as indicators of soil health, few efforts have been made to map these indicators, making it a challenge. However, digital soil mapping (DSM) techniques are available, which favor the construction of maps with high spatial resolution and low uncertainty, even at low sampling densities (Mendes et al., 2022). These DSM techniques combine soil point data with statistically correlated auxiliary data (covariates) (McBratney et al., 2003) and together with reflectance spectroscopy analyses associated with soil microbiological and mineral enzymatic activity, it is possible to generate methodologies for mapping, classification and monitoring (Di Iorio et al., 2019), of these soil properties, mainly by means of machine learning techniques.

Mathematical models related to DSM approaches have the ability to predict soil properties based on environmental covariates, through algorithms associated with these machine learning techniques (Hengl et al., 2015). Among these algorithms, the most widely used is Random Forest (RF) (Zeraatpisheh et al. 2020, Padarian et al. 2020), because it is a robust model that is composed of multiple decision trees that are not correlated with each other, which gives it a high accuracy in predictions with low possibility of overfitting (Wadoux et al., 2020).

The quality of DSM products is conditional on the environmental covariates used for modeling the attribute of interest. Usually environmental covariates representing physical and chemical processes associated with soil spatial variation (McBratney et al. 2003, Wadoux et al. 2020), and/or representing soil formation factors (Viscarra Rossel, 2011) are used, so

digital terrain elevation models, vegetation indices, climatic covariates, geological maps, satellite images of surface reflectance are commonly used (Ma et al., 2019). Recently, bare soil reflectance images have been included as environmental covariates (Rosin et al. 2023, Safanelli et al 2021b). These images are obtained using the GEOS3 (Geospatial Soil Sensing System) technique developed by Dematte et al. (2018), which allows capturing bare soil reflectance from historical series of Landsat images by adding pixels that were exposed at least once throughout the time series to a synthetic soil image (SYSI). Silvero et al. (2021), Rizzo et al. (2020), Rosin et al. (2023) have demonstrated the importance of this covariate in soil mapping due to its strong correlation with edaphic attributes.

This work is based on identifying the importance of mineralogy and microbiological activity in soil C sequestration potential, additionally mapping the importance of individual minerals that compose the clay fraction and microbiological activity and proposes the creation of microbiological activity indices based on reflectance spectroscopy, thus creating a cost-effective technique that contributes to the monitoring of indicators of the health of agricultural soils in Brazil.

Therefore, the thesis was divided into 4 scientific articles (Chapters) based on the main objectives:

- To analyze the individual contribution of each mineral composing the clay fraction in the sequestration potential of new SOC, through quantification, modeling and mapping of this potential in different pedogenetic soils of Brazil, using remote sensing products and the equation of Feller and Beare (1997) to obtain the theoretical potential of SOC sequestration.

- Mapping microbiological activity in different pedogenetic soils by reflectance spectroscopy and DSM approach and fitting the C sequestration potential model by including these microbiological variables in the spatial regression models and considering the interaction of these variables with mineralogy.

- Develop a strategy to analyze microbiological activity at the microscale by spectroscopic detection, through the interpretation of Vis NIR SWIR and Mid IR spectra of SOM fractions (particulate organic matter (POM) and SOM associated with the mineral fraction (MAOM)) based on the relationship of the spectra with the quantification of microbial biomass carbon (MBC) and enzymatic activity of beta-glucosidase, urease and phosphatase (which interfere in the carbon, nitrogen and phosphorus cycles).

- To develop a technique based on Vis NIR-SWIR and Mid IR reflectance spectroscopy products, DSM approach and machine learning techniques, to create and spatialize beta-glucosidase, phosphatase and urease enzyme activity indices for the Brazilian agricultural

territory, having as covariables the Synthetic Soil Image (SYSI), variables associated to relief, climate, biomes and mineralogical maps.

References

- ADAMCZYK, B., SIETIÖ, O.M., STRAKOVÁ, P., PROMMER, J., WILD, B., HAGNER, M., PIHLATIE, M., FRITZE, H., RICHTER, A. & J. HEINONSALO. 2019. Plant roots increase both decomposition and stable organic matter formation in boreal forest soil. *NATURE COMMUNICATIONS* (2019) 10:3982. <https://doi.org/10.1038/s41467-019-11993-1>
- ADETUNJI, A., LEWU, F., MULIDZI, R. AND B. NCUBE. 2017. The biological activities of β -glucosidase, phosphatase and urease as soil quality indicators: a review. *Journal of soil science and plant nutrition*, 17(3), 794-807. <https://dx.doi.org/10.4067/S0718-95162017000300018>.
- ADHIKARI, K., HARTEMINK, A., MINASNY, B., BOU KHEIR, R., GREVE, M. & M.H. GREVE. 2014. Digital mapping of soil organic carbon contents and stocks in Denmark *PLoS ONE*, 9 (8) Akpa et al. 2016
- ALBALADEJO, J., ORTIZ, R., GARCIA-FRANCO, N., NAVARRO, A., ALMAGRO, M., PINTADO, J. & M. MARTÍNEZ-MENA. 2013. Land use and climate change impacts on soil organic carbon stocks in semi-arid Spain. *J. Soils Sediments*, 13 (2) (2013), pp. 265-277, 10.1007/s11368-012-0617-7
- AMMANN, A. & H. BRANDL. 2011. Detection and differentiation of bacterial spores in a mineral matrix by Fourier transform infrared spectroscopy (FTIR) and chemometrical data treatment. *BMC Biophysics* 2011, 4:14. <http://www.biomedcentral.com/2046-1682/4/14> (14 July 2011)
- Angst et al., 2018
- ASHTON, N., TYE, A., PATTRICK, R. & B. VAN DONGEN. 2016. Carbon sequestration in the soils of Northern Ireland: potential based on mineralogical controls. In M.E. Young (ed.), *Unearthed: impacts of the Tellus surveys of the north of Ireland*. Dublin. Royal Irish Academy.
- BANDICK, A.K. AND R.P. DICK. 1999. Field management effects on soil enzyme activities. *Soil Biology and Biochemistry*. 31, 1471-1479.
- BUI, E., HENDERSON, B. & K. VIERGEVER. 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Glob. Biogeochem. Cycles*, 23 (4) (2009)
- CARVALHO, L., MONIZ, R., SOUZA, E., VIEIRA, G., SCHAEFER, C. & E. FERNANDES. 2019. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma* 340 (2019) 337–350.
- CHAPLOT, V., BOUAHOM, B. & C. VALENTIN. 2010. Soil organic carbon stocks in Laos: spatial variations and controlling factors *Glob. Chang. Biol.*, 16 (4) (2010), pp. 1380-1393
- CHODAK, M. 2011. Near-infrared spectroscopy for rapid estimation of microbial properties in reclaimed mine soils. *Journal of Plant Nutrition and Soil Science* 174, 702-709.
- COHEN, M.J., PRENGER, J.P. & W.F. DEBUSK. 2005. Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. *Journal of Environmental Quality* 34, 1422e1434.

- COÛTEAUX, M.-M., BERG, B. AND P. ROVIRA. 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. *Soil Biology & Biochemistry*. 2003; 35:1587–1600.
- DALMOLIN, R.S.D., GONÇALVES, C.N., KLAMT, E. & D.P. DICK. 2005. Relação entre os constituintes do solo e seu comportamento espectral. *Ciência Rural* 35, 481–489. <https://doi.org/10.1590/s0103-84782005000200042>.
- DE BEECK, M., PERSSON, P. & A. TUNLID. 2021. Fungal extracellular polymeric substance matrices – Highly specialized microenvironments that allow fungi to control soil organic matter decomposition reactions. *Soil Biology and Biochemistry* 159 (2021) 108304. <https://doi.org/10.1016/j.soilbio.2021.108304>
- DEMATTÊ, J.A.M., BELLINASSO, H., ROMERO, D.J. AND C.T. FONGARO. 2014. Morphological interpretation of reflectance spectrum (MIRS) using libraries looking towards soil classification. *Sci. Agric.* 71, 509–520. <https://doi.org/10.1590/0103-9016-2013-0365>.
- DEMATTÊ, J., FONGARO, C., RIZZO, C. & J.L. SAFANELLI. 2018. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.*, 212 (2018), pp. 161-175.
- DE MASTRO, F., TRAVERSA, A., COCOZZA, C., PALLARA, M. & G. BRUNETTI. 2020. Soil Organic Carbon Stabilization: Influence of Tillage on Mineralogical and Chemical Parameters. *Soil Syst.* 2020, 4(3), 58
- DI IORIO, E., CIRCELLI, L., LORENZETTI, R., COSTANTINI, E., PERL, S. & C. COLOMBO. 2019. Estimation of andic properties from Vis-NIR diffuse reflectance spectroscopy for volcanic soil classification. *Catena* 182 (2019) 104109
- DOS REIS, C., PINHEIRO, D., CALDAS, J. & C. BAYER. 2014. Carbon sequestration in clay and silt fractions of Brazilian soils under conventional and no-tillage systems. *Soils and Plant Nutrition. Sci. agric. (Piracicaba, Braz.)* 71 (4).
- DORAN, J.W., ZEISS, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. *Applied Soil Ecology*. Vol. 15
- ESWARAN, H., VANDENBERG, E. & P. REICH. 1993. Organic-carbon in soils of the world *Soil Sci. Soc. Am. J.*, 57 (1993), pp. 192-194
- FELLER, C. & M.H. BEARE. 1997. Physical control of soil organic matter dynamics in the tropics. *Geoderma*, 79 (1–4) (1997), pp. 69-116
- FISSORE, C., DALZELLB, J., BERHE, A., VOEGTLE, M., EVANS, M. & A. WU. 2017. Influence of topography on soil organic carbon dynamics in a Southern California grassland. *Catena*, 149 (2017), pp. 140-149
- FOLLETT, R., KIMBLE, J. & R. LAL. 2001. The potential of US grazing lands to sequester soil carbon. Lewis Publishers, Boca Raton, Florida (2001), pp. 401-430
- FORKUOR, G., HOUNKPATIN, O., WELP, G. & M. THIEL. 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) (2017)

- GOUDA, S., KERRY, R., DAS, G., PARAMITHIOTIS, S., SHIN, H. & J. PATRA. 2018. Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture. *Microbiological Research* 206 (2018) 131–140. <http://dx.doi.org/10.1016/j.micres.2017.08.016>
- GRAY, J., BISHOP, T. & B.R. WILSON. 2015. Factors controlling soil organic carbon stocks with depth in Eastern Australia *Soil Sci. Soc. Am. J.*, 79 (6) (2015), pp. 1741-1751
- GRÜNEBERG, E., SCHÖNING, I., HESSENMÖLLER, D., SCHULZE, E. & W.W. WEISSER. 2013. Organic layer and clay content control soil organic carbon stocks in density fractions of differently managed German beech forests. *For. Ecol. Manag.*, 303 (2013), pp. 1-10
- GUEVARA, M., ARROYO, C., BRUNSELL, N., CRUZ, C.O., DOMKE, G., EQUIHUA, J., ET AL. 2020. Soil organic carbon across Mexico and the conterminous United States (1991–2010). *Global Biogeochem. Cy.* 34, e2019GB006219. [10.1029/2019GB006219](https://doi.org/10.1029/2019GB006219).
- HART, M. M., CROSS, A. T., D'AGUI, H. M., DIXON, K. W., VAN DER HEYDE, M., MICKAN, B., HORST, C., GREZ, B. M., VALLIERE, J. M., VISCARRA ROSSEL, R. A., WHITELEY, A., WONG, W. S., ZHONG, H. & P. NEVILL. 2020. Examining assumptions of soil microbial ecology in the monitoring of ecological restoration, *Ecol. Solut. Evid.*, 1, e12031, <https://doi.org/10.1002/2688-8319.12031>, 2020
- HASSINK, J. 1994. Effect of soil texture on the size of the microbial biomass and on the amount of c and n mineralized per unit of microbial biomass in dutch grassland soils. *Soil Biol Biochem* 26: 1573-1581
- HASSINK, J. 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant and Soil*, 191 (1997), pp. 77-87, [10.1023/A:1004213929699](https://doi.org/10.1023/A:1004213929699)
- HEGLER, F., POSTH, N., JIANG, J. & A. KAPPLER. 2008. Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. *FEMS Microbiol Ecol* 66 (2008) 250–260. DOI:10.1111/j.1574-6941.2008.00592.x
- HENGL, T., HEUVELINK, G., KEMPEN, B., LEENAARS, J., WALSH, M., SHEPHERD, K., SILA, A., MACMILLAN, R., MENDES DE JESUS, J., TAMENE, L. & J.E. TONDOH. 2015. Mapping soil properties of Africa at 250 m resolution: Random Forests significantly improve current predictions. *PLoS One*, 10 (6) (2015), Article e0125814, [10.1371/journal.pone.0125814](https://doi.org/10.1371/journal.pone.0125814)
- HENGL, T., MENDES, J., HEUVELINK, G., GONZALEZ, M., KILIBARDA, M., BLAGOTIĆ, A. et al. 2017. SoilGrids250m: Global gridded soil information based on Machine Learning *PLoS ONE*, 12 (2) (2017), Article e0169748, [10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748)
- HOBLEY, E., WILSON, B., WILKIE, A., GRAY, J AND T. KOEN. 2015. Drivers of soil organic carbon storage and vertical distribution in Eastern Australia. *Plant Soil*, 390 (2015), pp. 111-127
- HOUGHTON, R.A. 2003. Why are estimates of the terrestrial carbon balance so different?. *Global Change Biology*. 9, 500-509
- HOUNKPATIN, O., DE HIPT, F., BOSSA, A., WELP, G. AND W. AMELUNG. 2018. Soil organic carbon stocks and their determining factors in the Dano catchment(Southwest Burkina Faso). *Catena* 166 (2018) 298–309. <https://doi.org/10.1016/j.catena.2018.04.013>
- INGRAM, J. & E.C.M. FERNANDES. 2001. Managing carbon sequestration in soils: concepts and terminology *Agric. Ecosyst. Environ.*, 87 (1) (2001), pp. 111-117

- JIANG, W., SAXENA, A., SONG, B., WARD, B.B., BEVERIDGE, T.J. AND S.C. MYNENI. 2004. Elucidation of functional groups on gram-positive and gram-negative bacterial surfaces using infrared spectroscopy. *Langmuir* 20, 11433–11442.
- JOBÁGY, E. & R.B. JACKSON. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation *Ecol. Appl.*, 10 (2000), pp. 423-436
- KARACA, A., CETIN, S.C., TURGAY, O.C. AND R. KIZILKAYA. 2010. Soil enzymes as indication of soil quality. *Soil enzymology*. Springer.
- KAISER, K. & W. ZECH, W. 2000. ‘Dissolved organic matter sorption by mineral constituents of subsoil clay fractions’, *Journal of Plant Nutrition and Soil Science*, 163, 5, 531–5.
- KALBITZ, K., SCHWESIG, D., RETHEMEYER, J. & E. MATZNER. 2005. Stabilization of dissolved organic matter by sorption to the mineral soil. *Soil Biol. Biochem.* 37, 1319-1331.
- KIMBLE, J.M., HEATH, L.S., BIRDSEY, R.A., LAL, R. 2003. *The Potential of U.S. Forest Soils to Sequester Carbon and Mitigate the Greenhouse Effect*. CRC Press, New York.
- KIRSTEN, M., MIKKUTA, R., VOGEL, C., THOMPSON, A., MUELLER, C., KIMARO, D., BERGSMAN, H., FEGER, K. & K. KALBITZ. 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. *Scientific Reports* | (2021) 11:5076
- KOBLER, J., ZEHETGRUBER, B., DIRNBÖCK, T., JANDL, R., MIRTL, M. & A. SCHINDLBACHER. 2019. Effects of aspect and altitude on carbon cycling processes in a temperate mountain forest catchment *Landscape Ecol.*, 34 (2019), pp. 325-340
- LADD, B., LAFFAN, S., AMELUNG, W., PERI, P., LUCAS C.R. SILVA, P., GERVASSI, P., BONSER, S., NAVALL, M. & D. SHEIL. 2013. Estimates of soil carbon concentration in tropical and temperate forest and woodland from available GIS data on three continents. *Glob. Ecol. Biogeogr.*, 22 (2013), pp. 461-469
- LAL R. 2004. Soil carbon sequestration impacts on global climate change and food security *Science*, 304 (2004), pp. 1623-1627
- LAMICHHANE, S., KUMAR, L. & B. WILSON. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* Volume 352, 15 October 2019, Pages 395-413
- LI, M., Han, X., Du. And L. Li. 2019. Profile stock of soil organic carbon and distribution in croplands of Northeast China. *Catena* 174 (2019) 285–292. <https://doi.org/10.1016/j.catena.2018.11.027>
- MA, Y., MINASNY, B., MALONE, B.P. AND A.B. MCBRATNEY. 2019. Pedology and digital soil mapping (DSM). *Eur J Soil Sci* 70, 216–235. <https://doi.org/10.1111/EJSS.12790>
- MA, C.Y., CI, K.D., ZHU, J., SUN, Z.L., LIU, Z.X., LI, X.Y., TANG, C., WANG, P. & Z.M. LIU. 2021. Impacts of exogenous mineral silicon on cadmium migration and transformation in the soil-rice system and on soil health. *Sci Total Environ.* 759, 143501.
- B. MARSCHNER, S. BRODOWSKI, A. DREVES, G. GLEIXNER, A. GUDE, P.M. GROOTES, U. HAMER, A. HEIM, G. JANDL, R. JI. 2008. How relevant is recalcitrance for the stabilization of organic matter in soils? *J. Plant Nutr. Soil Sci.*, 171 (2008), pp. 91-110

- MAYER, L. 1994. Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. *Chem. Geol.*, 114 (3) (1994), pp. 347-363
- MCBRATNEY, A., SANTOS, M. & B. MINASNY. 2003. On digital soil mapping. *Geoderma*, 117 (2003), pp. 3-52
- MENDES, W., DEMATTE, J.A.M., ROSIN, N., TERRA, F., POPPIEL, R., URBINA-SALAZAR, D., BOECHAT, C., SILVA, E., CURTI, N., SILVA, S., SANTOS, U. & G. VALLADARES. 2022. The Brazilian soil Mid-infrared Spectral Library: The Power of the Fundamental Range. *Geoderma* 415 (2022) 115776. <https://doi.org/10.1016/j.geoderma.2022.115776>
- MINASNY, B., MCBRATNEY, A. & B.P. MALONE. 2013. Wheeler Digital mapping of soil carbon. *Adv. Agron.*, 118 (2013), pp. 1-47
- 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. & L. WINOWIECKI. 2017. Soil carbon 4 per mille. *Geoderma* 292, 59–86. <https://doi.org/10.1016/J.GEODERMA.2017.01.002>
- MUELLER, K., HOBBIE, S., CHOROVER, J., REICH, P., EISENHAUER, N., CASTELLANO, M., CHADWICK, O., DOBIES, T., HALE, C., JAGODZIŃSKI, A., KAŁUCKA, I., KIELISZEWSKA-ROKICKA, B., MODRZYŃSKI, J., ROŻEN, A., SKORUPSKI, M., SOBCZYK, L., STASIŃSKA, M., TROCHA, L., WEINER, J., WIERZBICKA, A. & J. OLEKSYN. 2015. Effects of litter traits, soil biota, and soil chemistry on soil carbon stocks at a common garden with 14 tree species. *Biogeochemistry*, 123 (2015), pp. 313-327, 10.1007/s10533-015-0083-6
- MÜLLER, B., 2015. Experimental interactions between clay minerals and bacteria: a review. *Pedosphere* 25, 799–810.
- NATH, D., LAIK, R., MEENA, V., PRAMANICK, B. & SINGH. 2021. Can mid-infrared (mid-IR) spectroscopy evaluate soil conditions by predicting soil biological properties?. *Soil Security* 4 (2021) 100008. <https://doi.org/10.1016/j.soisec.2021.100008>
- NICOLAS, C., MARTINBERTELSEN, T., FLOUDAS, D., BENTZER, J., SMITS, M.M., JOHANSSON, T., TROEIN, C., PERSSON, P. & A. TUNLID. 2019. The soil organic matter decomposition mechanisms in ectomycorrhizal fungi are tuned for liberating soil organic nitrogen. *The ISME Journal* 13, 977–988.
- NYSSSEN, J., TEMESGEN, H., LEMENIH, M., ZENEBE, A., HAREGEWEYN, N. & M. HAILE. 2008. Spatial and temporal variation of soil organic carbon stocks in a lake retreat area of the Ethiopian Rift Valley. *Geoderma*, 146 (1–2) (2008), pp. 261-268
- OJEDA, J.J., ROMERO-GONZALEZ, M.E., POURAN, H.M. & S.A. BANWART. 2008. In situ monitoring of the biofilm formation of *Pseudomonas putida* on hematite using flow-cell ATR-FTIR spectroscopy to investigate the formation of inner-sphere bonds between the bacteria and the mineral. *Mineral. Mag.* 72, 101–106.
- OUESLATI, I., ALLAMANO, P., BONIFACIO, E. & P. CLAPS. 2013. Vegetation and topographic control on spatial variability of soil organic carbon. *Pedosphere*, 23 (1) (2013), pp. 48-58

- PADARIAN, J., MINASNY, B. & A.B. MCBRATNEY. 2020. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* 6, 35–52. <https://doi.org/10.5194/soil-6-35-2020>
- PARIKH, S., GOYNE, K., MARGENOT, A., MUKOME, F. & F. CALDE-RON. 2014. Soil Chemical Insights Provided through Vibrational Spectros-copy. Publications from USDA-ARS / UNL Faculty. Paper 1471.
- POEPLAU, C., DON, A., VESTERDAL, L., LEIFELD, J., VAN WESEMAEL, B., SCHUMACHER, J. & A. GENSIOR. 2011. Temporal dynamics of soil organic carbon after land-use change in the temperate zone — carbon response functions as a model approach. *Glob. Chang. Biol.*, 17 (2011), pp. 2415-2427
- POWERS, J., CORRE, M., TWINE, T. & E. VELDKAMP. 2011. Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation. *Proc. Natl. Acad. Sci.*, 108 (15) (2011), pp. 6318-6322
- PROUT, J.M., SHEPHERD, K.D., MCGRATH, S.P., KIRK, G.J. & S.M. HAEFELE. 2021. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. *European Journal of Soil Science*, 72(6), pp.2493-2503. <https://doi.org/10.1111/ejss.13012>
- RASSE, D., RUMPEL, C. AND M.F. DIGNAC. 2005. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant and Soil*, 269 (2005), pp. 341-356, 10.1007/s11104-004-0907-y
- RASCHE, F., MARHAN, S., BERNER, D., KEIL, D., KANDELER, E. & G. CADISCH. 2013. midDRIFTS-based partial least square regression analysis allows predicting microbial biomass, enzyme activities and 16S rRNA gene abundance in soils of temperate grasslands. *Soil Biology & Biochemistry* 57 (2013) 504-512
- RASCHE, F., KNAPP, D., KAISER, C., KORANDA, M., KITZLER, B., ZECHMEISTER-BOLTENSTERN, S., RICHTER, A. AND A. SESSITSCH. 2011. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *The ISME Journal* 5, 389-402.
- RENTSCHLER, T., GRIES, P., BEHRENS, T., BRUELHEIDE, H., KUHN, P., SEITZ, S., ET AL. 2019. Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. *PLoS ONE* 14, 8, e0220881, 2019. 10.1371/journal.pone.0220881
- RIAL, M., MARTÍNEZ CORTIZAS, A. & L. RODRÍGUEZ-LADO. 2017. Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils *Sci. Total Environ.*, 609 (2017), pp. 1411-1422
- RINNAN, R. & A. RINNAN. 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. *Soil Biol. Biochem.* 39 (7), 1664–1673. <https://doi.org/10.1016/j.soilbio.2007.01.022>.
- RIZZO, R., MEDEIROS, L.G., MELLO, D.C. DE, MARQUES, K.P.P., MENDES, W. DE S., QUIÑONEZ SILVERO, N.E., DOTTO, A.C., BONFATTI, B.R., DEMATTÊ, J.A.M. 2020. Multi temporal bare surface image associated with transfer functions to support soil classification and mapping in southeastern Brazil. *Geoderma* 361, 114018. <https://doi.org/10.1016/J.GEODERMA.2019.114018>

- RONG, X., CHEN, W., HUANG, Q., CAI, P. & W. LIANG. 2010. Pseudomonas putida adhesion to goethite: studied by equilibrium adsorption, SEM, FTIR and ITC. *Colloids Surf. B Biointerfaces* 80, 79–85.
- ROSIN, N.A., DEMATTÊ, J.A.M., POPPIEL, R.R., SILVERO, N.E.Q., RODRIGUEZ ALBARRACIN, H.S., ROSAS, J.T.F., GRESCHUK, L.T., BELLINASSO, H., MINASNY, B., GOMEZ, C., MARQUES JÚNIOR, J., FERNANDES, K. 2023. Mapping Brazilian soil mineralogy using proximal and remote sensing data. *Geoderma* 432, 116413. <https://doi.org/10.1016/J.GEODERMA.2023.116413>
- SAFANELLI, J.L., DEMATTÊ, J.A.M., DOS SANTOS, N.V., ROSAS, J.T.F., SILVERO, N.E.Q., BONFATTI, B.R. AND W. DE S. MENDES. 2021b. Fine scale soil mapping with Earth Observation data: a multiple geographic level comparison. *Rev Bras Cienc Solo* 45, e0210080. <https://doi.org/10.36783/18069657RBCS20210080>
- SARKHEIL, H., TAHERY, B., RAYEGANI, B., RAMEZANI, J., GOSHTASB, H. & A. JAHANI. 2020. Evaluating the current status of the national health, safety, and environment management system for integration, harmonization, and standardization of environmental protection. *Health Risk Analysis* 18–24.
- SARKHEIL, H., NOUGHABI, K., AZIMI, Y. & S. RAHBARI. 2023. Fuzzy soil quality index using resistivity and induced polarization for contamination assessment in a lead and zinc drainage irrigation field study. *Ecological Indicators* 152 (2023) 110362
- SHARMA, V., CHAUHAN, R. & R. KUMAR. 2021. Spectral characteristics of organic soil matter: A comprehensive review. *Microchemical Journal* 171 (2021) 106836. <https://doi.org/10.1016/j.microc.2021.106836>
- SCHINDLBACHER, a., DE GONZALO, C., DÍAZ-PINÉS, E., GORRÍA, P., MATTHEWS, B., INCLÁN, R., ZECHMEISTER-BOLTENSTERN, S., RUBIO, AND R. JANDL. 2010. Temperature sensitivity of forest soil organic matter decomposition along two elevation gradients *J Geophys Res* (2010), 10.1029/2009JG001191
- SHI, Z., ALLISON, S. D., HE, Y., LEVINE, P. A., HOYT, A. M., BEEM-MILLER, J., ZHU, Q., WIEDER, W. R., TRUMBORE, S., AND RANDERSON, J. T.: The age distribution of global soil carbon inferred from radiocarbon measurements, *Nature Geoscience*, <https://doi.org/10.1038/s41561-020-0596-z>, 2020.
- SINGH, M., SARKAR, B., SARKAR, S., CHURCHMAN, J., BOLAN, N., MANDAL, S., MENON, M., PURAKAYASTHA, T. J., & BEERLING, D. J. 2018. Stabilization of soil organic carbon as influenced by clay mineralogy. *Advances in Agronomy*, 148, 33–84.
- SILVERO, N.E.Q., DEMATTE, J.A.M., AMORIM, M.T.A., SANTOS, N.V. DOS, RIZZO, R., SAFANELLI, J.L., POPPIEL, R.R., MENDES, W. DE S., BONFATTI, B.R., 2021. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. *Remote Sens. Environ.* 252, 112117. 10.1016/j.rse.2020.112117.
- SILVERO, N.E.Q., DEMATTE, J.A.M., MINASNY, B., ROSIN, N.A., NASCIMENTO, J.G., RODRIGUEZ-ALBARRACIN, H.S., BELLINASSO, H. AND A.M. GÓMEZ. 2023. Chapter Three - Sensing technologies for characterizing and monitoring soil functions: A review. Editor(s): Donald L. Sparks. *Advances in Agronomy*, Academic Press, Volume 177, 2023, Pages 125-168, ISSN 0065-2113, ISBN 9780443192586, <https://doi.org/10.1016/bs.agron.2022.08.002>.

- SIX, J., CONANT, R., PAUL, E. & K. PAUSTIAN. 2002. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils *Plant Soil*, 241 (2002), pp. 155-176
- SOTHE, C., GONSAMO, A., ARABIAN, J. & J. SNIDER. 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. *Geoderma* Volume 405, 1 January 2022, 115402
- VESTERDAL, L., ELBERLING, B., CHRISTIANSEN, J., CALLESEN, I. & I. SCHMIDT. 2012. Soil respiration and rates of soil carbon turnover differ among six common European tree species. *Forest Ecology and Management*, 264, 185–196
- VISCARRA ROSSEL, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *J. Geophys. Res.* 116, F04023. <https://doi.org/10.1029/2011JF001977>.
- VISCARRA ROSSEL, R.A. AND W.S. HICKS. 2015. Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. *European Journal of Soil Science* 66, 438–450. <https://doi.org/10.1111/ejss.12237>.
- VISCARRA ROSSEL, R. A., BEHRENS, T., BEN-DOR, E., BROWN, D., DEMATTÊ, J., SHEPHERD, K., SHI, Z., STENBERG, B., STEVENS, A., ADAMCHUK, V., AICHI, H., BARTHES, B., BARTHOLOMEUS, H., BAYER, A., BERNOUX, M., BOTTCHEK, K., BRODSKY, L., DU, C., CHAPPELL, A., FOUAD, Y., GENOT, V., GOMEZ, C., GRUNWALD, S., GUBLER, A., GUERRERO, C., HEDLEY, C., KNADEL, M., MORRAS, H., NOCITA, M., RAMÍREZ LÓPEZ, L., ROUDIER, P., CAMPOS, E., SANBORN, P., SELLITTO, V., SUDDUTH, K., RAWLINS, B., WALTER, C., WINOWIECKI, L., HONG, S. AND 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>.
- VISCARRA-ROSSEL, R.A., YANG, Y., BISSETT, A., BEHRENS, T., DIXON, K., NEVIL, P. & S. LI. 2022. Environmental controls of soil fungal abundance and diversity in Australia's diverse ecosystems. *Soil Biology and Biochemistry* 170 (2022) 108694. <https://doi.org/10.1016/j.soilbio.2022.108694>
- VON LÜTZOW, M., KÖGEL-KNABNER, I., LUDWIG, B., MATZNER, E., FLESSA, H., EKSCHEMITT, K., GUGGENBERGER, G., MARSCHNER, B. & K. KALBITZ. 2008. Stabilization mechanisms of organic matter in four temperate soils: development and application of a conceptual model. *Journal of Plant Nutrition and Soil Science*, 171 (2008), pp. 111-124, 10.1002/jpln.200700047
- WADOUX, A.-M.-J.-C., MINASNY, B. AND A.B. MCBRATNEY. 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Rev.* 210, 103359 <https://doi.org/10.1016/j.earscirev.2020.103359>.
- WAGAI, R., MAYER, L.M., KITAYAMA, K. AND H. KNICKER. 2008. Climate and parent material controls on organic matter storage in surface soils: a three- pool, density-separation approach. *Geoderma*, 147 (2008), pp. 23-33, 10.1016/j.geoderma.2008.07.010
- WEBER, K. A., ACHENBACH, L. A. & J. COATES. 2006. Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat. Rev. Microbiol.* 4, 752–764. doi: 10.1038/nrmicro1490
- WEIL, R. AND N.C. BRADY. 2016. *The Nature and Properties of Soils* (15th edition), Pearson Education (2016)

- WIESMEIER, M., BARTHOLD, F., BLANK, B. & I. KÖGEL-KNABNER. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem *Plant Soil*, 340 (1–2) (2011), pp. 7–24
- WIESMEIER, M., SPOERLEIN, P., GEUSS, U., HANGEN, E., HAUG, S., REISCHL, A., SCHILLING, B., VON LUETZOW, M. & I. KOEGEL-KNABNER. 2012. Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling Depth. *Glob. Chang. Biol.*, 18 (2012), pp. 2233–2245
- WIESMEIER, M., HUBNER, R., BARTHOLD, F. ET AL. 2013. Amount, distribution and driving factors of soil organic carbon and nitrogen in cropland and grassland soils of southeast Germany (Bavaria). *Agriculture Ecosystems & Environment*, 176, 39–52
- XU, X., LIU, W., ZHANG, C. & G. KIELY. 2011. Estimation of soil organic carbon stock and its spatial distribution in the Republic of Ireland. *Soil Use Manag.*, 27 (2011), pp. 156–162
- YANG, J., ZHANG, X., BOURG, A. AND H. STONE. 2021. 4D imaging reveals mechanisms of clay-carbon protection and release. *Nat Commun* 12, 622 (2021). <https://doi.org/10.1038/s41467-020-20798-6>
- YANG, Y., SHEN, Z., BISSETT, A. & R. VISCARRA-ROSSEL. 2022. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. *SOIL*, 8, 223–235, 2022. <https://doi.org/10.5194/soil-8-223-2022>
- ZERAATPISHE, M. & F. KHORMALI. 2012. Carbon stock and mineral factors controlling soil organic carbon in a climatic gradient, Golestan province. *Journal of soil science and plant nutrition*, 12(4), 637–654.
- ZERAATPISHEH, M., JAFARI, A., BODAGHABADI, M.B., AYOUBI, S., TAGHIZADEH-MEHRJARDI, R., TOOMANIAN, N., KERRY, R. & M. XU. 2020. Conventional and digital soil mapping in Iran: past, present, and future *Catena*, 188 (2020), p. 104424
- ZHANG, L., CHEN, X., XU, Y., JIN, M., YE, X., GAO, H., CHU, W., MAO, J. & M. THOMPSON. 2020. Soil labile organic carbon fractions and soil enzyme activities after 10 years of continuous fertilization and wheat residue incorporation. *Scientific Reports* (2020) 10:11318. <https://doi.org/10.1038/s41598-020-68163-3>.
- ZORNOZA, R., GUERRERO, C., MATAIX-SOLERA, J., SCOW, K., ARCENEGUI, V. & J. MATAIX-BENEYTO. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem.* 2008 July; 40(7): 1923–1930. doi:10.1016/j.soilbio.2008.04.003.

2. POTENTIAL OF SOIL MINERALS TO SEQUESTER SOIL ORGANIC CARBON

Abstract

The capacity of soil to sequester carbon (C) is a key process that promotes the reduction of CO₂ in the atmosphere. Soils can absorb as much as 20% of anthropogenic carbon emissions, which can contribute to mitigate climate change. This capacity relies on the organo-mineral association, which includes different minerals, Fe and Al oxides, which have a critical soil organic carbon (SOC) sorption surface. Based on an equation of the potential C saturation deficit of fine soil particles (<20 µm/silt and clay fractions) for tropical regions, this study investigated the SOC sequestration potential of the clay fraction for soils in Piracicaba region, São Paulo State, Brazil as influenced by the clay minerals. This potential was fitted to a spatial regression model for soil depths 0 - 20 cm and 80 to 100 cm. In the surface layer, the sequestration potential was mostly explained by the relative abundance of soil minerals (Kaolinite, Hematite, Goethite and Gibbsite) determined using Vis-NIR-SWIR spectroscopy. A direct relationship was observed with Goethite and Gibbsite, indicating that low concentrations would reduce the sequestration potential. At 80 to 100 cm depth, Kaolinite and Hematite explained most variation in SOC sequestration potential. Additionally, the C associated with the mineral fraction and the C saturation potential as a function of minerals were modeled and a strong importance of hematite in the C sequestration and stabilization cycle was identified at both depths. The individual mineral contribution to SOC sequestration potential was also mapped, which identified high contributions of goethite and gibbsite for deep soil layers. The influence of land use on the carbon sequestration potential of minerals was observed, with the greatest potential being found in areas with pasture and cropping mosaics and grassland and forest mosaics, with a high presence of kaolinite and hematite. These minerals have a greater potential for carbon sequestration at greater depths and, therefore, could be critical in climate change mitigation strategies.

Keywords: Spatial regression, Digital soil mapping, Sustainability, Soil security, Soil spectroscopy, Saturation deficit

Published as: RODRÍGUEZ-ALBARRACÍN, H., DEMATTÊ, J., ROSIN, N., DARGHAN, E., SILVERO, N., CERRI, C., MENDES, W. & M. TAYEBI. 2023. Potential of soil minerals to sequester soil organic carbon. *Geoderma* 436 (2023) 116549. <https://doi.org/10.1016/j.geoderma.2023.116549>

2.1. Introduction

The ability of soil to sequester carbon is considered a cost-effective and plausible method to reduce the concentration of CO₂ in the atmosphere (Houghton 2003, Kimble et al. 2003). This is because global soils have the potential to absorb about 20% of anthropogenic carbon emissions (Yang et al., 2021). Therefore, carbon sequestration is a phenomenon that can help to partially mitigate climate change (Padarian et al., 2022), as for greenhouse gas emissions (Minasny et al., 2017).

Hassink (1997) and Loveland and Webb (2003) proposed that soils have a limited capacity to retain carbon, which is based on the reactive capacity of mineral surfaces (Churchman et

al., 2020; Prout et al., 2021). Therefore, it is understood that the clay fraction has a finite carbon storage capacity (Diekow et al., 2005), and the search for this storage limit justifies the determination of carbon (C) sequestration potential (Six et al. 2002; Goh 2004; Stewart et al. 2008 and Chung et al. 2008), which also depends on the limited potential of the soil to stabilize soil organic carbon (SOC) against microbial mineralization.

The SOC stabilization is based on selective preservation associated with recalcitrance or chemical resistance (Singh et al., 2018). It is also related to the spatial inaccessibility of soil organic matter (SOM) by occlusion in soil aggregates, especially in microaggregates (Hoffland et al., 2020). Furthermore, the interaction with mineral surfaces favors the formation of organic-mineral complexes (Sollins et al. 1996, Von Lützow et al. 2006), especially with clay minerals and metal ions (Oades 1988, Arrouays et al. 2006, Singh et al., 2018).

Hassink (1997) indicated that the potential C saturation is associated with silt and clay particles. Angers et al. (2011) also pointed out that fine-textured soils have a higher retention capacity than sandy soils, due to a larger specific surface area available for organo-mineral interactions present in the silt and clay fractions (Stewart et al., 2008). Similarly, Zeraatpishe and Khormali (2012) observed that the SOC can be adsorbed by coarse aggregates, fine aggregates, and particles smaller than 0.053 mm. Brodowski et al. (2006) found that particles such as clay and silt alone cannot retain much SOC, and more than 90% is stored in the aggregates, which protects SOC from microbial decomposition (Hoffland et al. 2020, Baldock & Skjemstad 2000, Lützow et al. 2006).

The SOC dynamics is closely related to the soil development process and, therefore, to its formation factors (Hobley et al., 2015). There are active factors, such as climate and organisms, and passive factors, such as parent materials, relief, and time. To make predictions of the terrestrial carbon cycle, it is necessary to understand the complete processes related to the sequestration and release of carbon (Marschner et al., 2008).

Ingram and Fernandes (2001) and Weil and Brady (2016) indicated the importance of mineralogy on the potential for SOC storage, especially soils in the deeper layers (Gray et al. 2015, Wiesmeier et al. 2011). In contrast, land use presents a more significant influence in the superficial layers (Adhikari et al. 2014, Hobley et al. 2015). For Ashton et al. (2016), clay content alone is not related to the increase or decrease of C content. The concentration is influenced by soil mineralogy and the geochemistry of the soil, thus depending on the presence of clay minerals and iron oxides of high specific surface area and pH. The clay fraction is key to C sequestration because it involves different minerals and varying amounts

of pedogenic Fe and Al oxides (Yang et al., 2021; Kirsten et al., 2021; Prout et al., 2021). Minerals such as iron and aluminum oxides, specifically goethite and gibbsite are recognized as critical C sorption surfaces (Kaiser and Zech, 2000; Dos Reis et al., 2014).

Feng et al. (2005) highlighted the adsorption of anionic cations by SOM through ionic linkages related to hydrogen, cation and anion exchange, ligand exchange and cation bridges, likewise electrostatic attractions or van der Waals linkages alone can occur. Soil oxides, oxyhydroxides and hydroxides present electrostatic attractions and ionic linkages between the hydroxyl groups of the oxides and the carboxyl or hydroxyl groups of the SOM, with additional strong relationships between iron manganese and humic substances (De Mastro et al., 2020). This adsorption of SOM by minerals favors SOC stabilization, reducing microbial mineralization (Kalbitz et al., 2005), due to the specific surface area of these minerals and their surface charge that favors these linkages and stabilizes SOM.

Interactions of the reactive phases of poorly crystalline Fe and Al oxides with SOM are an essential mechanism in the long-term stabilization of SOC (Kögel-Knabner et al. 2008, Percival et al. 2000). According to Weber et al. (2006), the biogeochemical cycling of iron is closely related to the dynamics of SOM. Lalonde et al. (2012) noted that about 21.5% of global SOC is associated with reactive forms of Fe. Crystalline Fe and Al oxides present reactive sites on the surface that can adsorb SOC (De Mastro et al., 2020), however Duiker et al. (2003) and De Mastro et al. (2020) observed that Fe oxides of low degree of crystallinity stabilize SOM more effectively than crystalline Fe oxides or oxyhydroxides, because they present higher specific surface area and density of hydroxyl sites compared to crystalline ones, increasing their chelation capacity (Wen et al., 2019). Zeraatpishe and Khormali (2012), to the contrary, reported that amorphous and crystalline iron oxides and hydroxides retain 50-70% of total carbon.

Poorly crystalline Fe minerals have a specific surface area of around $800 \text{ m}^2 \text{ g}^{-1}$, for example, for ferrihydrite, higher than crystalline forms of Fe found around $200 \text{ m}^2 \text{ g}^{-1}$, as in the case of goethite (Eusterhues et al., 2005). According to Churchman and Velde (2019), SOC shows a preference for weakly crystalline oxides as well as Fe and Al silicates. Bonds with 2:1 phyllosilicates are going to depend on their relative surface reactivities. Therefore, minerals such as smectite that are more reactive present greater potential to retain C (Churchman et al., 2020). However, a more significant effect occurs with poorly crystalline Fe oxyhydroxides (Rasmussen et al., 2007). The combined effect of silicates and oxides is involved in SOC stabilization, for example, ferrihydrite and goethite can favor the sorption capacity of kaolinite (De Mastro et al., 2020).

Kirsten et al. (2021) indicated that clay minerals and oxyhydroxides (pedogenic metal oxides) are the most reactive and control the persistence of SOC. However, most studies did not explicitly evaluate the contribution of clay minerals on SOC. For example, Weismeyer et al. (2013) estimated the SOC sequestration potential of silt and clay particles, in soils from Germany, without specifying the clay mineral type. Ashton et al. (2016) analyzed SOC concentrations in different parent materials and clay mineralogies, evaluating total concentrations without determining the specific contribution of each mineral. Yang et al. (2021) evaluated the spatio-temporal dynamics of C adsorption and release in aggregates of a transparent smectite clay, also relating enzymatic decomposition, through 4D imaging on a microfluidic chip. It should be noted that this study was not performed on a specific soil. Kirsten et al. (2021) determined the contribution of kaolinite, gibbsite, goethite, and hematite to C storage in soils under forests and agricultural lands, evaluating only the variation of C and clay mineral concentrations, but keeping the mineral types invariant.

Mapping the C sequestration potential that clay minerals composing the clay fraction has is important for the understanding of their dynamics and soil management (Padarian et al., 2022), remembering that SOC contents and their stable forms vary in relation to the amount and type of mineral (Yang et al., 2021; Kirsten et al., 2021; Prout et al., 2021; Dos Reis et al., 2014). However, in general, soil mapping has limitations in terms of spatial delimitation (Teng et al., 2018), and depends on the estimation and description of their properties. These conventional methods require complex laboratory chemical analysis and expertise for their description, incurring more time and costs (Shi et al., 2015), and are generally developed with limited information on quantity, volume and spatial coverage (Soil Survey Staff, 2017).

Information systems and remote sensing techniques have facilitated the acquisition of spatial information, specifically by digital soil mapping (DSM) approaches, which combine point soil data with statistically correlated auxiliary data (covariates) (McBratney et al., 2003), mainly through machine learning techniques (Padarian et al., 2019). Additionally, reflectance spectroscopy is a technology that has improved DSM (Teng et al., 2018), as the evaluation of infrared spectral curves provides useful indicators to map, classify and monitor different soil properties (Di Iorio et al., 2019), as is the case of mineral quantification (Mendes et al., 2021).

This study has the objective of analyzing the individual contribution of each mineral that composes the clay fraction in the sequestration potential of new SOC, through the quantification, modeling and mapping of this potential in different pedogenetic soils of Brazil, using remote sensing products and the equation of Feller and Beare (1997) to obtain the theoretical SOC sequestration potential. Considering the importance of the minerals that

compose the clay fraction and that this fraction in tropical soils is dominated by kaolinite, gibbsite, hematite and goethite (Kämpf and Curi 2003, Schaefer et al. 2008) and the high sorption power of Fe and Al oxides for organic molecules, it is expected that the SOC sequestration potential is directly related to the concentration of these minerals.

2.2.Methodology

2.2.1. Study area

The study area is in Piracicaba region, São Paulo State, Brazil, with approximately 2.598 km² (Figure 1). The climate of the region, according to the Köppen system, is classified as subtropical Cwa, with a dry winter and a rainy summer, with an average annual temperature ranging from 20 to 22.5 ° C and annual rainfall between 1200 and 1400 mm (Alvares et al., 2013). In relation to the topography, undulating highlands and rolling hills with altitudes ranging from 450 to 950 m are characteristic. Agricultural land uses such as sugarcane and pasture are dominant under no-till and conventional tillage management systems. The main soil types are Cambisol, Gleysol, Ferralsol, Nitosol, Lixisol, Leptsol, Arenosol and Planosol, according to the World Reference Base (IUSS Working Group WRB, 2015). Geologically, there are diverse parent materials, such as siltstones, tillites, varvites, conglomerates, sandstones, limestones, siltstones, flint, dolomite, siltite, pyrombetuminosite, schists, diabase, and basalt (Bonfatti et al., 2020).

The soil observations used are from the Brazilian Soil Spectral Library (BSSL) (Demattê et al., 2019). A total of 2354 observations from 0 to 20, 40 to 60, and 80 to 100 cm depths were used (Figure 1). The soil organic carbon (SOC) content and particle size were analyzed by the Walkey-Black method (wet digestion) (Walkey and Black, 1934) and the hydrometer method (Bouyoucus and John, 1962), respectively.

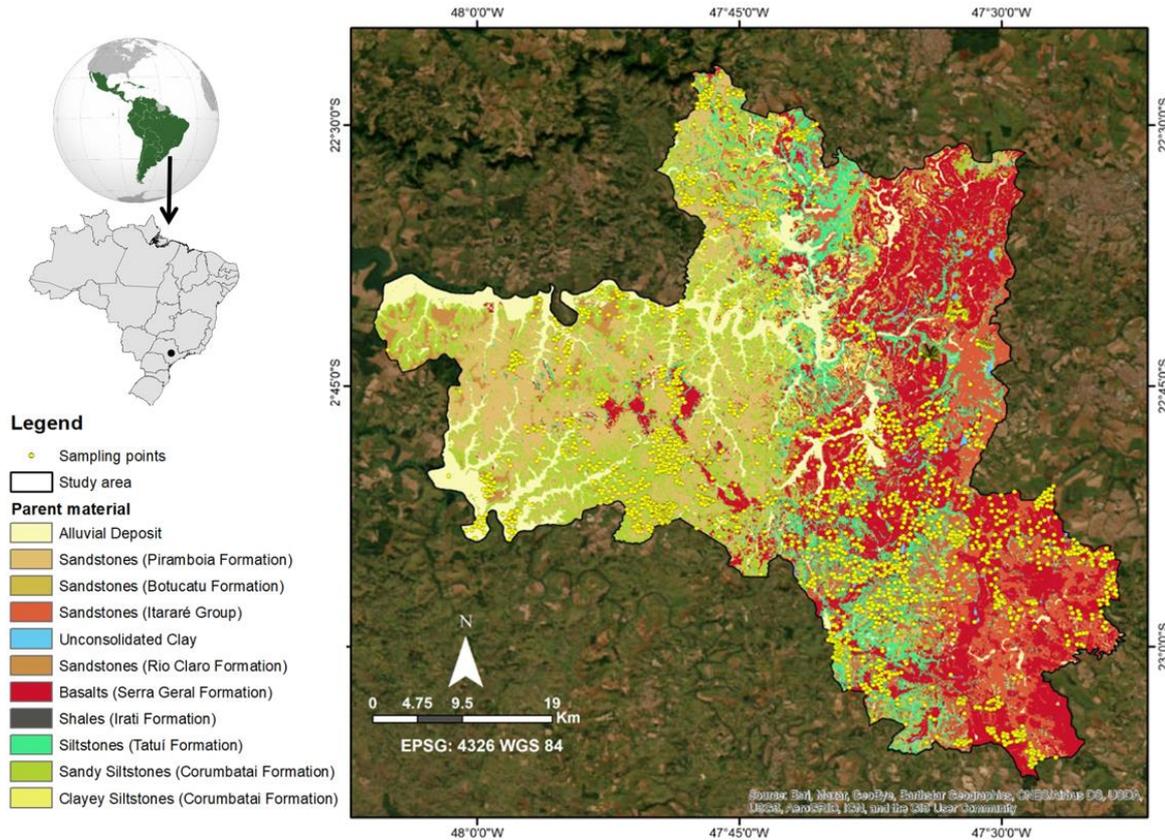


Figure 1. Location of the study area (Piracicaba region, São Paulo state). The geology map is from Bonfatti et al. (2020).

2.2.2. Point modeling of soil carbon sequestration potential

2.2.2.1. Carbon saturation potential

Hassink (1997) proposed the following equation to determine the C saturation potential of fine soil particles (<20 μm /silt and clay fractions):

$$\text{C}_{\text{sat-pot}} = 4.09 + 0.37 * \text{Particles} \leq 20 \mu\text{m} (\%)$$

where $\text{C}_{\text{sat-pot}}$ corresponds to the potential C saturation (mg g^{-1}), referred to as the theoretical maximum SOC that is stabilized in fine particles and allows estimating the SOC storage potential (Fujisaki et al., 2018). However, considering the study area, the modified equation by Feller and Beare (1997) for tropical soils was used, as it included samples of Brazilian clay soils with iron and aluminum oxyhydroxides, some located in our study area (Figure 2):

$$\text{C}_{\text{sat-pot}} = 3.2 + 0.29 * \text{Particles} \leq 20 \mu\text{m} (\%) \quad (r = 0.95, p < 0.001)$$

In this study, the value of the percentage of fine soil particles (particles $\leq 20 \mu\text{m}$ (%)) was replaced by the percentage of clay. The silt fraction, both coarse (20-53 μm) and fine fraction (2-20 μm), was not used in this study, as it was reported that the silt fraction of highly

weathered soils with high contents of kaolinite and iron and aluminum oxides store a small amount of carbon, representing about 4.8% of the total C content (Rodrigues et al., 2022). Figure 2 shows the methodological sequence of the point modeling to determine the C sequestration potential and to generate an equation to predict this potential based on spectral information associated with the minerals of the clay fraction.

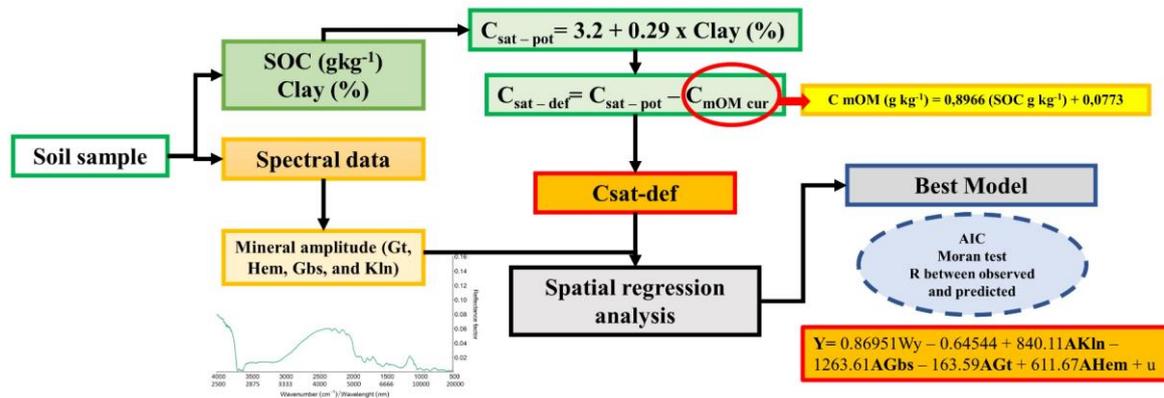


Figure 2. Methodological scheme of the point modeling of the carbon sequestration potential of minerals that compose the clay fraction. sat-pot: potential C saturation, sat-def: C saturation deficit, CmOM cur: actual concentration of C in mineral-associated organic matter, Gt: goethite, Hem: hematite, Gbs: gibbsite, Kln: kaolinite, A: relative abundance of soil minerals.

The determination of the C saturation deficit requires the actual concentration of C in the fine soil particles (<20 μm). For this, 35 representative soil samples were selected from the study area, based on the conditioned Latin hypercube sampling method, which corresponds to a stratified random sampling procedure where the selected samples follow multivariate characteristics according to the indicated covariables (Yang et al., 2020). Here, the Soil Synthetic Image (SYSI), soil type and variability in clay and C content were considered as covariates.

In these representative samples, fractionation was performed to quantify the C in particulate organic matter and the C in organic matter associated with the mineral fraction (mOM) (Cotrufo et al., 2019) following the methodology described by Jindaluang et al. (2013). The soil was dispersed using 5% sodium hexamethasphate solution and considering the low contribution of the coarse and fine silt fraction to the total C storage described by Rodrigues et al. (2022) the separation of sand and clay/silt fraction was performed by 53 μm sieving. Subsequently, a linear regression model based on the total SOC content and mOM

was developed to predict the C content of mOM in the remaining samples, with an R² fit of 0.98 (Figure 3). Therefore, the C saturation deficit, corresponds to the expression:

$$C_{\text{sat-def}} = C_{\text{satpot}} - (0.8966 \times \text{SOC total} + 0.0773)$$

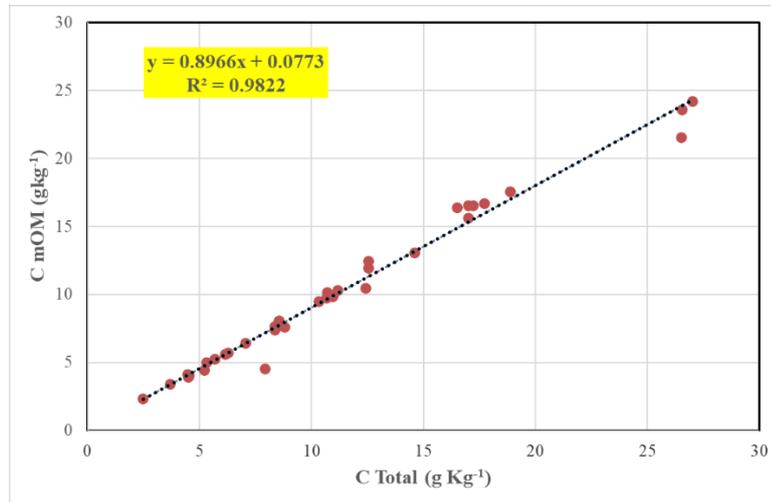


Figure 3. Simple linear regression between total soil C (x-axes) and C in organic matter associated with the mineral fraction (mOM) (y-axes).

Subsequently, the C reserve or stock of this difference is calculated from this difference, using the Benites et al. (2007) equations:

$$\text{SOC-stock} = (\text{SOC} \times D \times \text{BD}) \times 10$$

Where: SOC-stock = Soil Organic Carbon Stock (g m⁻²), SOC = Soil Organic Carbon content (g kg⁻¹), D = soil thickness (cm), BD = bulk density (g cm⁻³). The bulk density is calculated from:

$$\text{BD (g cm}^{-3}\text{)} = 1.5688 - 0.0005 \times \text{clay (g kg}^{-1}\text{)} - 0.0090 \times \text{SOC (g kg}^{-1}\text{)}$$

2.2.2.2. Spatial regression analysis

As pointed out by Marschner et al. (2008), predictions of SOC storage potential require an understanding of the processes related to SOC sequestration and release. Several authors (Hassink, 1997; Yang et al., 2021; Kirsten et al., 2021), highlighted the importance of minerals associated with the clay fraction in C sequestration because they involve variable amounts of minerals that have an affinity with organic molecules, according to the interaction on their surfaces that favors the formation of organic-mineral complexes (Sollins et al. 1996, Von Lützwow et al. 2006), reflecting in higher stability of SOC (De Mastro et al., 2020). Thus, this study focused on specifically demonstrating the response of variability in mineralogy on

SOC sequestration potential; therefore, regression models were built to predict the theoretical potential SOC saturation deficit (Sat-def), potential C saturation (Sat-pot), and C in mineral-associated organic matter (CmOM), each as a function of the mineral contents that make up the clay fraction. According to Okunlola et al. (2021), the presence and quantity of a specific mineral content depend on the spatial location. Therefore, it was necessary to perform a regression analysis considering the geographic coordinates of the samples and the spatial dependence of soil variability (Webster and Oliver 2007), so a weight matrix associated to all neighbors was generated.

The spatial regression analysis considered as dependent variables the potential SOC saturation deficit or potential SOC sequestration and potential C saturation (Sat-pot), calculated from the equation of Feller and Beller (1997) and the C in mineral-associated organic matter (CmOM) determined in the laboratory, as explanatory variables were considered the relative abundance of soil minerals represented by their infrared spectral amplitudes (calculated from diffuse reflectance spectroscopy data (Vis-NIR-SWIR)). The mineral amplitudes correspond to the difference between maxima and minima of the Savitzky-Golay second derivative curves obtained from the Kubelka-Munk absorption curves of the original spectra. These amplitudes were obtained from the study of Mendes et al. (2020), in which the bands associated with goethite (Gt, 422/450nm), hematite (Hem, 535/575 nm), gibbsite (Gbs, 2265/2285 nm) and kaolinite (Kln, 1415/2205 nm) are defined. A data set of 1248 samples was taken for the 0 to 20 cm depth and 833 for the 80 to 100 cm depth.

Spatial regression models, such as spatial autocorrelation models (SAC; referred to in the literature as SARAR), spatially lagged models (SLM) and spatial error model (SEM and SDEM) were fitted to predict the potential SOC sequestration spatially (Elhorts, 2014). SARAR is a double autoregressive model that includes the autoregressive component of the response and the residuals, allowing to explain the spatial dependence of the residuals.

The models are expressed in the following equation:

$$\begin{aligned} Y &= \lambda WY + \alpha \iota_n + X\beta + u; & |u| &< 1 \\ u &= \rho Wu + \varepsilon & |\rho| &< 1 \end{aligned}$$

Where, Y Where, Y represents the potential saturation deficit of SOC (Sat-def) or potential saturation of C (Sat-pot) or the CmOM, X represents the matrix of explanatory variables associated with the amplitude of minerals, W corresponds to the matrix of weights in relation to the distances of the nearest neighbor centroids in the polygons generated by tessellation of the soil sampling points, λ represents the spatial autoregressive coefficient, ρ the spatial autocorrelation coefficient, α corresponds to the intercept, β represents the

parameters linked to the explanatory variables, u is associated with the vector of residuals with spatial dependence and $\varepsilon \sim N(0, \sigma^2 I)$, where I is an identity matrix.

The choice of the model that best explains the statistical relationship of the experimental data was based on the lowest value of the Akaike information criterion (AIC) and on the fulfilment of the assumption of independence of the residuals based on the Moran Index Test (MIT), with the matrix of weights of all neighbors (Liu and Chen, 2021), where the p-value of the test must be greater than 0.05. In the case that more than one model satisfied the above assumptions, the highest correlation (r) between the observed values of response Y and the values estimated by the model was used as a criterion (Hoge et al., 2018).

Once the spatial regression model has been selected, it is important to evaluate and interpret the impact of the explanatory variables to determine the most important ones, however, in some spatial regression models, such as the autoregressive ones, it is not possible to perform this interpretation directly with the coefficients of the model as it is evaluated in classical regression models, therefore, according to Elhorst (2014), strategies are proposed for the estimation and interpretation of these coefficients, dividing them into direct, indirect and total impacts, which are obtained from the impacts function of the `spatialreg` library of R (Mendez, 2020). For the present study, the direct impacts are analyzed and the relative importance of each explanatory variable is calculated according to the total impacts.

2.2.3. Spatialization of carbon sequestration potential

The carbon sequestration potential of the clay fraction was spatialized by applying the equations described in the point modeling on the SOC and clay predicted maps for the different depths, as described in Figure 4.

For SOC and clay mapping, covariables (predictors) associated with relief and a Synthetic Soil Image (SYSI) were used. The relief attributes included elevation, slope, aspect, curvatures, valleys, hills, orientation, and topographic wetness index as described by Carvalho et al. (2019) and Sabetizade et al. (2021). The terrain variables were from a digital elevation model (DEM) of the Radar Topography Mission – SRTM (USGS, 2018), at 30 m spatial resolution. The SYSI in turn corresponds to a mosaic of the bare soil surfaces obtained from the Landsat images collection from 1984 to 2020. The SYSI images contain six bands in the Vis-NIR-SWIR spectral range (blue, green, red, NIR, SWIR1 and SWIR2) and were obtained by applying the Geospatial Soil Sensing System (GEOS3), developed by Demattê et al. (2018).

The Random Forest (RF) algorithm was chosen for spatial prediction, as it was reported as the best performing predictive algorithm in SOC mapping (Khaledian and Miller 2020, Zeraatpisheh et al. 2020, Lamichhane et al. 2019; Padarian et al., 2020). RF is a nonparametric model that performs classification and regression of sets through the construction of several decision trees in the training stage, where each tree is generated by a random vector (Breiman, 2001). The subdivisions within each tree are determined based on predictor variables chosen randomly from the set of variables (Coelho et al., 2020). Its strength is based on bootstrapping randomization of data and random input selection (Sothe et al., 2022) with replacement of the original data and internal validation with data not used in the bootstrap procedure (Khaledian and Miller 2020, Zeraatpisheh et al. 2020). The samples ($n = 2354$) were randomly divided into 70% and 30% for calibration and validation, respectively. The adjusted coefficient of determination (R^2) was used as a model evaluation metric.

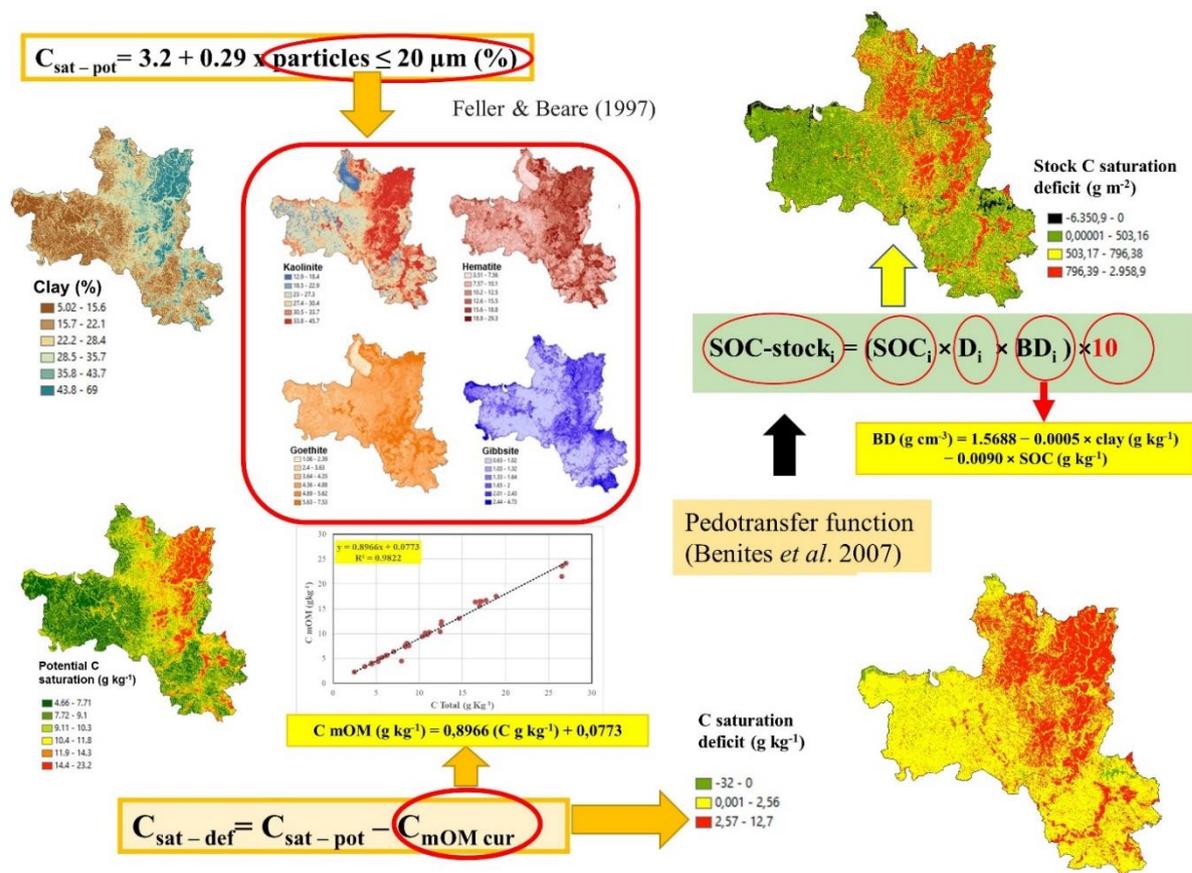


Figure 4. Methodological scheme for the spatialization of the carbon sequestration potential of the minerals that compose the clay fraction, based on map algebra.

For the spatialization of the C sequestration potential of each of the minerals that compose the clay fraction (goethite, hematite, gibbsite and kaolinite), we used the mineral maps elaborated by Mendes et al. (2021), which were obtained by digital soil mapping, using diffuse reflectance spectroscopy (Vis-NIR-SWIR) to estimate mineral abundance at specific locations and environmental covariates for spatialization. As for the clay fraction, the equations described in the point model were applied using map algebra, where "Particles $\leq 20 \mu\text{m}$ (%)" was replaced by the abundance map of each mineral at different depths, leaving fixed the predicted SOC maps for the different depths. According to Sothe et al. (2022) in the use of machine learning models for SOC prediction it is possible to use the same model keeping some covariates fixed to identify the influence of the variable of interest in the SOC prediction. Such spatialization will allow us to observe a spatial approximation of the individual contribution of the minerals that compose the clay fraction in the C sequestration potential, and together with the predictive model of this potential obtained from the spatial regression, will help to understand the dynamics of the potential of the mineralogy of the clay fraction to sequester new carbon.

2.3.Results

2.3.1. Point modeling of soil carbon sequestration potential

For the selection of the best fit models of the relationship between the response associated with the C sequestration potential or potential saturation deficit of SOC (Sat-def), potential saturation of C (Sat-pot) and CmOM with the explanatory variables related to mineral amplitude, the pure spatial autoregressive regression models (PAR), the spatial lag model (SLM), the spatial error (SEM), the spatial double autoregressive model (SARAR) and the spatial Durbin error (SDEM) were used were used, however the latter was excluded from Sat-pot and CmOM because a fit was not achieved (Tables 1 and 2).

Table 1. Fitted spatial regression models and its related statistics for potential saturation deficit of SOC (Sat-def). λ = autoregressive parameters, ρ = spatial autocorrelation coefficient, r = correlation coefficient, MIT= Moran Index test.

Model	λ	ρ	AIC	r	MIT	Explanatory variables
0-20 cm						
PAR	0.98 (2.22E ⁻¹⁶)		6848	0.51	6.41E ⁻¹³	Sat-def-a
SEM	0.97 (2.22E ⁻¹⁶)		6708	0.56	2.15E ⁻⁰⁸	AKln + AGbs + AGt + AHem
SLM		0.95 (2.22E ⁻¹⁶)	6720	0.55	1.53E ⁻¹⁰	AKln + AGbs + AGt + AHem
SLMA		0.97 (2.22E ⁻¹⁶)	6700	0.56	9.90E ⁻⁰⁶	AKln + AGbs + AGt + AHem
SARAR	0.87 (7.53E⁻¹³)	0.66 (7.18E⁻⁰⁴)	6692	0.56	0.39407	AKln + AGbs + AGt + AHem
SDEM	0.97 (2.22E ⁻¹⁶)		6701	0.56	2.98E ⁻⁰⁵	AKln + AGbs + AGt + AHem
80-100 cm						
PAR	0.99 (2.22E ⁻¹⁶)		4520	0.62	2.00E ⁻¹⁵	Sat-def-c
SEM	0.98 (2.22E ⁻¹⁶)		4455	0.65	1.98E ⁻¹⁰	AKln + AGbs + AGt + AHem
SLM		0.98 (2.22E ⁻¹⁶)	4445	0.65	1.65E ⁻⁰⁶	AKln + AGbs + AHem
SLMA		0.96 (2.22E ⁻¹⁶)	4443	0.65	3.98E ⁻⁰⁵	AKln + AGbs + AGt + AHem
SARAR	0.82 (4.88E⁻⁰²)	0.85 (1.21E⁻⁰²)	4429	0.66	0.37	AKln + AHem
SDEM	0.98 (2.22E ⁻¹⁶)		4449	0.65	1.15E ⁻⁰⁷	AKln + AGbs + AGt + AHem

Sat-def = potential carbon saturation deficit or carbon sequestration potential. A= amplitude of the different minerals AKln (kaolinite), AGt (goethite), AHem (hematite), AGbs (gibbsite).

Table 2. Fitted spatial regression models and its related statistics for C in organic matter associated with the mineral fraction (CmOM) and Potential C saturation (Sat-pot). λ = autoregressive parameters, ρ = spatial autocorrelation coefficient, r = correlation coefficient, MIT= Moran Index test.

Model	λ	ρ	AIC	r	MIT	Explanatory variables
C in organic matter associated with the mineral fraction (CmOM)						
0-20 cm						
PAR	0.98 (2.22E ⁻¹⁶)		6823.3	0.56	0	CmOM-a
SEM	0.99 (2.22E ⁻¹⁶)		6700.3	0.59	0	AKln + AGbs + AGt + AHem

SLM	0.98 (2.22E ⁻¹⁶)	6695.5	0.59	2.87E ⁻¹⁵	AKln + AGbs + AGt + AHem
SLMA	0.98 (2.22E ⁻¹⁶)	6693.1	0.59	2.31E ⁻¹¹	AKln + AGbs + AGt + AHem
SARAR	0.88 (0.015)	6659.1	0.60	0.1723	AKln + AGbs + AGt + AHem
80-100 cm					
PAR	0.99 (2.22E ⁻¹⁶)	3895.2	0.64	0	CmOM -c
SEM	0.99 (2.22E ⁻¹⁶)	3783.8	0.68	0	AGbs + AGt + AHem
SLM	0.98 (2.22E ⁻¹⁶)	3783.7	0.67	0	AGbs + AGt + AHem
SLMA	0.98 (2.22E ⁻¹⁶)	3778.9	0.68	0	AGbs + AGt + AHem
SARAR	0.94 (1.59E-06)	3734.3	0.70	5.05E⁻⁰²	AGbs + AGt + AHem
Potential carbon saturation (Sat-pot)					
0-20 cm					
PAR	0.99 (2.22E ⁻¹⁶)	7053.1	0.65	0	Sat-pot-a
SEM	0.99 (2.22E ⁻¹⁶)	6620	0.75	0	AKln + AHem
SLM	0.89 (2.22E ⁻¹⁶)	6623.9	0.74	1.15E ⁻¹⁴	AKln + AHem
SLMA	0.98 (2.22E ⁻¹⁶)	6610.6	0.43	4.68E ⁻¹¹	AKln + AHem
SARAR	0.89 (2.22E⁻¹⁶)	6586.9	0.76	0.28	AKln + AHem
80-100 cm					
PAR	0.99 (2.22E ⁻¹⁶)	4711.8	0.75	0	Sat-pot-c
SEM	0.99 (2.22E ⁻¹⁶)	4517.5	0.79	0	AKln + AGbs + AHem
SLM	0.98 (2.22E ⁻¹⁶)	4501.1	0.79	1.57E ⁻⁰⁸	AKln + AGbs + AHem
SLMA	0.99 (2.22E ⁻¹⁶)	4534.2	0.79	0	AKln + AGbs + AHem
SARAR	0.86 (8.54E-05)	4481.5	0.66	0.37	AKln + AGbs + AHem

Where A= amplitude of the different minerals AKln (kaolinite), AGt (goethite), AHem (hematite), AGbs (gibbsite).

Based on the MIT evaluation (Tables 1 and 2), only the SARAR model, for each of the dependent variables, satisfied the criteria. Based on AIC and r, the SARAR model also

performed best in all three models for the depth of 0-20 cm (Sat-def: AIC = 6692 y r= 0,56; CmOM: AIC=6659 y r=0.60; Sat-pot: AIC=6587 y r=0.76) and 80-100 cm (Sat-def: AIC = 4429 y r= 0,66; CmOM: AIC=3734 y r=0.70; Sat-pot: AIC=4481 y r=0.66). For 0-20 cm, the SARAR model of carbon sequestration potential (Sat-def) includes all minerals, whereas, for the latter depth, it includes only kaolinite and hematite. CmOM is also explained by all minerals at the first depth, and the importance of kaolinite was lost in the last layer. On the contrary, Sat-pot at the first depth was only explained by kaolinite and hematite, and at the last depth kaolinite, gibbsite and hematite were considered.

Table 3. Parameters of the SARAR spatial regression model for depths 0 to 20 cm and 8 to 100 cm for potential saturation deficit of SOC (Sat-def)

Coefficients	Estimate	Asymptotic Std. Error	z value	Pr(> z)	Impact Direct	Impact p-value
0-20 cm						
(Intercept)	-0.64	0.89	-0.72	0.47		
AKln	840	142	5.92	3.26E⁻⁰⁹	845	1.90E⁻⁰⁹
AGbs	-1264	376	-3.35	0.000786	-1272	0.000563
AGt	-164	40	-4.08	4.48E ⁻⁰⁵	-165	2.94E ⁻⁰⁵
AHem	612	90	6.80	1.02E⁻¹¹	616	6.10E⁻¹³
80-100 cm						
(Intercept)	-1.66	1.28	-1.30	0.19		
AKln	600	101	5.94	2.92E⁻⁰⁹	611	2.72E⁻⁰⁸
AHem	127	66.9	1.90	0.05	130	3.35E⁻⁰²

Where A= amplitude of the different minerals AKln (kaolinite), AGt (goethite), AHem (hematite), AGbs (gibbsite). Pr (>|z|) is related to the significance of each variable in the model, with lower values highlighting greater importance.

Table 4. Parameters of the SARAR spatial regression model for depths 0 to 20 cm and 8 to 100 cm for C in organic matter associated with the mineral fraction (mOM) and Potential carbon saturation (Sat-pot)

Coefficients	Estimate	Asymptotic Std. Error	z value	Pr(> z)	Impact Direct	Impact p-value
C in organic matter associated with the mineral fraction (mOM)						
0-20 cm						
(Intercept)	-0.81	1.41	-0.57	0.57		
AKln	-278.9	137.1	-2.03	0.041	-283.7	0.061
AGbs	1283.4	364.8	3.52	0.0004	1305.4	0.0007
AGt	164.7	38.4	4.29	1.75E⁻⁰⁵	167.6	8.94E⁻⁰⁶
AHem	183.2	85.9	2.13	0.032	186.3	0.028
80-100 cm						
(Intercept)	-1.71	1.78	-0.95	0.34		
AGbs	720.1	234.1	3.07	0.002	739.8	0.003
AGt	66.6	22.9	2.91	0.003	68.4	0.009
AHem	188.4	57.8	3.26	0.001	193.6	0.0009
Potential carbon saturation (mg g⁻¹) (Sat-pot)						
0-20 cm						
(Intercept)	-0.46	1.3	-0.35	0.72		
AKln	616.6	121.6	5.07	3.95E⁻⁰⁷	621.6	4.95E⁻⁰⁷
AHem	824.2	65.9	12.5	2.22E⁻¹⁶	830.9	2.22E⁻¹⁶
80-100 cm						
(Intercept)	-3.03	1.15	-2.61	0.009		
AKln	716.4	110.8	6.47	9.98E⁻¹¹	729.5	0.0002
AGbs	886.4	385.4	2.30	0.021	902.7	0.035
AHem	346.4	78.4	4.42	1.004E⁻⁰⁵	352.8	6.36E⁻⁰⁵

Satdef = potential carbon saturation deficit or carbon sequestration potential. A= amplitude of the different minerals AKln (kaolinite), AGt (goethite), AHem (hematite), AGbs (gibbsite). Pr (>|z|) is related to the significance of each variable in the model, with lower values highlighting greater importance.

The models for each depth described in the tables 2 and 3 can be expressed from the matrix point of view as shown in the following equations:

- 1) For depth from 0 to 20cm:

$$\begin{aligned}
 C_{sat} - def &= 0.87Wy - 0.64 + 840AKln - 1264AGbs - 164AGt + 612AHem + u; \\
 u &= 0.66Wu + \varepsilon \\
 C_{mOM} &= 0.88Wy - 0.81 - 278.9AKln + 1283.4AGbs + 164.7AGt + 183.2AHem + u; \\
 u &= 0.88Wu + \varepsilon \\
 C_{SatPot} &= 0.89Wy - 0.46 + 616.6AKln + 824.2AHem + u; \\
 u &= 0.71Wu + \varepsilon
 \end{aligned}$$

- 2) For depth from 80 to 100cm:

$$\begin{aligned}
 C_{sat} - def &= 0.82Wy - 1.66 + 600AKln + 127AHem + u; \\
 u &= 0.85Wu + \varepsilon
 \end{aligned}$$

$$\begin{aligned}
 C_{mOM} &= 0.94Wy - 1.71 + 720.1AGbs + 66.6AGt + 188.4AHem + u; \\
 u &= 0.92Wu + \varepsilon \\
 C_{SatPot} &= 0.86Wy - 3.03 + 716.4AKln + 886.4AGbs + 346.4AHem + u; \\
 u &= 0.86Wu + \varepsilon
 \end{aligned}$$

Where, Sat-def = potential carbon saturation deficit or carbon sequestration potential, CmOM= C in organic matter associated with the mineral fraction, Sat-pot= Potential carbon saturation, A= amplitude of the different minerals AKln (kaolinite), AGt (goethite), AHem (hematite), AGbs (gibbsite), W corresponds to the matrix of weights, u is associated with the vector of residuals with spatial dependence and $\sim N(0, \sigma^2 I)$, where I is an identity matrix.

The spatial modeling results show that the carbon sequestration potential (sat-def) for 0-20 cm depth could be explained by the relative contents of kaolinite, gibbsite, goethite and hematite (Table 3). Where kaolinite and hematite had the largest direct positive impact. On the contrary, a direct but negative impact was observed for goethite and gibbsite, which could indicate that an increase in the concentration of these minerals reduces the C sequestration potential of the soil, however these minerals have the highest affinity for organic molecules (Kaiser and Zech, 2000; Dos Reis *et al.*, 2014), so they tend to saturate first compared to kaolinite and hematite, and stabilize more efficiently the sequestered C (Kalbitz *et al.*, 2005; Dos Reis *et al.*, 2014), in that sense, such negative impacts could then be translated as the higher concentration of these minerals, the greater stabilization of organic molecules may occur, that is, higher current COS content and lower potential to sequester new carbon. This explains the results of the model for potential C saturation (sat-pot, Table 4), corresponding to the theoretical maximum of SOC, which in the 0-20 cm depth was only explained by kaolinite and hematite, indicating that it is these minerals that have the potential to sequester new C.

It is also important to highlight that the CmOM model for the 0 to 20 cm depth (Table 4) shows greater importance in gibbsite and goethite (Figure 5), due to the potential for stabilization of organic molecules presented by these minerals, which corroborates that these are the ones who contribute most to the current C, and contrary to the Sat-def model (Table 3), the negative impacts were presented in kaolinite, since, as mentioned above, the C associated with this mineral is related to the potential for sequestering new C. On the other hand, the C sequestration potential (sat-def) for the 80 to 100 cm depth (Table 3) was mainly explained by the contents of kaolinite and hematite, with the greatest impact of kaolinite (Figure 5). CmOM at this depth (Table 4) was mainly explained by gibbsite, goethite and hematite, with greater impact of gibbsite and hematite (Figure 5). On the contrary, Sat-pot was explained by kaolinite, gibbsite and hematite with higher impact of kaolinite and hematite. It is important to highlight the importance of hematite in the C sequestration and stabilization cycle at the two

depths, since in all three models it is a variable of high importance (Figure 5). According to Georgiou et al. (2022) increasing mineral-associated C is key to long-lasting C sequestration, and for the soils of the study region hematite responds to these additional spaces to sequester and stabilize new C along the soil profile.

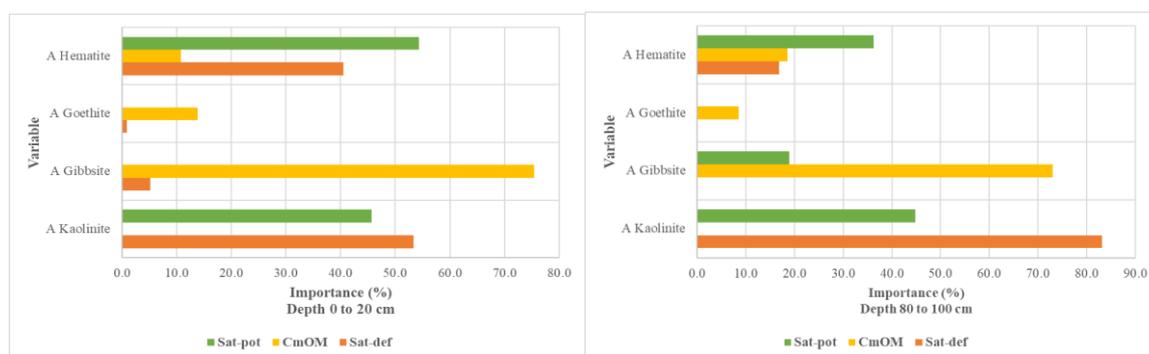


Figure 5. Importance of the explanatory variables of the SARAR spatial regression model for depths 0 to 20 cm and 8 to 100 cm for the potential carbon saturation deficit or carbon sequestration potential (sat-def), C in organic matter associated with the mineral fraction (CmOM) and Potential carbon saturation (sat-pot)

In general, the participation of goethite and gibbsite in explaining the C sequestration potential (sat-def) was low in the 0 to 20 cm depth and null in the 80 to 100 cm depth (Figure 5), with greater importance of kaolinite compared to hematite, whose difference was not so marked for the 0 to 20 cm depth.

2.3.2. Carbon sequestration potential mapping

Carbon and clay maps were obtained for the different depths using DSM with R^2 of 0.6 and 0.7, respectively. Areas with higher clay content had a higher carbon sequestration potential (areas in red), that is, the minerals that compose this fraction had the potential to retain more carbon, and these were related to agricultural areas (Figure 6 and 13, Table 5). On the contrary, areas with more than 15 years under the same land use, such as pastures and forests (Figure 6 and 13, Table 5), had less potential for additional carbon sequestration.

When evaluating the individual contribution of each mineral (Figure 7), it was observed that the zones with the highest C sequestration potential in Figure 6 corresponded to areas that were saturated, highlighting the importance of the individual analysis of the minerals that make up the clay fraction, because evidently not all of them have the potential to sequester

new carbon, being kaolinite and hematite those that still have space to store new carbon in the 0 to 20 cm depth, which was consistent with the result of the spatial regression models in Tables 3 and 4 related to Sat-def and Sat-pot. Spatially, a high C sequestration potential was evidenced for kaolinite in the areas related to pasture and agricultural mosaics with more than 15 years (Figure 13, Table 5), and the agricultural zone presented a low to null C sequestration potential (Figure 7, Figure 13). For gibbsite and goethite, zero sequestration potential was observed, indicating C saturation since the major contribution of these minerals translates into the current C associated with the mineral fraction (CmOM), confirming their low importance in the Sat-def spatial regression model (Figure 5). On the other hand, areas dominated by hematite had a medium to high sequestration potential. Areas under agriculture and agricultural mosaics with pastures and forests showed the highest potential for carbon sequestration by hematite.

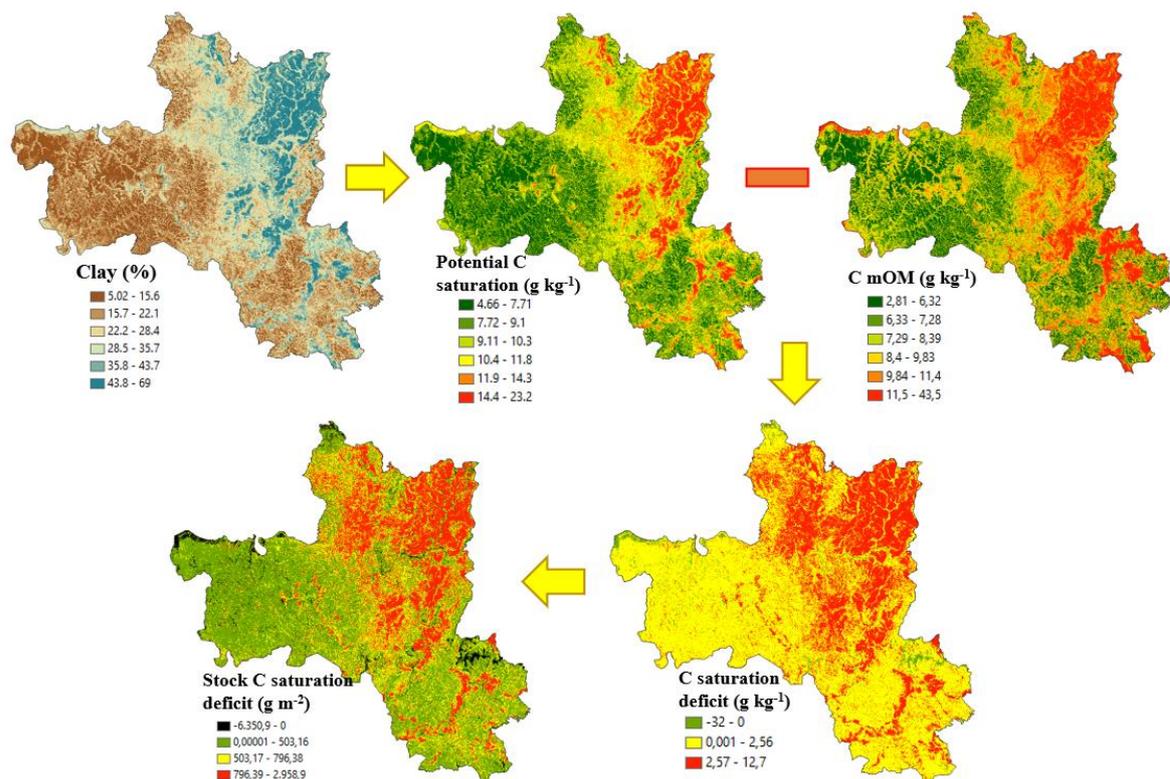


Figure 6. Carbon saturation potential, C in organic matter associated with the mineral fraction (mOM), saturation deficit, and deficit stock associated with clay contents at soil depth of 0 to 20 cm.

For the 40-60 cm and 80-100 cm depths, an increase in C sequestration potential was observed compared to the 0-20 cm depth (Figure 8, 10 and 12). Specifically, the increase in

potential in relation to the first depth was observed in the agricultural zones, from the north and southwest of the study area, with a considerable improvement in the zones that had mixed pasture and cropping (Figure 13, Table 5).

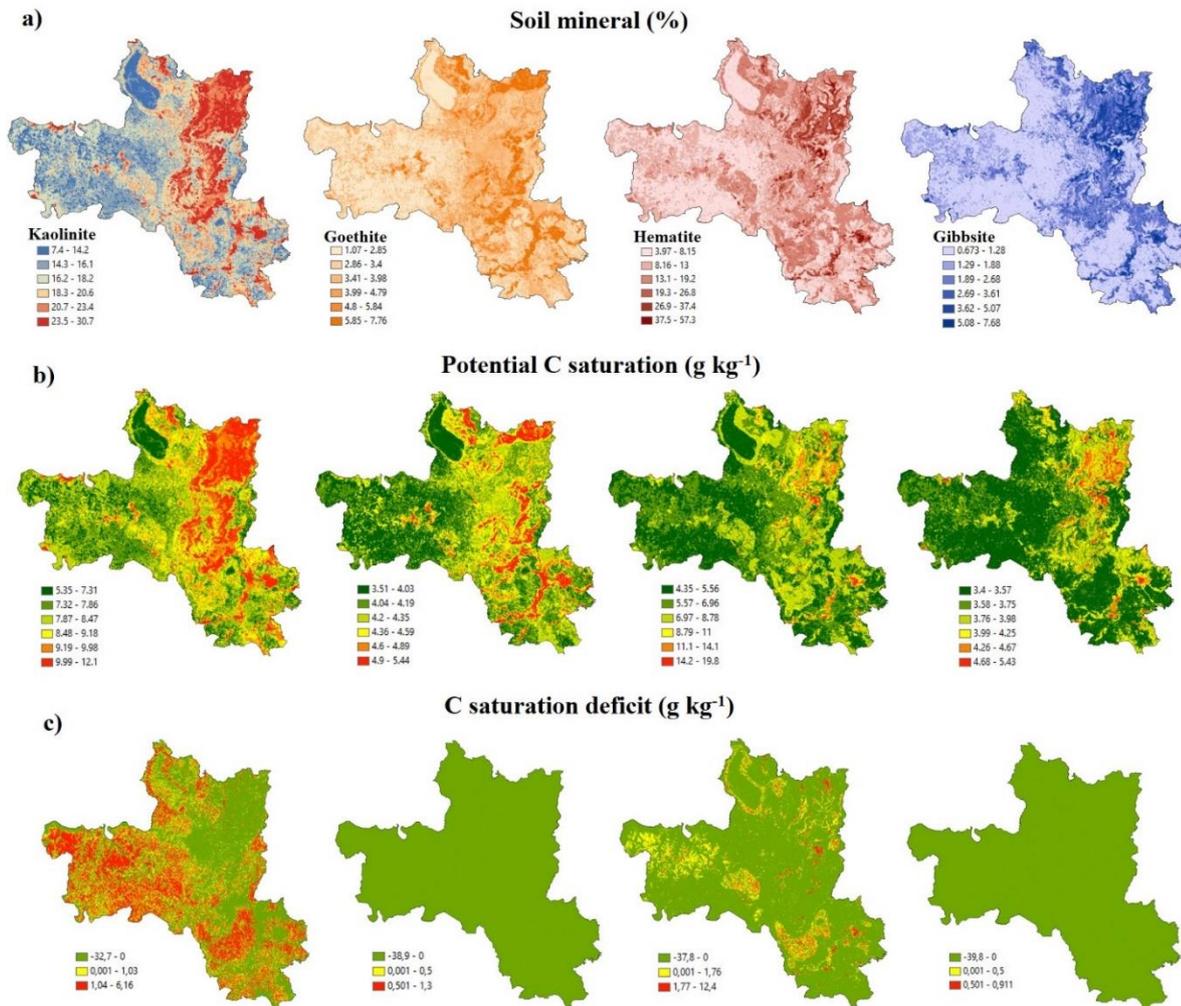


Figure 7. Percentage of kaolinite, goethite, hematite, and gibbsite minerals (a), carbon saturation potential (b) and saturation deficit (c) for 0 to 20 cm.

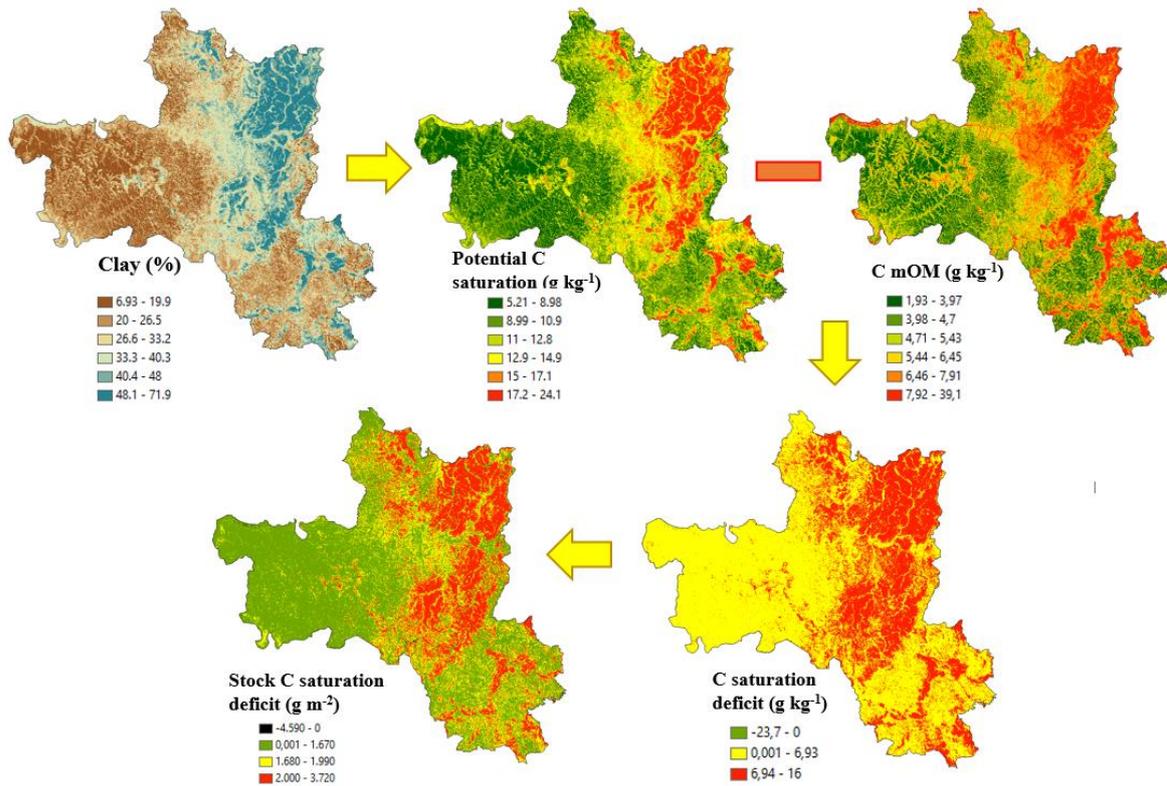


Figure 8. Carbon saturation potential, C in organic matter associated with the mineral fraction (mOM), saturation deficit and deficit stock associated with clay contents at soil depth of 40 to 60 cm.

Higher mineral contents were observed in the deeper layers (Figures 9 and 11), especially kaolinite. However, the contents of other minerals showed a reduction at 40-60 cm and a considerable increase at 80-100 cm depth, which was reflected in an increase in C sequestration potential. It should be noted that even with the reduction of iron and aluminum oxide minerals contents at depth 40-60 cm (Figure 9), a considerable improvement in the C sequestration potential of hematite and kaolinite were observed. A slight improvement in the potential of gibbsite and goethite was also observed, where gibbsite maintains a low potential in most of the area, with a slight improvement in the proximity of pasture, forest and cropping mosaics. The increase in C sequestration potential for goethite was observed in areas with crop and pasture mosaics. For hematite, low C sequestration potential was maintained in the northeastern part of the study area, corresponding to areas with more than 15 years in agriculture (Figure 13, Table 5).

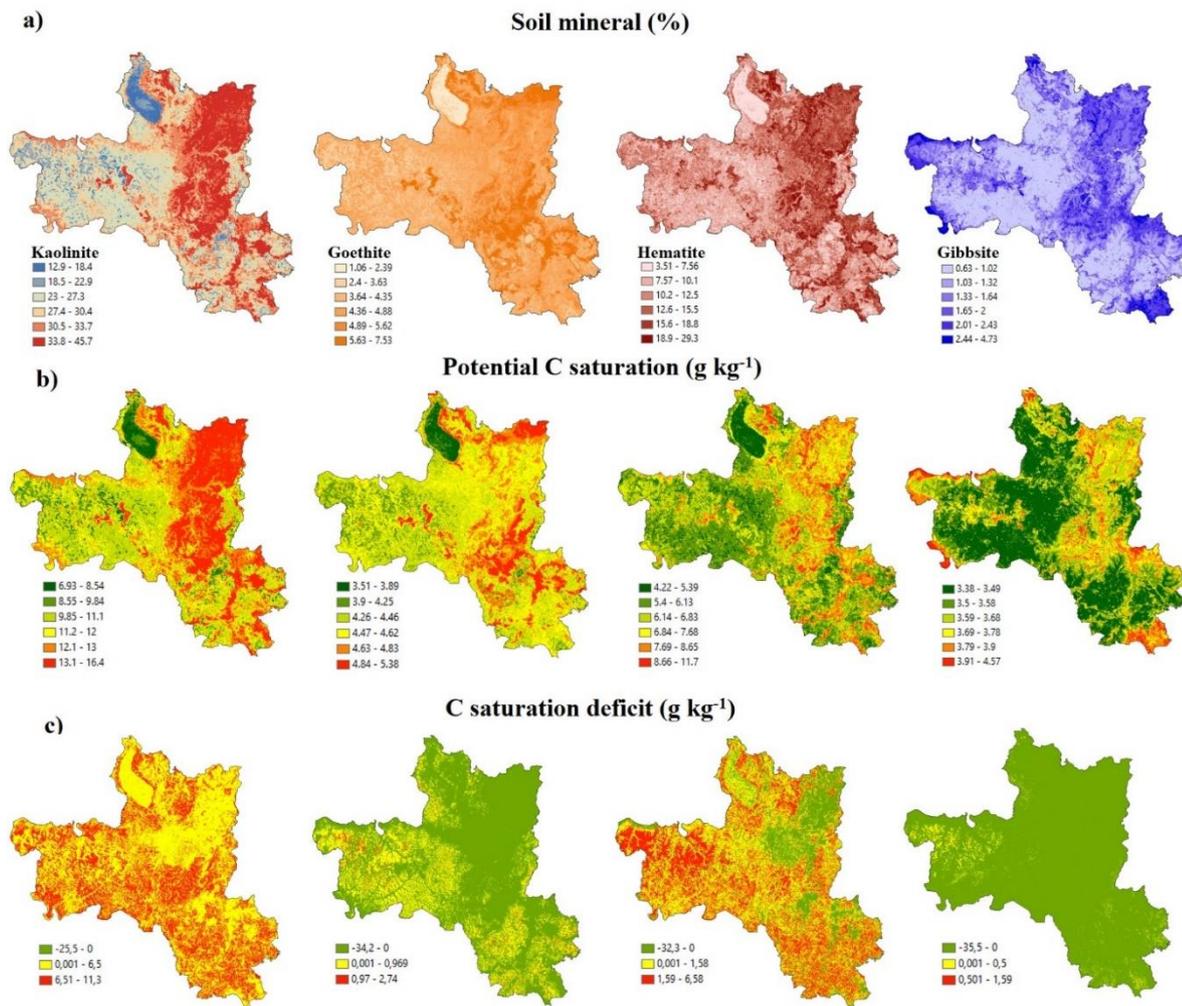


Figure 9. Percentage of kaolinite, goethite, hematite, and gibbsite minerals (a), C saturation potential (b) and saturation deficit (c) for 40 to 60 cm.

For 80-100 cm, the results of Sothe et al. (2022) were confirmed, showing that the kaolinite and the iron and aluminum oxides were not fully saturated (Figures 10 and 11). C sequestration potential was higher, observing an increase in the potential for hematite in the areas under agriculture, with an increase in hematite potential observed in the areas under agriculture, which at depths 40-60 cm still showed low sequestration potential. Similarly, an increase in the sequestration potential for goethite and gibbsite were observed in the areas with pasture and cropping mosaics, maintaining a low potential in the areas with agricultural use for more than 15 years (Figure 13, Table 5). Statistically, the point modelling highlighted the importance of kaolinite and hematite in the carbon sequestration potential of this depth (Table 3), however, it did not consider this contribution of gibbsite and goethite.

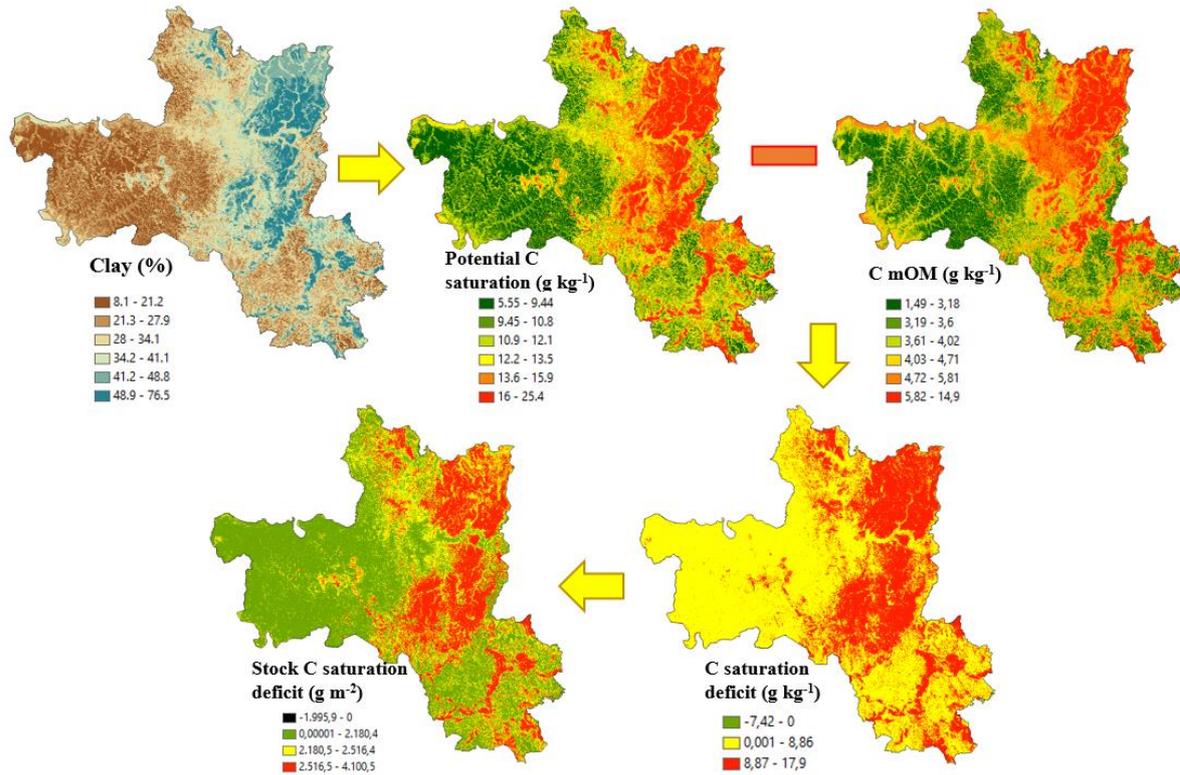


Figure 10. Carbon saturation potential, C in organic matter associated with the mineral fraction (mOM), saturation deficit and deficit stock associated with clay contents at a soil depth of 80 to 100 cm.

In general, it was observed that as the depth increases, there is a greater potential for sequestration of new C (Figure 12), because there is less current C content in the mineral fraction, as Georgiou et al. (2022) mention, the greater the depth, the greater the subsaturation of C associated with minerals, therefore it is possible to consider that there is a potential C pool that could be exploited with the inclusion of shrub and tree crops whose root system reaches deeper into the soil.

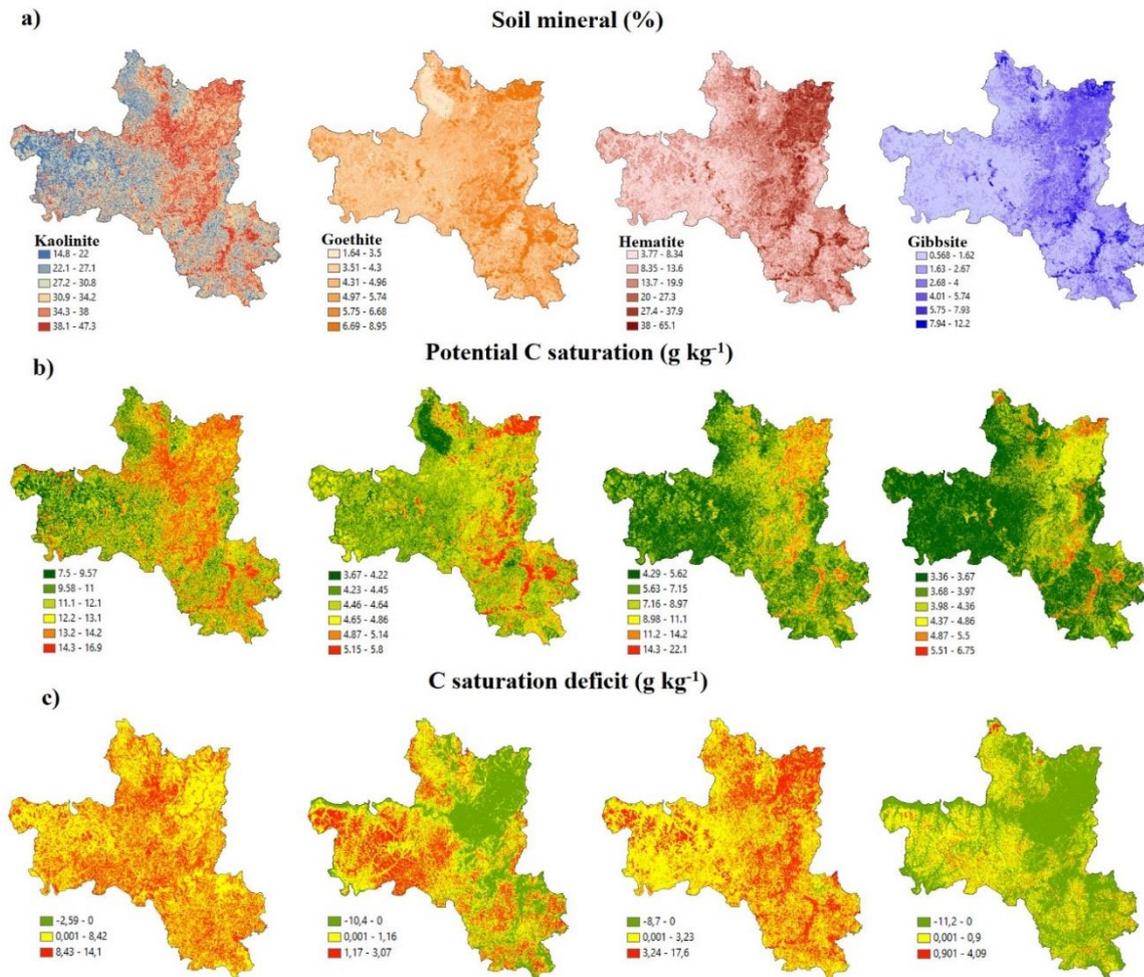


Figure 11. Percentage of kaolinite, goethite, hematite, and gibbsite minerals (a), carbon saturation potential (b) and saturation deficit (c) for 80 to 100 cm.

2.4. Discussion

According to Boddey et al. (2010), the analysis of C storage potential requires the evaluation of deeper soil layers because studies from 0 to 100 cm depth reveal 59% more storage in relation to a study from 0 to 30 cm. That is, the inclusion of depth allows adequate prediction of SOC concentration (Sothe et al., 2022), since at shallower depths the mineral particles are more saturated with SOC. Therefore, depth allows for improved analysis of SOC storage potential (Hobley et al., 2015). This was confirmed in the present study, where with increasing depth a higher C sequestration potential was observed (Figure 12), due to lower C saturation in clay fraction minerals such as kaolinite, hematite, goethite and gibbsite (Figures 7, 9 and 11) and to the increase in the content of these minerals with depth, as they are more commonly found in highly weathered soils, with greater homogeneity in depth (Berg and

Oliveira, 2000) and according to Georgiou et al. (2022) the maximum C content associated with minerals depends on the amount and type of mineral.

Ingram and Fernandes (2001), Weil and Brady (2016), indicated the importance of clay mineralogy on the potential of a soil to store organic C, especially in the deeper layers (Gray et al. 2015, Wiesmeier et al., 2011). In the present study, the spatial regression models showed a clear difference in the minerals contributing to C sequestration potential at depth 0 to 20 and 80 to 100 cm (Table 2), with low contribution of gibbsite and goethite at the deeper depth. However, in Figures 9 and 11, the contribution of goethite and gibbsite in this overall contribution of new carbon sequestration that is not seen in the statistical model was observed.

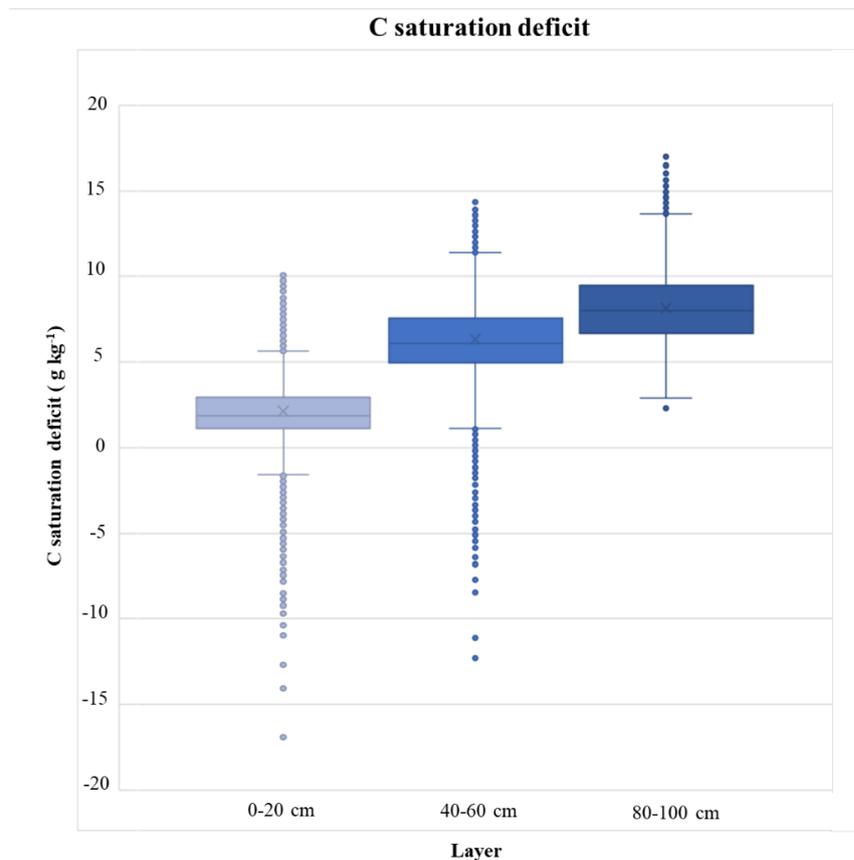


Figure 12. Variation of the content of the potential C saturation deficit or C sequestration potential, at different soil depths

Interpreting the individual contribution of the C sequestration potential of the clay fraction minerals is difficult due to their coexistence (Kirsten et al. 2021). Georgiou et al. (2022) points out the importance of generating mathematical models that allow inferring the C associated with the mineral fraction, however, their study was based on the limit line analysis

where the determination of the C saturation potential is based on the highest C stocks and C contents in soils with presence of 2:1 clays and poorly crystalline minerals, which theoretically have higher capacity to stabilize C. For the present study this theoretical maximum limit was calculated as the potential C saturation based on the equation of Feller and Beare (1997) and spatial regression models were used to explain this potential C saturation (Sat-pot), the C associated with the mineral fraction (CmOM) and the potential for sequestration of new C (Sat-def) as a function of the relative abundance of the minerals that compose the clay fraction, finding the best fits with the SARAR double autoregressive spatial regression model, highlighting that the main contributing minerals in the C sequestration potential correspond to kaolinite and hematite (Table 3), with a low contribution of goethite and gibbsite in the 0 to 20 cm depth (Figure 5). These minerals have direct impacts that indicate that a reduction in their concentration could reduce the C sequestration potential of the study area, since they are the minerals that contribute most to C stabilization and to the CmOM content.

According to Schaefer et al. (2008), the clay fraction of Brazilian soils is dominated by kaolinite and low crystallinity Fe and Al oxides, typically corresponding to gibbsite, hematite, goethite and maghemite (Kämpf and Curi, 2003). This low crystallinity of these iron oxides in Brazilian soils translates into more effective OM stability than crystalline Fe oxides or oxyhydroxides (Schaefer et al., 2008), because they exhibit electrostatic attractions and ionic bonds between the hydroxyl groups of the oxides and the carboxyl or hydroxyl groups of the OM (Duiker et al. 2003, De Mastro et al. 2020). In our study area, the presence of parental material associated with basalt (Figure 1) allows locating ferruginous minerals of low degree of crystallinity according to Tombácz et al. (2004) and Ashton et al. (2016). Such localization coincides with the concentration of iron oxides with higher affinity for the MO (Ashton et al., 2016), which highlights that the reduction of goethite and gibbsite contents affects the stabilized MO, since as observed in the CmOM model, these minerals account for approximately 90% of the importance in explaining the current C of the mineral fraction (Figure 5).

According to Guzmán et al. (1994), goethite usually presents greater affinity for OM because it presents a double network of octahedra, where Fe^{+3} occupies only half of the spaces (Bigham et al., 2002). It gives a greater specific surface area for this oxide, compared to hematite, which presents an occupation of Fe^{+3} in 66% of the oxygens, reducing its specific surface area (Bigham et al. 2002) considerably. Kaiser and Guggenberger (2000) indicated that goethite and gibbsite have a high density of reactive sorptive sites, allowing more

effective organo-mineral interactions. Hematite has a denser structure and lower surface area compared to goethite, having lower reactivity of hydroxyl groups on its surface (Dos Reis et al., 2014). Therefore, C had affinity and preference for goethite (Figures 7 and 9). Similarly, it was found that gibbsite also had higher C saturation potential.

A decrease in pH increases the positive charges of iron oxides, increasing OC sorption (Ashton et al., 2016). However, the high presence of kaolinite could generate an increase in the negative charges neutralizing the positive charges of the oxides. In turn, it could lead to a reduction of stabilized OC (Kirsten et al., 2021), which is evidenced in the greater C retention potential, being clear the role played by pH (Ashton et al., 2016), as it influences the surface charge and therefore the adsorption capacity of organic compounds (Saidy et al., 2013). The pH control of the protonation and deprotonation of hydroxyl groups (Wang et al., 2020). For the study area, the more significant presence of kaolinite was related to a high hematite content, especially in the eastern region, where pH ranged from strongly acidic (5 to 5.5) to moderately acidic (5.6 to 6) (Figure 13), which could also explain the high potential for C sequestration in these areas.

Land use type also influences SOC content due to differences in vegetation and C input. Agricultural and highly degraded soils have considerable potential to store additional SOC (Wiesmeier et al. 2013, Georgiou et al. 2022), as a marked depletion of SOC stocks is observed (Paustian et al. 1997, Lal 2004, Smith 2004, Follett et al. 2001, Padarian et al. 2022). Sothe et al. (2022), reported a higher concentration of SOC in crops than in grazing land. Areas with more than 15 years of agriculture had both the lowest and the highest carbon sequestration potential (Table 5), that is, minerals such as goethite and gibbsite at depth 0 to 20 cm present a low potential to sequester new C in the study area and at greater depth the low potential is concentrated in these areas with traditional agricultural use. On the other hand, at shallower depths, minerals such as hematite and kaolinite had a higher sequestration potential, being higher for kaolinite in areas of pasture and cropping mosaics, and higher for hematite in areas with cropping mosaics of pasture and forest. With increasing depth, unlike goethite and gibbsite, kaolinite and hematite had high C sequestration potential in areas with traditional agricultural use. In agricultural areas, management practices that favor sequestration are related to the promotion of organic inputs, conservation/minimum tillage, conversion of cropland to pasture, introduction of perennials, proper management of cultivated peatlands, and organic farming (Sauerbeck 2001, Vleeshouwers and Verhagen 2002, Freibauer et al. 2004, Lal 2004, Johnson et al. 2007, Smith 2012). Rabbi et al. (2015)

and Ashton et al. (2016), reported that conversion of cropland to grassland could increase C sequestration, that coincides with the observed results.

Afforestation and pasture improvements could contribute with soil C storage increase (Zeraatpishe and Khormali 2012, Nave et al. 2013). It has also been reported when conversion from forest to managed pasture and from cropland to pasture occurred (Poeplau and Don, 2013). The areas under exclusively forest use presents low sequestration potential and the minerals such as goethite were highly saturated. On the other hand, areas with cropping and pasture mosaics presents a medium potential for C sequestration. However, it was evident that as the depth increases, the C sequestration potential improves in the forest uses, especially in agriculture and forest mosaics. For Minasny et al. (2013), historical land use is a variable that influences the explanation of C concentrations in deeper soil layers. Land use change can favor C sequestration because it results in a variation of organic compounds reaching the soil and mechanization can reactivate the C cycle, where bacteria in the environment take advantage of the released C, however, there are residues of this microbial decomposition that can be retained by minerals (Kirsten et al., 2021).

Acosta-Martinez et al. (2004) concluded that continuous monoculture systems had a negative impact on soil function and sustainability. Cultivation and tillage reduce and change the distribution of SOC, while appropriate crop rotation can increase or maintain the quantity and quality of SOM, improving soil chemical and physical properties (Liu et al., 2006). Crop intensity or frequency affects SOC storage by modifying the amount of time the soil is supporting a crop, thereby increasing annual production and C input to the soil (Ogle et al., 2005). Areas with the same land use for more than 15 years were those with higher saturation of minerals, such as goethite and gibbsite (higher affinity for SOM), as well kaolinite and hematite (low affinity for SOM). However, it was clear that those areas with agriculture and pasture mosaics, and pasture and forest mosaics, had the greater potential to retain new C, and its potential increases with soil depth (Table 5).

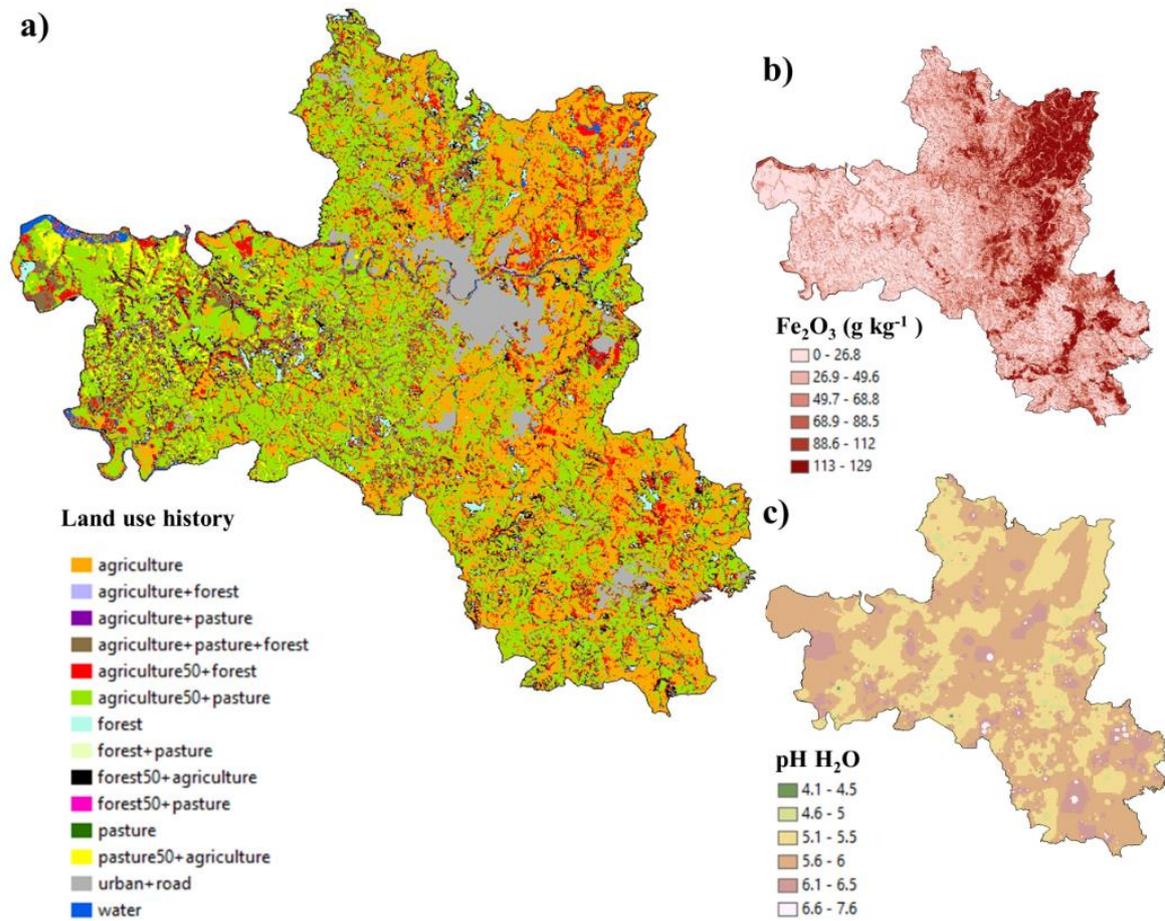


Figure 13. Additional variables, Land use history (1985-2015) based on Tayebi et al., (2021) (a), iron oxide (Fe_2O_3) (b) and pH in water (c).

These areas with higher retention potential due to mineralogy are key to promote CO_2 sequestration by agroforestry and silvopastoral systems, because as evidenced, it is important to exploit the potential of goethite and gibbsite at depth (Figure 11, Table 5), since, as indicated by Georgiou et al. (2022), the deeper the soil minerals are, the less saturated they become. Additionally, crop rotation or cover crops to exploit the potential of the most superficial layers of the soil is also important. The results presented could contribute to climate change mitigation strategies, as described by Minasny et al. (2017), who pointed out that at the 21st Conference of the Parties to the United Nations Framework Convention on Climate Change in Paris (COP21) the strategy "4 per thousand soils for food security and climate" was unveiled. This strategy aims to increase global soil organic matter stocks by 4 per 1000 (or 0.4%) per year considering SOC sequestration as a possible solution to mitigate climate change by taking atmospheric CO_2 and converting it into long-lived soil C.

2.5. Conclusions

The C sequestration potential prediction models obtained in the present study confirm the importance of the minerals that compose the clay fraction in the C sequestration potential of the soil. The prediction of this potential was fitted to a spatial regression model SARAR (Spatial AutoRegressive-AutoRegressive model) for depths of 0 to 20 and 80 to 100 cm, where at a depth of 0 to 20 cm the sequestration potential is explained by the content of kaolinite, hematite, goethite and gibbsite, with kaolinite and hematite being the most important explanatory variables. On the other hand, goethite and gibbsite had a direct but negative impact, indicating that an increase in the concentration of these minerals reduces the potential for sequestration of new C, due to the affinity they have with organic molecules, so they tend to saturate reducing their potential to store new C, but translates into greater stability of organic molecules and higher current COS content. For the 80 to 100 cm depth, the prediction of C sequestration potential was explained by the content of kaolinite and hematite, with greater importance of kaolinite. Hematite is a mineral of importance in C sequestration and stabilization since it was a variable of high importance in explaining mineral-associated C (CmOM), potential C saturation (Sat-pot) and C sequestration potential (Sat-def) at different depths.

Soil carbon sequestration potential by mineralogy is strongly influenced by land use. Areas of pasture and crop on soils with high kaolinite and hematite content presented greater potential to sequester carbon. In addition, areas with lower pH and higher kaolinite and hematite content also have a high potential for carbon sequestration, which can be enhanced by land use change.

Gibbsite and goethite had a higher sorption power of organic molecules; therefore, they had a lower potential for sequestration of new carbon in areas with the same land use for more than 15 years, because they are the first minerals to become saturated, especially in the surface layers. However, their potential increases in cropping and pasture areas at greater depths because the concentration of SOM was lower. Soils at greater depths had the greatest potential for carbon sequestration and could be key for climate change mitigation strategies.

Table 5. Areas expressed in m² with high, medium and low potential saturation of COS according to historical land use, type of mineral and soil depth A) from 0 to 20 cm, B) from 40 to 60 cm, C) from 80 to 100 cm.

Depth	Mineral	C retention potential	Agriculture	Agriculture +Forest	Agriculture +Pasture	Agriculture +pasture+ forest	Agriculture re50+forest	Agriculture e50+pasture	Forest	Forest+ pasture	Forest50+ agriculture	forest50+ pasture	Pasture	Pasture50+ agriculture	Urban+ road	Water
A	Kaolinite	Low	52396.6	465.6	262.7	23562.2	20265.8	38032.0	4393.6	1.6	7323.5	215.1	15.9	1842.2	12186.4	1802.7
		Medium	13677.1	42.0	146.2	6874.3	3763.8	29507.4	473.8	0.4	858.5	19.2	17.6	2106.3	1782.6	36.8
		High	7360.7	19.5	104.4	2856.0	1661.7	19954.7	201.0	0.0	325.6	4.5	16.4	1649.2	648.8	15.6
	Goethite	Low	73434.3	527.1	513.4	33292.5	25691.2	87493.8	5068.4	2.0	8507.6	238.7	50.0	5597.7	14617.8	1855.1
		Medium	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		High	0.1	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Hematite	Low	66288.9	485.9	460.3	30888.7	23123.2	77702.7	4761.1	2.0	8002.5	232.5	43.2	4990.3	14130.4	1728.4
		Medium	3688.8	17.2	37.8	1507.7	1270.6	6256.7	174.5	0.1	275.2	3.5	5.7	495.5	284.7	29.9
		High	3456.6	24.0	15.2	896.1	1297.4	3534.7	132.9	0.0	229.9	2.7	1.1	111.9	202.7	96.8
B	Kaolinite	Low	71.4	12.8	6.2	163.2	81.5	133.5	15.0	0.0	28.6	0.9	0.0	9.5	123.6	107.7
		Medium	48898.5	435.4	341.2	25196.8	19562.4	50230.5	3977.6	1.7	6771.1	203.7	33.1	3593.9	10811.2	1666.4
		High	24464.5	78.9	166.0	7932.5	6047.4	37130.0	1075.8	0.3	1707.9	34.1	16.9	1994.2	3683.0	80.9
	Goethite	Low	40753.6	385.4	131.9	11137.0	14632.5	17110.1	2298.7	0.5	3909.2	112.9	0.7	224.8	9804.2	1749.9
		Medium	29688.4	141.0	342.2	21566.7	10708.3	61886.3	2769.4	1.6	4591.3	125.9	45.4	4814.1	4653.6	105.2
		High	2992.3	0.6	39.3	588.8	350.4	8497.7	0.3	0.0	7.1	0.0	3.8	558.7	160.0	0.0
	Hematite	Low	28685.1	348.7	144.8	13647.5	13272.1	19183.1	2379.2	0.6	4280.7	130.3	3.0	589.5	8785.2	1678.6
		Medium	23448.0	112.6	186.1	11982.5	7155.1	35365.3	1624.2	1.2	2648.0	70.4	23.4	2413.4	3562.1	120.4
		High	21301.2	65.7	182.5	7662.5	5264.0	32945.6	1065.0	0.2	1578.8	38.0	23.6	2594.8	2270.6	56.1
C	Kaolinite	Low	5.2	0.0	0.0	1.5	1.4	0.9	0.0	0.0	0.3	0.1	0.0	0.2	1.4	0.0
		Medium	50559.7	354.1	386.9	24990.2	18462.5	57869.0	4162.6	1.7	6751.2	199.0	42.3	4554.3	9986.5	1302.4
		High	22869.4	173.0	126.5	8300.8	7227.3	29624.2	905.9	0.3	1756.0	39.7	7.7	1043.2	4629.9	552.7
	Goethite	Low	16045.7	38.5	22.1	1323.4	4608.5	3656.2	96.6	0.0	209.0	5.9	0.2	52.1	2363.4	115.8
		Medium	33551.0	431.8	228.7	18888.6	14494.6	29235.3	3986.0	1.3	6420.1	183.0	9.5	1522.5	10177.2	1706.4
		High	23837.6	56.7	262.6	13080.4	6588.1	54602.6	985.9	0.7	1878.5	49.9	40.3	4023.0	2077.2	32.9
	Hematite	Low	3983.1	89.4	42.0	3303.2	2218.1	3529.6	572.7	0.2	1152.4	32.1	1.5	267.2	1649.2	613.1
		Medium	40247.3	294.0	374.5	24427.0	14854.8	62214.4	3759.0	1.6	6058.4	179.0	43.0	4681.5	9522.6	818.3
		High	29204.0	143.7	96.9	5562.3	8618.3	21750.1	736.7	0.2	1296.8	27.7	5.5	649.0	3446.0	423.6

The land uses described correspond to a historical use analysis for a period of 30 years. The number 50 in the heading relates to the middle of the period under this use. The highlighted numbering indicates the largest areas with high, medium and low carbon saturation potential.

References

- ACOSTA-MARTINEZ, V., ZOBECK, T.M. & V. ALLEN. 2004. Soil microbial, chemical and physical properties in continuous cotton and integrated crop-livestock systems. *Soil Sci. Soc. Am. J.*, 68 (2004), pp. 1875-1884
- ADHIKARI, K., HARTEMINK, A., MINASNY, B., BOU KHEIR, R., GREVE, M. & M.H. GREVE. 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE*, 9 (8)
- ALVARES, C.A., STAPE, J.L., SENTELHAS, P.C., DE MORAES GONÇALVES, J.L. & G. SPAROVEK. 2013. Koppen's climate classification map for Brazil. *Meteorol. Z.* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- ANGERS, D.A.; ARROUAYS, D.; SABY, N.P.A. & C. WALTER. 2011. Estimating and mapping the carbon saturation deficit of French agricultural topsoils. *Soil Use and Management* 27:448-452.
- ASHTON, N., TYE, A., PATTRICK, R. & B. VAN DONGEN. 2016. Carbon sequestration in the soils of Northern Ireland: potential based on mineralogical controls. In M.E. Young (ed.), *Unearthed: impacts of the Tellus surveys of the north of Ireland*. Dublin. Royal Irish Academy.
- ARROUAYS, D., SABY, N., WALTER, C., LEMERCIER, B. & C. SCHVARTZ. 2006. Relationships between particle-size distribution and organic carbon in French arable topsoils. *Soil Use and Management*, 22, 48–51.
- BALDOCK, J.A. & J. O. SKJEMSTAD. 2000. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. *Organic Geochemistry*, 31, 7–8, 697–710
- BENITES, V., MACHADO, P., FIDALGO, E., COELHO, M. & B. MADARI. 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. *Geoderma*, 139 (2007), pp. 90-97
- BERG, M. & J. OLIVEIRA. 2000. Variability of apparently homogeneous soils in São Paulo State, Brazil: II. Quality Of Soil Maps. *R. Bras. Ci. Solo*, 24:393-407, 2000
- BIGHAM, J.M.; FITZPATRICK, R.W. & D. SCHULZE. 2002. Iron oxides. p. 323-366. In: Dixon, J.B.; Schulze, D.G., eds. *Soil mineralogy with environmental applications*. Soil Science Society of America Book Series, Madison, WI, USA.
- BODDEY, R.M., JANTALIA, C.P., CONCEIÇÃO, P.C., ZANATTA, J.A., BAYER, C., MIELNICZUK, J., DIECKOW, J., SANTOS, H.P., DENARDIN, J.E., AITA, C., GIACOMINI, S.J., ALVES, B.J.R. & S. URQUIAGA. 2010. Carbon accumulation at depth in Ferralsols under zero-till subtropical agriculture. *Global Change Biology* 16:784-795.
- BONFATTI, B.R., DEMATTÊ, J.A.M., MARQUES, K.P.P., POPPIEL, R.R., RIZZO, R., MENDES, W., SILVERO, N.E.Q. & J.L. SAFANELLI. 2020. Digital mapping of soil parent material in a heterogeneous tropical area. *Geomorphology*, 367 (2020), p. 107305, [10.1016/j.geomorph.2020.107305](https://doi.org/10.1016/j.geomorph.2020.107305)
- BOUYOUCOS, G.J. 1962. Hydrometer method improved for making particle size analysis of soils. *Agronomy Journal*, 54, 464-465. <http://dx.doi.org/10.2134/agronj1962.00021962005400050028x>
- BREIMAN, L. 2001. Random forests. *Machine Learning*, 45 (2001), pp. 5-32

- BRODOWSKI, S., JOHN, B., FLESSA, H. & W. AMELUNG. 2006. Aggregate- occluded black carbon in soil. *Eur. J. Soil. Sci.* 57, 539- 546.
- CARVALHO, L., MONIZ, R., SOUZA, E., VIEIRA, G., SCHAEFER, C. & E. FERNANDES. 2019. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma* 340 (2019) 337–350.
- CARVALHO FILHO, A., INDA, A.V., FINK, J.R. & N. CURI. 2015. Iron oxides in soils of different lithological origins in ferriferous quadrilateral (Minas Gerais, Brazil) *Appl Clay Sci*, 118 (2015), pp. 1-7, 10.1016/j.clay.2015.08.037
- CHUNG, H., GROVE, J.H. & J. SIX. 2008. Indications for soil carbon saturation in a temperate agroecosystem. *Soil Science Society of America Journal*, 72, 1132–1139
- CHURCHMAN, G. J., & B. VELDE. 2019. *Soil Clays, Linking Geology, Biology, Agriculture, and the Environment*. Boca Raton, Florida, USA: CRC Press.
- CHURCHMAN, G.J., SINGH, M., SCHAPPEL, A., SARKAR, B. & N. BOLAN. 2020. Clay minerals as the key to the sequestration of carbon in soils. *Clays Clay Miner.* 68, 135–143 (2020).
- COELHO, F., GIASSON, E., CAMPOS, A., TIECHER, T., COSTA, J. & J. COBLINSKI. 2020. Digital soil class mapping in Brazil: a systematic review. *Soils and Plant Nutrition. Sci. agric. (Piracicaba, Braz.)* 78 (5) 2021. <https://doi.org/10.1590/1678-992X-2019-0227>.
- COTRUFO, M.F., RANALLI, M.G., HADDIX, M.L., SIX, J. & E. LUGATO. 2019. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nature Geoscience*, 12(12), pp.989-994. <https://doi.org/10.1038/s41561-019-0484-6>
- DE MASTRO, F., TRAVERSA, A., COCOZZA, C., PALLARA, M. & G. BRUNETTI. 2020. Soil Organic Carbon Stabilization: Influence of Tillage on Mineralogical and Chemical Parameters. *Soil Syst.* 2020, 4(3), 58.
- DEMATTÊ, J., FONGARO, C., RIZZO, C. & J.L. SAFANELLI. 2018. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.*, 212 (2018), pp. 161-175.
- 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., BELLINASSO, H., FRANCELINO, M.R., SCHAEFER, C.E.G.R., VICENTE, L.E., DOS SANTOS, U.J., DE SÁ BARRETTO SAMPAIO, E. V., MENEZES, R.S.C., DE SOUZA, J.J.L.L., ABRAHÃO, W.A.P., COELHO, R.M., GREGO, C.R., LANI, J.L., FERNANDES, A.R., GONÇALVES, D.A.M., SILVA, S.H.G., DE MENEZES, M.D., CURI, N., COUTO, E.G., DOS ANJOS, L.H.C., CEDDIA, M.B., PINHEIRO, É.F.M., GRUNWALD, S., VASQUES, G.M., MARQUES JÚNIOR, J., DA SILVA, A.J., BARRETO, M.C. D. V., NÓBREGA, G.N., DA SILVA, M.Z., DE SOUZA, S.F., VALLADARES, G.S., VIANA, J.H.M., DA SILVA TERRA, F., HORÁK-TERRA, I., FIORIO, P.R., DA SILVA, R.C., FRADE JÚNIOR, E.F., LIMA, R.H.C., ALBA, J.M.F., DE SOUZA JUNIOR, V.S., BREFIN, M.D.L.M.S., RUIVO, M.D.L.P., FERREIRA, T.O., BRAIT, M.A., CAETANO, N.R., BRINGHENTI, I., DE SOUSA MENDES, W., SAFANELLI, J.L., GUIMARÃES, C.C.B., POPPIEL, R.R., E SOUZA, A.B., QUESADA, C.A. & DO COUTO, H.T.Z. 2019. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma* 354, 113793. <https://doi.org/10.1016/j.geoderma.2019.05.043>

- DI IORIO, E., CIRCELLI, L., LORENZETTI, R., COSTANTINI, E., PERL, S. & C. COLOMBO. 2019. Estimation of andic properties from Vis-NIR diffuse reflectance spectroscopy for volcanic soil classification. *Catena* 182 (2019) 104109
- DIEKOW, J., MIELNICZUK, J., KNICKER, H., BAYER, C., DICK, D.P. & I. KÖGEL-KNABNER. 2005. Carbon and nitrogen stocks in physical fractions of a subtropical Acrisol as influenced by long-term no-till cropping systems and N fertilization. *Plant and Soil* 268:319-328.
- DOS REIS, C., PINHEIRO, D., CALDAS, J. & C. BAYER. 2014. Carbon sequestration in clay and silt fractions of Brazilian soils under conventional and no-tillage systems. *Soils and Plant Nutrition. Sci. agric. (Piracicaba, Braz.)* 71 (4).
- DUIKER, S. W., RHOTON, F. E., TORRENT, J., SMECK, N. E. & R. LAL. 2003. Iron (hydr)oxide crystallinity effects on soil aggregation. *Soil Sci. Soc. Am. J.* 67, 606–611.
- ELHORST, J.P. 2014. *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer, Heidelberg New York Dordrecht London, <http://www.springer.com/economics/regional+science/book/978-3-642-40339-2>
- EUSTERHUES, K., RUMPEL, C. & I. KOGEL-KNABNER. 2005. Organo-mineral associations in sandy acid forest soils: importance of specific surface area, iron oxides and micropores. *Eur. J. Soil Sci.* 56, 753–763. doi: 10.1111/j.1365-2389.2005.00710.x
- FENG, X., SIMPSON, A.J. & M. SIMPSON. 2005. Chemical and mineralogical controls on humic acid sorption to clay mineral surfaces. *Org. Geochem.* 36, 1553-1566.
- FELLER, C. & M.H. BEARE. 1997. Physical control of soil organic matter dynamics in the tropics. *Geoderma*, 79 (1–4) (1997), pp. 69-116
- FOLLETT, R., KIMBLE, J. & R. LAL. 2001. The potential of US grazing lands to sequester soil carbon. Lewis Publishers, Boca Raton, Florida (2001), pp. 401-430
- FREIBAUER, A., ROUNSEVELL, MDA., SMITH, P. & J. VERHAGEN. 2004. Carbon sequestration in the agricultural soils of Europe. *Geoderma*, 122,1–23
- FUJISAKI, K., CHAPUIS-LARDY, L., ALBRECHT, A., RAZAFIMBELO, T., CHOTTE, J.L. & t. CHEVALLIER. 2018. Data synthesis of carbon distribution in particle size fractions of tropical soils: Implications for soil carbon storage potential in croplands. *Geoderma* 313 (2018) 41–51.
- GEORGIU, K., JACKSON, R. B., VINDUŠKOVÁ, O., ABRAMOFF, R. Z., AHLSTRÖM, A., FENG, W., HARDEN, J., PELLEGRINI, A., POLLEY, H., SOONG, J., RILEY, W. & M. S. TORN. 2022. Global stocks and capacity of mineral-associated soil organic carbon. *Nature communications*, 13(1), 3797.
- GRAY, J., BISHOP, T. & B.R. WILSON. 2015. Factors controlling soil organic carbon stocks with depth in Eastern Australia *Soil Sci. Soc. Am. J.*, 79 (6) (2015), pp. 1741-1751
- GUO, P., LI, M., LUO, W., TANG, Q., LIU, Z. & Z.-M. LIN. 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach. *Geoderma*, 237 (2015), pp. 49-59
- GUZMÁN, G., ALCANTARA, E., BARON, V. & J. TORRENT. 1994. Phytoavailability of phosphate adsorbed on ferrihydrite, hematite, and goethite. *Plant and Soil*, 159:219-225, 1994.

- HASSINK, J. 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant and Soil*, 191 (1997), pp. 77-87, 10.1023/A:1004213929699
- HOBLEY, E., WILSON, B., WILKIE, A., GRAY, J & T. KOEN. 2015. Drivers of soil organic carbon storage and vertical distribution in Eastern Australia. *Plant Soil*, 390 (2015), pp. 111-127
- HOFFLAND, E., KUYPER, T., COMANS, R. & R. CREAMER. 2020. Eco-functionality of organic matter in soils. *Plant Soil* (2020) 455:1–22. <https://doi.org/10.1007/s11104-020-04651-9>
- HOGUE, M., WOHLING, T. & W. NOWAK. 2018. A primer for Model Selection: The Decisive Role of Model Complexity. *Water Resources Research*, 54, 1688-1715. <https://doi.org/10.1002/2017WR021902>
- HOUGHTON, R.A. 2003. Why are estimates of the terrestrial carbon balance so different?. *Global Change Biology*. 9, 500-509
- INGRAM, J. & E.C.M. FERNANDES. 2001. Managing carbon sequestration in soils: concepts and terminology *Agric. Ecosyst. Environ.*, 87 (1) (2001), pp. 111-117
- IUSS WORKING GROUP WRB. 2015. World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome, Italy (2015)
- JINDALUANG, W., KHEORUENROMNE, I., SUDDHIPRAKARN, A., SINGH, B. & B. SINGH. 2013. Influence of soil texture and mineralogy on organic matter content and composition in physically separated fractions soils of Thailand. *Geoderma* 195–196 (2013) 207–219. <http://dx.doi.org/10.1016/j.geoderma.2012.12.003>
- JOHNSON, J., FRANZLUEBBERS, A., WEYERS, S. & D. REICOSKY. 2007. Agricultural opportunities to mitigate greenhouse gas emissions. *Environmental Pollution*, 150,107–124.
- KAISER, K. & W. ZECH, W. 2000. ‘Dissolved organic matter sorption by mineral constituents of subsoil clay fractions’, *Journal of Plant Nutrition and Soil Science*, 163, 5, 531–5.
- KAISER, K. & G. GUGGENBERGER. 2000. The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. *Organic Geochemistry* 31:711-725.
- KALBITZ, K., SCHWESIG, D., RETHEMEYER, J. & E. MATZNER. 2005. Stabilization of dissolved organic matter by sorption to the mineral soil. *Soil Biol. Biochem.* 37, 1319-1331.
- KÄMPF, N. & N. CURI. 2003. Argilominerais em solos brasileiros. In: CURI, N. et al. *Tópicos em ciência do solo*. Viçosa: Sociedade Brasileira de Ciência do Solo, 2003. v.3. p.1-54.
- KHALEDIAN, Y. & B.A. MILLER. 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>
- KIMBLE, J.M., HEATH, L.S., BIRDSEY, R.A., LAL, R. 2003. *The Potential of U.S. Forest Soils to Sequester Carbon and Mitigate the Greenhouse Effect*. CRC Press, New York.
- KIRSTEN, M., MIKKUTA, R., VOGEL, C., THOMPSON, A., MUELLER, C., KIMARO, D., BERGSMA, H., FEGER, K. & K. KALBITZ. 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. *Scientific Reports* | (2021) 11:5076

- KÖGEL-KNABNER, I., GUGGENBERGER, G., KLEBER, M. ET AL. 2008. Organo-mineral associations in temperate soils: integrating biology, mineralogy, and organic matter chemistry. *Journal of Plant Nutrition and Soil Science*, 171,61–82
- LAL R. 2004. Soil carbon sequestration impacts on global climate change and food security *Science*, 304 (2004), pp. 1623-1627
- LALONDE, K., MUCCI, A., OUELLET, A. & Y. GÉLINAS. 2012. Preservation of organic matter in sediments promoted by iron. *Nature* 483, 198–200. doi: 10.1038/nature10855
- LAMICHHANE, S., KUMAR, L. & B. WILSON. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* Volume 352, 15 October 2019, Pages 395-413
- LIU, K., ZHOU, Q., WU, W., XIA, T. & H. TANG. 2016. Estimating the crop leaf area index using hyperspectral remote sensing *J. Integr. Agric.*, 15 (2) (2016), pp. 475-491, 10.1016/S2095-3119(15)61073-5
- LIU, X., HERBERT, S.J., HASHEMI, A.M., ZHANG, X. & G. DING. 2006. Effects of agricultural management on soil organic matter and carbon transformation – a review *Plant Soil Environ.*, 52 (2006), pp. 531-543, 10.17221/3544-PSE
- LIU, X. & J. CHEN. 2021. Variable Selection for the Spatial Autoregressive Model with Autoregressive Disturbances. *Mathematics* 2021, 9, 1448. <https://doi.org/10.3390/math9121448>
- LOVELAND, P. & WEBB, J. 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review, *Soil and Tillage research*, 70, 1–18, 2003.
- MARSCHNER, B., BRODOWSKI, S., DREVES, A., GLEIXNER, G., GUDE, A., GROOTES, P.M., HAMER, U., HEIM, A., JANDL, G. & R. JI. 2008. How relevant is recalcitrance for the stabilization of organic matter in soils? *J. Plant Nutr. Soil Sci.*, 171 (2008), pp. 91-110
- MENDES, W., DEMATTÊ, J.A.M., BONFATTI, B., RESENDE, M., CAMPOS, L. & A. DA COSTA, A. 2021. A novel framework to estimate soil mineralogy using soil spectroscopy. *Applied Geochemistry*, Volume 127, 2021, 104909, ISSN 0883-2927, <https://doi.org/10.1016/j.apgeochem.2021.104909>.
- MENDEZ, C. 2020. Spatial regression analysis in R. R Studio/RPubs. Available at <https://rpubs.com/quarcs-lab/tutorial-spatial-regression>.
- MINASNY, B., MCBRATNEY, A. & B.P. MALONE. 2013. WheelerDigital mapping of soil carbon. *Adv. Agron.*, 118 (2013), pp. 1-47
- 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. & L. WINOWIECKI. 2017. Soil carbon 4 per mille. *Geoderma* 292, 59–86. <https://doi.org/10.1016/J.GEODERMA.2017.01.002>

- MCBRATNEY, A., SANTOS, M. & B. MINASNY. 2003. On digital soil mapping. *Geoderma*, 117 (2003), pp. 3-52
- NAVE, E., SWANSTON, C., MISHRA, U. & K.J. NADELHOFFER. 2013. Afforestation effects on soil carbon storage in the United States: a synthesis. *Soil Sci. Soc. Am. J.*, 77 (3) (2013), pp. 1035-1047
- OADES, J. 1988. The retention of organic matter in soils. *Biogeochemistry*, 5,35–70
- OGLE, S.M., BREIDT, F.J. & K. PAUSTIAN. 2005. *Biogeochemistry*, 72 (2005), p. 87, 10.1007/s10533-004-0360-2
- OKUNLOLA, O.A., ALOBID, M., OLUBUSOYE, O.E., AYINDE, K., LUKMAN, A. & I. SZUCS. 2021. Spatial regression and geostatistics discourse with empirical application to precipitation data in Nigeria. *Sci Rep* 11, 16848 (2021). <https://doi.org/10.1038/s41598-021-96124-x>
- PADARIAN, J., MINASNY, B. & A. MCBRATNEY. 2019. Using deep learning for digital soil mapping. *Soil*, 5(1), pp.79-89. <https://doi.org/10.5194/soil-5-79-2019>
- PADARIAN, J., MINASNY, B. & A.B. MCBRATNEY. 2020. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* 6, 35–52. <https://doi.org/10.5194/soil-6-35-2020>
- PADARIAN, J., MINASNY, B., MCBRATNEY, A. & P. SMITH. 2022. Soil carbon sequestration potential in global croplands. *PeerJ*, 10, p.e13740. <https://doi.org/10.7717/peerj.13740>
- PAUSTIAN, K., ANDREN, O., JANZEN, H. ET AL. 1997. Agricultural soils as a sink to mitigate CO₂ emissions. *Soil Use and Management*, 13, 230–244
- PERCIVAL, H. J., PARFITT, R. L. & N. SCOTT. 2000. Factors controlling soil carbon levels in New Zealand grasslands: Is clay content important? *Soil Science Society of America Journal*, 64, 1623–1630.
- POEPLAU, C. & A. DON. 2013. Sensitivity of soil organic carbon stocks and fractions to different land-use changes across Europe. *Geoderma*, 192 (2013), pp. 189-201
- PROUT, J.M., SHEPHERD, K.D., MCGRATH, S.P., KIRK, G.J. & S.M. HAEFELE. 2021. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. *European Journal of Soil Science*, 72(6), pp.2493-2503. <https://doi.org/10.1111/ejss.13012>
- RABBI, S., TIGHE, M., DELGADO-BAQUERIZO, M., COWIE, A., ROBERTSON, F., DALAL, R., PAGE, K., CRAWFORD, D., WILSON, B., SCHWENKE, G., MCLEOD, M., BADGERY, W., DANG, Y., BELL, M., O'LEARY, G., LIU, D. & J. BALDOCK. 2015. Climate and soil properties limit the positive effects of land use reversion on carbon storage in Eastern Australia *Sci. Rep.*, 5 (2015), Article 17866
- RASMUSSEN, C., MATSUYAMA, N., DAHLGREN, R. A., SOUTHARD, R. J. & N. BRAUER. 2007. Soil genesis and mineral transformation across and environmental gradient on andesitic lahar. *Soil Science Society of America Journal*, 71, 225–237.
- RODRIGUES, L., DIECKOW, J., GIACOMINI, S., OTTONELLI, A., ZORZO, G. & C. BAYER. 2022. Carbon sequestration capacity in no-till soil decreases in the long-term due to saturation of fine silt plus clay-size fraction. *Geoderma* 412 (2022) 115711.

- SABETIZADE, M., GORJI, M., ROUDIER, P., ZOLFAGHARI, A. A. & A. KESHAVARZI. 2021. Combination of MIR spectroscopy and environmental covariates to predict soil organic carbon in a semi-arid region, *CATENA*, Volume 196, 2021, 104844, ISSN 0341-8162, <https://doi.org/10.1016/j.catena.2020.104844>.
- SAIDY, A.R., SMERNIK, R.J., BALDOCK, J.A., KAISER, K. & J. SANDERMAN. 2013 'The sorption of organic carbon onto differing clay minerals in the presence and absence of hydrous iron oxide', *Geoderma*, 209–10, 15–21.
- SAMUEL-ROSA, A., DALMOLIN, R.S.D., MOURA-BUENO, J.M., TEIXEIRA, W.G. & J.M. ALBA. 2020. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. *Sci. Agric.* 77. <https://doi.org/10.1590/1678-992x-2017-0430>
- SAUERBECK, D. 2001. CO₂ emissions and C sequestration by agriculture – perspective and limitations. *Nutrient Cycling in Agroecosystems*, 60, 253–266
- SCHAEFER, C., FABRIS, J. D. & J.C. KER. (2008). Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. *Clay Minerals*, 43:137-154, 2008
- SINGH, M., SARKAR, B., SARKAR, S., CHURCHMAN, J., BOLAN, N., MANDAL, S., MENON, M., PURAKAYASTHA, T. J., & BEERLING, D. J. 2018. Stabilization of soil organic carbon as influenced by clay mineralogy. *Advances in Agronomy*, 148, 33–84.
- SIX, J., CONANT, R., PAUL, E. & K. PAUSTIAN. 2002. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils *Plant Soil*, 241 (2002), pp. 155-176
- SOIL SCIENCE DIVISION STAFF. 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). 4th ed. Agriculture Handbook 18. Government Printing Office, Washington, D.C. Retrieved May 26, 2020, from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054262
- STEWART, C., PAUSTIAN, K., CONANT, R., PLANTE, A. & J. SIX. 2008. Soil carbon saturation: evaluation and corroboration by long-term incubations. *Soil Biology & Biochemistry*, 40, 1741–1750
- SHI, Z., JI, W., VISCARRA ROSSEL, R., CHEN, S. & Y. ZHOU. 2015. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library *Eur. J. Soil Sci.*, 66 (4) (2015), pp. 679-687
- SMITH, P. 2012. Agricultural greenhouse gas mitigation potential globally, in Europe and in the UK: what have we learnt in the last 20 years? *Global Change Biology*, 18, 35–43
- SMITH, P. 2004. Carbon sequestration in croplands: the potential in Europe and the global context. *European Journal of Agronomy*, 20, 229–236
- SOLLINS, P., HOMANN, P. & B. CALDWELL. 1996. Stabilization and destabilization of soil organic matter: mechanisms and controls. *Geoderma*, 74, 65–105
- SOTHE, C., GONSAMO, A., ARABIAN, J. & J. SNIDER. 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. *Geoderma* Volume 405, 1 January 2022, 115402

- TAYEBI, M., FIM ROSAS, J.T., MENDES, W.D.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. & J. A.M. DEMATTÊ. 2021. Drivers of Organic Carbon Stocks in Different LULC History and along Soil Depth for a 30 Years Image Time Series. *Remote Sens.* 2021, 13, 2223. <https://doi.org/10.3390/rs13112223>
- TENG, H., VISCARRA, R., SHI, A. & T. BEHRENS. 2018. Updating a national soil classification with spectroscopic predictions and digital soil mapping. *Catena* 164 (2018) 125–134
- USGS. 2018. USGS EROS archive - digital elevation - Shuttle radar Topography mission (SRTM) 1 arc-second global [WWW document] *Earth Resour. Obs. Sci. Cent* <https://doi-org.ezproxy.unal.edu.co/10.5066/F7PR7TFT> (2018)
- VLEESHOUWERS, L. & A. VERHAGEN. 2002. Carbon emission and sequestration by agricultural land use: a model study for Europe. *Global Change Biology*, 8, 519–530
- VON LÜTZOW, M., KÖGEL-KNABNER, I., EKSCHEMITT, K., MATZNER, E., GUGGENBERGER, G., MARSCHNER, B. & H. FLESSA. 2006. Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions—a review *Eur. J. Soil Sci.*, 57 (2006), pp. 426-445
- Wang, Y., Liu, Y., Ynag, K. AND D. Lin. 2020. Reciprocal interference of clay minerals and nanoparticulate zero-valent iron on their interfacial interaction with dissolved organic matter. *Science of the Total Environment* 739 (2020) 140372. <https://doi.org/10.1016/j.scitotenv.2020.140372>
- WALKLEY, A. & I.A. BLACK. 1934. An examination of the Degtjareff method for determining soil organic matter, and proposed modification of the chromic acid titration method. *Soil Science* 37: 29-38
- WEBER, K. A., ACHENBACH, L. A. & J. COATES. 2006. Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat. Rev. Microbiol.* 4, 752–764. doi: 10.1038/nrmicro1490
- WEBSTER, R. & M. OLIVER. 2007. *Geostatistics for Environmental Scientists*. Editorial Jhon Wiley & Sons, Ltd. 333 p.
- WEIL, R. & N.C. BRADY. 2016. *The Nature and Properties of Soils* (15th edition), Pearson Education (2016)
- WEN, Y., XIAO, J., GOODMAN, B. & X. HE. 2019. Effects of Organic Amendments on the Transformation of Fe (Oxyhydr)Oxides and Soil Organic Carbon Storage. *Frontiers in Earth Science*. Volume 7, 2019. <https://www.frontiersin.org/article/10.3389/feart.2019.00257>
- WIESMEIER, M., BARTHOLD, F., BLANK, B. & I. KÖGEL-KNABNER. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem *Plant Soil*, 340 (1–2) (2011), pp. 7-24
- WIESMEIER, M., HUBNER, R., BARTHOLD, F. ET AL. 2013. Amount, distribution and driving factors of soil organic carbon and nitrogen in cropland and grassland soils of southeast Germany (Bavaria). *Agriculture Ecosystems & Environment*, 176, 39–52

- YANG, Y., SHEN, Z., BISSET, A. & R. VISCARRA ROSSEL. 2021. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. SOIL Discuss. [preprint], <https://doi.org/10.5194/soil-2021-79>, in review, 2021.
- YANG, L., LI, X., SHIB, J., SHEN, F., QI, F., GAO, B., CHEN, Z., ZHU, A. & C. ZHOU. 2020. Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. *Geoderma* 369 (2020) 114337.
- ZERAATPISHE, M. & F. KHORMALI. 2012. Carbon stock and mineral factors controlling soil organic carbon in a climatic gradient, Golestan province. *Journal of soil science and plant nutrition*, 12(4), 637-654.
- ZERAATPISHEH, M., JAFARI, A., BODAGHABADI, M.B., AYOUBI, S., TAGHIZADEH-MEHRJARDI, R., TOOMANIAN, N., KERRY, R. & M. XU. 2020. Conventional and digital soil mapping in Iran: past, present, and future *Catena*, 188 (2020), p. 104424

3. SOIL ORGANIC CARBON SEQUESTRATION POTENTIAL EXPLAINED BY MINERALOGY AND MICROBIOLOGICAL ACTIVITY BY SPECTRAL TRANSFER FUNCTIONS

Abstract

The ability of soil to sequester carbon and reduce atmospheric CO₂ concentrations is limited and depends on the soil minerals and their interaction with the microbiota. Microbial activities are closely associated with the types and amounts of soil organic matter (SOM) and clay minerals that have functional groups that interact with energy in Vis NIR-SWIR and Mid-IR wavelengths. The main objective of this research was to determine, based on these spectral ranges, the relation between mineralogical and organic compounds, as their sequestration and specialization in soils from Brazil. It was possible to map microbiological activity by spectral transfer functions and digital soil mapping reaching R² from 0.77 to 0.85. Multiple regression equations were constructed to quantify enzymatic activity, microbial biomass carbon (MBC), particulate organic matter (POM), and resistant forms of carbon, and SOM associated with the mineral fraction (MAOM). All these properties were detected by specific bands obtained with the recursive feature elimination (RFE) algorithm, reaching correlations from 0.64 to 0.98 in specific ranges. The prediction model of the carbon sequestration potential was adjusted with microbiological and mineralogical variables from Vis-NIR-SWIR and the Mid-IR spectral range. A SARAR double autoregressive model was adjusted with r 0.61 and to a spatial error model (SEM) with r 0.7. The explanatory variables were associated with kaolinite, hematite, goethite, gibbsite, and the abundance of fungi, actinomycetes, vesico-arbuscular mycorrhizal fungi, enzymatic activity of beta-glucosidase, urease and phosphatase, and POM. Among the microbiological variables, the general abundance of fungi was the most important, in contrast to enzymatic activity that was the least important. The interaction between the different maps constructed and historical land use allowed the identification of areas that contribute to sequestering new carbon and could be the key to climate change mitigation strategies.

Keywords: Mineralogy, Spectroscopy, Spectral ranges, Soil health, Proximal sensing

3.1. Introduction

Soil dynamics respond to direct interactions between microorganisms, biomineralization, and synergistic co-evolution with plants (Gouda et al., 2018); they allow soil to perform several ecosystem services such as carbon (C) sequestration. This dynamic is considered a method to reduce the concentration of CO₂ in the atmosphere, since soils can absorb about 20% of anthropogenic C emissions and thus help to mitigate climate change (Yang et al., 2021; Padarian et al., 2022). The ability of the soil to sequester C depends especially on the mineralogy and its interactions with microorganisms, mainly in clay fractions, which includes different minerals such as variable amounts of pedogenic Fe and Al oxides that have a critical soil organic carbon (SOC) sorption surface (Kirsten et al., 2021; Dos Reis et al., 2014). Microorganisms are responsible for 80% to 95% of C mineralization (Hassink, 1994), as they

make use of C through respiration (Follett et al., 2001) and increase its concentration through their dead remains. Additionally, they participate in the formation of microaggregates that interfere in SOC stabilization (De Mastro et al., 2020) and contribute to C sequestration potential through extracellular activity that aids soil organic matter (SOM) decomposition and C stabilization (Nicolas et al., 2019).

In general, the degree of SOM decomposition is related to the extracellular enzymatic activity of microbiological organisms (De Beeck et al., 2021). Considering the microbial C-pump concept that Adamczyk et al. (2019) describe, stable forms of SOM result from microbiological activity. As microorganisms generate biomass through metabolic processing of plant residues, microbial residues are converted into stable forms by interacting with soil minerals. Rodriguez-Albarracín et al. (2023) identify the importance and varying capacity of soil minerals in predicting the potential to sequester new C. Aluminum and iron oxides, as well as clay minerals (especially smectites), control C storage and release. However, it is still unclear how they relate to the presence of microbes and extracellular enzymes that degrade SOM (Yang et al., 2021). Bacterial adhesion to clay minerals and oxides is mediated by interactions with proteins, extracellular enzymes, and hydrogen bonds (Parink et al., 2014).

Our understanding of the activity and diversity of exoenzymes and their interactions with minerals in SOC modeling needs to be enhanced, because microbial and extracellular enzyme activity directly affects the effectiveness of SOC mineral protection and promotes its release (Yang et al., 2021). Hart et al. (2020) propose the use of spectroscopy-focused technologies in association with molecular leads to model the presence of fungi and bacteria in soil samples. However, there are no specifics visible (Vis: 400 - 700 nm), near infrared (NIR: 700 - 1100 nm), shortwave infrared (SWIR: 1100 - 2500 nm), and mid-infrared (Mid-IR: 2500 - 25000 nm, $4000 - 400 \text{ cm}^{-1}$) absorbances assigned to microbial communities. Nevertheless, since the soil physicochemical environmental conditions and the abundance and function of microorganisms (Rasche et al., 2011) and the spectral response of soil samples allow predicting fundamental components like minerals, SOM, and water content (Yang et al. 2021; Viscarra Rossel et al., 2022) that are required by fungal and bacterial communities for their growth and obtaining energy (Müller, 2015), it is possible to relate wavelengths to microbiological characteristics. Iron, for example, is related to microbial energy generation for Fe-reducing microorganisms (Weber et al., 2006), as well as for phototrophic bacteria (Hegler et al., 2008); and the spectral response of iron oxides is recognized on faults 540, 640 and 900 nm, hematite on fault 550 nm, and goethite around 440 to 470 nm (Dematte et al., 2014).

The major difficulty to access soil microbiological properties directly by spectral frequencies is due to the low contents of microbial biomass and enzymatic activity that limits the induction of changes in the spectra (Rinnan and Rinnan, 2007). However, microbiological properties can be predicted due to the strong relationship with specific soil organic compounds that influence their spectral responses (Chodak 2011). As an example, extracellular enzymes produced by fungi for lignin and lignocellulose degradation release C into the soil solution (Nicolas et al., 2019), making it possible to interpret spectral information as an indicator of soil microbiology (Rasche et al., 2013; Viscarra Rossel et al., 2022). Not only the quantity but also the quality of organic carbon (OC) interferes in the prediction of microbiological properties (Rasche et al., 2013). Thus, the differentiation of organic functional groups is possible through Vis-NIR-SWIR and Mid-IR spectroscopy (Ojeda et al., 2008; Viscarra Rossel and Hicks, 2015; Viscarra Rossel et al., 2022).

NIR reflectance spectra respond to the concentration of compounds related to C-H, N-H, S-H, C=O and O-H chemical bonds, allowing a differentiation of the organic composition of a soil sample (Zornoza et al., 2008; Viscarra Rossel et al., 2022). These compounds are absorbed more strongly in the Mid-IR region (Burns and Ciurczak, 2001) making it possible to identify wavelengths related to the specific property to be predicted by multivariate regression analyses (Zornoza et al., 2008). In this way, Rasche et al. (2013) develop a partial least squares regression (PLSR) analysis based on Mid-IR reflectance spectroscopy to predict soil microbial biomass and enzyme activities.

The SOM fractionation analysis allows a better understanding of the destiny of the sequestered and stabilized C by the discrimination of labile and more stable forms of C (Demyan et al., 2012). Organic matter (OM) associated with the fine-sized mineral fraction contributes more to SOM compared to particulate OM (POM) (Kleber et al, 2015). The SOM associated with the fine mineral fraction (MAOM) contains the largest amount of transformed microbial SOM, including microbial biomass and necromass (Buckeridge et al., 2020; Liang et al., 2019), while the POM is the raw material for microbial activation. In general, microbial activities are closely associated with the types and amounts of MOS, and NIR and Mid IR wavelengths that respond to functional groups of different types of soil organic compounds (Viscarra Rossel and Hicks, 2015), iron oxides, and clay minerals (Yang et al., 2021) that are related to microbial activity.

Identifying the influence of microbial activity on the C sequestration potential of clay minerals allows an understanding of C dynamics and promotes sustainable soil management. Monitoring soil quality and health requires predictions of soil microbiological properties

using a simple and cost-effective method (Rasche et al., 2013). Geographic information systems and remote sensing techniques are key for this by facilitating the acquisition of spatial information specifically through digital soil mapping (DSM) approaches. This allows mapping soil attributes by combining soil point data with statistically correlated auxiliary data (covariates) (McBratney et al., 2003). Reflectance spectroscopy is also key, as it provides access to soil mineralogical and microbiological activity (Di Iorio et al., 2019). A clear example of the combination of these techniques is presented in Rodríguez-Albarracín et al. (2023), in which a prediction model of the soil's potential to sequester new C was adjusted and this potential was mapped by differentiating the storage capacity of each mineral using mainly machine learning techniques and soil mineralogical information from spectral data. In summary, there is a gap in our understanding of the energy interaction between mineralogical and organic compounds. Therefore, it is important to deepen our understanding of C sequestration involving microbiological effects mediated by the mineral component of the soil.

We aimed to map microbiological activity in different pedogenetic soils of the Piracicaba area (São Paulo, Brazil) by reflectance spectroscopy and a DSM approach and to fit the C sequestration potential model developed by Rodríguez-Albarracín et al. (2023), through the inclusion of these microbiological variables in the spatial regression models while considering the interaction of these variables with mineralogy. We expected that the detection of this microbiological activity by spectral readings would favor an understanding of its influence on C sequestration.

3.2. Material and Methods

3.2.1. Study area

The study area was located in São Paulo State, Brazil, including approximately 2,598 km², comprising eight municipalities (Figure 1). The region has a mean annual temperature varying between 20°C and 22.5°C, an annual rainfall between 1200 mm and 1400 mm, with a subtropical climate with a dry winter and a rainy summer, classified Cwa according to the Köppen classification (Alvares et al., 2013). The topographical relief is characterized by undulating highlands and undulating hills with altitudes ranging from 450 to 950 meters. The area is dominated by agricultural land use, principally sugarcane and pasture, under no-till and conventional tillage management systems. The main soil types from the World Reference Base system (IUSS Working Group WRB, 2015) are Cambisol, Gleysol, Ferralsol, Nitosol, Lixisol, Leptsol, Arenosol and Planosol (Oliveira and Prado, 1989).

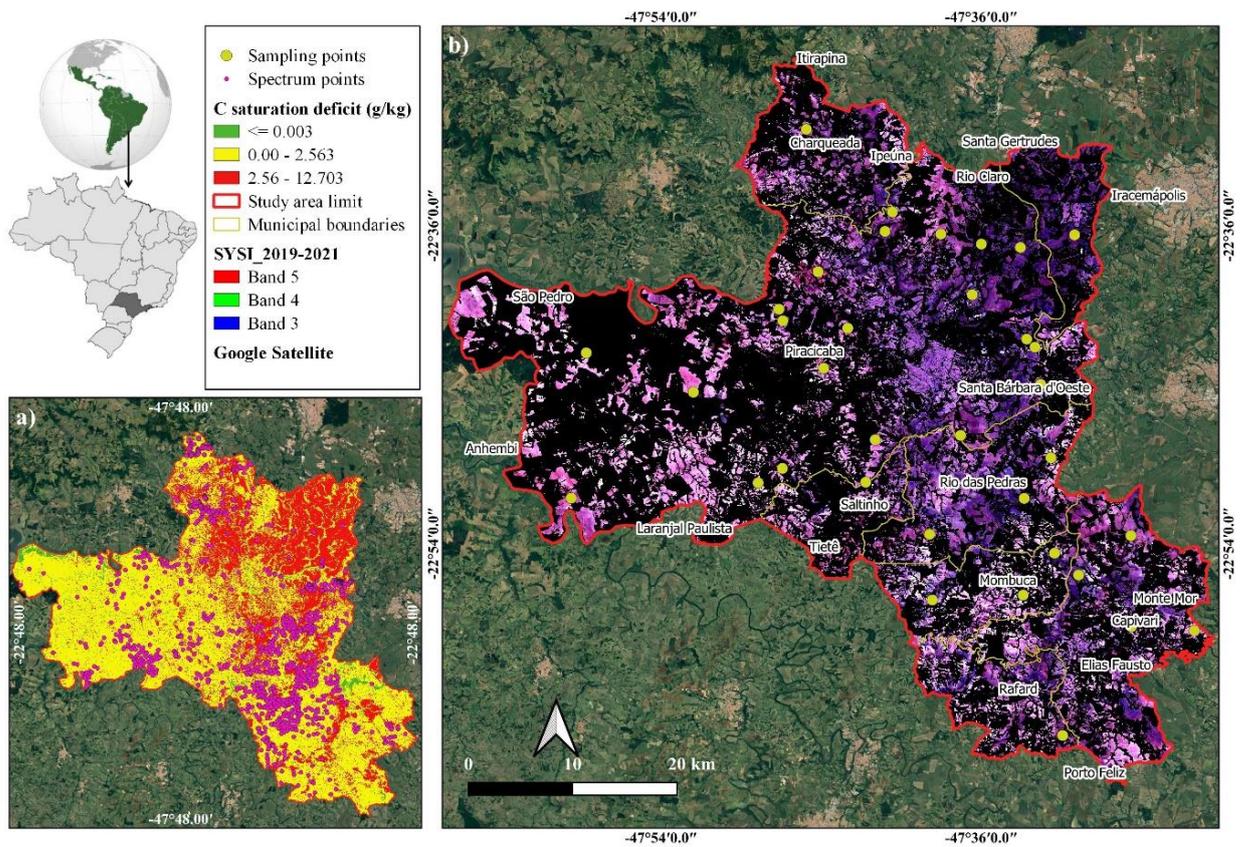


Figure 1. Study area and sample points. a) Map of potential C saturation deficit (C sequestration potential) for depths of 0 cm to 20 cm (Rodríguez-Albarracín et al., 2023); b) synthetic soil image (SYSI) RGB 543 (image from Geos3, Demattê et al., 2018).

3.2.2. Microbiological activity estimated by spectrotransfer functions

Two methods were developed for the prediction and mapping of microbiological activity. The first one considered bibliographic information associated with specific wave coverages for each variable described in Table 1 to obtain relative values in terms of reflectance. The second method was based on the development of specific prediction equations that allowed quantifying the enzymatic activity of beta-glucosidase, phosphatase, and urease, microbial biomass carbon (MBC), POM and MAOM, through regression analyses between laboratory measurements and the specific bands of each spectral range for each variable, identified with machine learning techniques.

Selection of sampling points

New 35 soil samples from representative sites of the study area were selected for soil sampling through the conditional Latin hypercube sampling method (Minasny and

McBratney, 2006) (Figure 2). This corresponded to a stratified random sampling procedure in which the selected samples follow multivariate characteristics according to the indicated covariates (Yang et al., 2020). In this case, we considered as covariates the synthetic soil image (SYSI), terrain elevation model (DEM), soil type, variability in clay and C content, and the maps obtained from method 1 of microbial biomass carbon (MBC), general abundance of fungi, Actinomycetes and Acidobacteria, enzymatic activity of beta-glucosidase, urease and acid phosphatase, and maps of labile and resistant C forms.

Laboratory analysis

For these 35 samples, we performed fractionation of SOM and microbiology analyses (Figure 2). We made the physical fractionation of SOM to quantify C in POM and MAOM (Cotrufo et al., 2019) following Jindaluang et al. (2013). We dispersed the soil sample with a 5% sodium hexametaphosphate solution, and we separated the sand and clay/silt fraction by sieving at 53 μm . The MBC quantification was performed according to Vance et al. (1987). We determined the enzymatic activity of beta-glucosidase, acid phosphatase, and urease following Tabatabai (1994) and Dick et al. (1996).

Spectral data acquisition

The Vis-NIR-SWIR data, from 350 to 2500 nm, were acquired using the FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, Col., USA) with a spectral resolution of 3 nm for the range between 400 nm and 700 nm and 10 nm for the range from 700 nm to 2500 nm. It was then resampled to 1 nm, obtaining 2,151 spectral bands (Demattê et al., 2019). Readings with this sensor required air-dried soil samples, ground and sieved at 2 mm. Mid-infrared data from 4000 to 600 cm^{-1} were obtained with the Alpha Sample Compartment RT-DLaTGS ZnSe sensor (Bruker Optik GmbH), equipped with a drift attachment (Souza et al., 2022). The Alpha sensor had a resolution of 1.2 nm and worked 64 scans per second, according to Mendes et al. (2022). Readings with this sensor required samples to be ground and sieved to 100 mesh (149 μm).

3.2.2.1. Method 1

We used legacy soil observations from the Brazilian Soil Spectral Library (BSSL) (Demattê et al., 2019; Mendes et al., 2022; Demattê et al., 2023) totaling 1,828 points at 0 cm to 20 cm depth (Figure 1). A total of 403 observations had spectral data in the Mid-IR range and 1,425 in the Vis-NIR-SWIR range. The wavelengths of the Vis-NIR-SWIR and Mid-IR ranges related to the microbiological activity variables were defined according to Zornoza et

al. (2008); Rasche et al. (2013); Parikh et al. (2014); and Zhang et al. (2022) (Figure 2, Table 1).

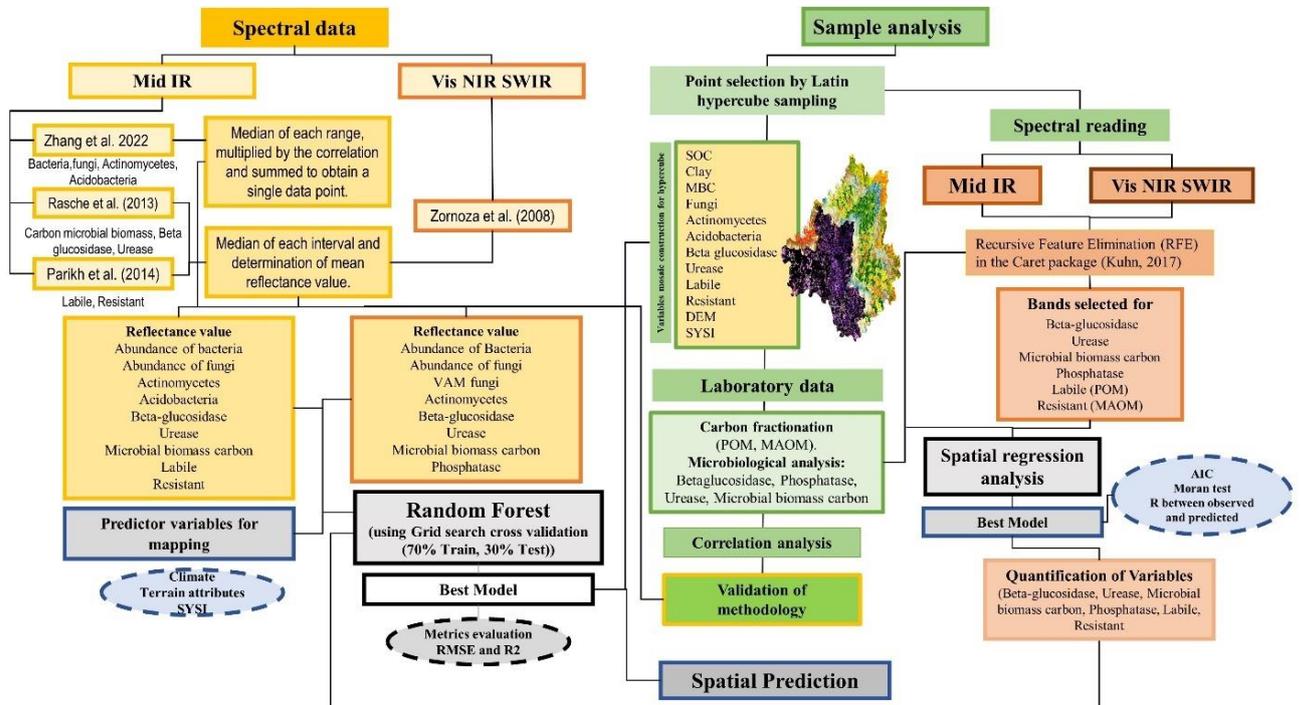


Figure 2. Flowchart of methodology.

Table 1. Wavelengths related to soil microbiological activity reported in literature. In the gray fields, no correspondence was found in the literature.

Variable	Mid IR wavelengths	Reference	Vis-NIR-SWIR wavelengths	Reference
<i>Microbial biomass carbon (MBC)</i>	3658-3317, 2301-939, 3658-3317, 2980-2399, 1959-939	Rasche et al. (2013)	1374-2092, 2270-2630	
<i>Beta-glucosidase</i>	3951-2399, 2301-1279, 3951-2978, 2640-2399, 2301-1618, 941-630	Rasche et al. (2013)	1372-2272	Zornoza et al. (2008)
<i>Urease</i>	3951-2978, 2640-2399, 1959-1279, 3658-2978, 2640-2399, 1959-1618	Rasche et al. (2013)	1195-1734, 1912-2094	
<i>Phosphatase</i>			1014-1195, 1732-2092, 2270-2630	
<i>Labile (POM)</i>	1727, 1650, 1160, 1127, 1050	Parikh et al. (2014)		
<i>Resistant (MAOM)</i>	2950, 2924, 2850, 1530, 1509, 1457, 1420, 779	Parikh et al. (2014)		
<i>Fungi</i>	3990-3930, 3920-3790, 3780-3710, 3520-3140, 3130-3050, 3040-3000, 2990-2860, 2850-2820, 2810-2740, 2730-2670, 2270-2180, 2170-2050, 2040-1810, 1800-1790, 1740-1730, 1360-1320, 1240-1230, 1220-1130, 1120-1040, 820	Zhang et al. (2022)	1193-1555, 1912-2453	
<i>Vesicular arbuscular mycorrhizal fungi</i>			1014-1195, 1374-1555, 2092-2274, 2452-2632	Zornoza et al. (2008)
<i>Bacteria</i>	3990-3840, 3830-3720, 3550-3520, 3510-3110, 3100-3030, 3020-2850, 2840-2780, 2370-2280, 2270-2160, 2150-1810, 1800-1780, 1500-1480, 1420-1300, 1230-1200, 1190-1060, 1050-1040, 820	Zhang et al. (2022)	1195-1913, 2092-2274, 2452-2632	
<i>Actinomycetes</i>	3990-3910, 3900-3730, 3720, 3560-3540, 3530-3120, 3110-3010, 3000-2950, 2940-2850, 2840-2810, 2370-2290, 2280-2150, 2140-1810, 1800-1780, 1530-1470, 1450-1430, 1420-1310, 1300-1290, 1230-1200, 1190-1050, 1040-1030	Zhang et al. (2022)	1014-1195, 2092-2274, 2452-2632	
<i>Acidobacteria</i>	3680-3660, 3650-3590, 2980-2920, 2910-2870, 2860-2840, 2830-2690, 2680-2650, 2640-2570, 2560-2540, 2490-2380, 1760-1720, 960-860, 760-710, 640-630	Zhang et al. (2022)		

For each spectral range reported by Zornoza et al. (2008), Rasche et al. (2013) and Parikh et al. (2014), the median was calculated and subsequently a mean reflectance value was determined as an activity index for each variable, as follows:

$$\bar{Y}_X = \frac{\sum MdW}{No. W}$$

Where X corresponds to the variables associated with microbiological activity described in Table 1; Y activity index (mean value of reflectance); Md is the median; W_i is the specific spectral range described in Table 1; $No.$ is the number.

The correlation values (a) between the microbiological variables and the ranges defined by Zhang et al. (2022) were available and used as weights in the weighted sum to define a single reflectance value for each variable, as follows:

$$\bar{Y}_x = \sum a * MdW$$

For mapping the average reflectance values Y_x of the different variables associated with microbiological activity, covariates associated with relief, climate, and the synthetic soil image (SYSI) were used (Figure 2). For topographical relief we used the elevation, slope, aspect, curvatures, valleys, hills, orientation, and topographic moisture index, obtained from the radar topography mission (SRTM) (USGS, 2018) with the digital terrain elevation model (DEM) at 30 meters spatial resolution (Safanelli et al., 2020). The climatic variables corresponding to the annual temperature trend, mean diurnal range, seasonality (coefficient of variation), maximum temperature of the warmest month, annual trend, and the seasonality of precipitation were taken from the 1970 to 2000 dataset of WorldClim2 with resolution of 1 km (Fick and Hijmans, 2014). They were resampled to 30 m in R software through the resampling function and the bilinear method (Gómez et al., 2023). The land surface temperature (LST) obtained from the median of the Landsat7 images collection from 1999 to 2020 (Ermida et al., 2020) was also included. Finally, a synthetic soil image (SYSI) that corresponds to a mosaic of bare soil pixel from the Landsat 5,7 and 8, images collected from 1984 to 2020, was used. The SYSI had six bands in the Vis-NIR-SWIR spectral range (blue, green, red, NIR, SWIR1 and SWIR2) and was obtained by the Geospatial Soil Sensing System (GEOS3) method, developed by Demattê et al. (2018).

The spatial prediction was made by the random forest (RF) algorithm, reported as the best-performing algorithm for the prediction and mapping of a wide range of soil attributes such as SOC (Gómez et al., 2023; Zeraatpisheh et al., 2020; Padarian et al., 2020). This algorithm is based on a combination of different decision trees in the training stage. A random sampling of the covariate values distributed identically for each tree is performed, allowing the obtention of an average value for the prediction of all trees (Breiman, 2001; Gómez et al., 2023). The samples were randomly split 70% and 30% for calibration and validation. The adjusted coefficient of determination (R^2) was used as a model evaluation metric.

3.2.2.1.1. Correlation analysis

The pixel information of microbiological maps obtained by methodology 1 was sampled for these 35 observation points to determine their correlation with truth laboratory analysis. The correlation coefficient was used to quantitatively evaluate the degree of correlation between variables (Zhao et al., 2022), and Spearman's correlation coefficient, which is one of the most suitable and accurate metrics for this type of analysis because it is applicable for data with normal and non-normal distributions and is effective for characterizing linear and non-linear correlations, robust and insensitive to outliers (Zhang and Wang, 2023).

3.2.2.2. Method 2

3.2.2.2.1. Spectral ranges selection

Spectral information is a high-dimensional dataset that may contain redundant and irrelevant information (Chen et al., 2018); therefore, it is necessary to select a more specific dataset according to its importance (Wang and Li, 2023). For this purpose, we applied the recursive feature elimination (RFE) algorithm developed by Guyon et al. (2002) and implemented in the "Caret" package of R, using the random forest (RF) machine learning method as an internal model (Kuhn, 2021).

RFE is a method with wide applicability that allows handling nonlinear and more complex relationships, considering in an integral way the relationships between the analyzed factors (Wang and Li, 2023). This algorithm is based on backward selection. Initially a model is created using all n predictors, the performance is calculated by k -fold cross-validation (RMSE) and the importance of the variable. Subsequently, the least important predictor is eliminated from the set, and the model is readjusted. Again, the performance is evaluated and the least important predictor is eliminated. This is done repeatedly until the optimal number of predictors is determined by taking the model with the best performance (RMSE) (Zhang et al., 2023). For this research, we applied the RFE algorithm for each variable and for each range of the electromagnetic spectrum, thus obtaining the most important Vis-NIR-SWIR and Mid-IR bands for each variable (Figure 2). Using this data, we built regression models to quantify the variables through spectral information.

3.2.2.2.2. Quantification and regression analysis with geo-referenced data

The regression analyses studied considered the geographical coordinates of the samples and the spatial dependence of soil variability (Webster and Oliver, 2007) and made use of a weight matrix associated to all neighbors. MBC, enzymatic activity of beta-glucosidase,

phosphatase and urease, POM, and the MAOM were considered as dependent variables. As explanatory variables, the reflectance values of the most important bands for each variable and for each Vis-NIR-SWIR and Mid IR electromagnetic spectrum were considered; regression models were performed for each range of the spectrum and each variable.

The spatial regression models applied for the prediction of each microbiological variable corresponded to the spatial autocorrelation models (SAC), the spatially lagged models (SLM), the spatial error model (SEM and SDEM), the SARAR double autoregressive model that includes the autoregressive component of the response and the residuals, allowing us to explain the spatial dependence of the residuals (Elhorts, 2014).

These models are expressed in matrix form as follows:

$$Y = \lambda WY + \alpha \iota_n + X + u; \quad |u| < 1$$

$$u = \rho Wu + \varepsilon \quad |\rho| < 1$$

Where Y represents each microbiological variable (MBC, beta-glucosidase, phosphatase, urease, POM, and MAOM), X represents the matrix of explanatory variables associated with the reflectance of the most important set of bands for each variable Y , W corresponds to the matrix of weights in relation to the distances of the nearest neighbor centroids in the polygons generated by tessellation of the soil sampling points, λ represents the spatial autoregressive coefficient, ρ the spatial autocorrelation coefficient, α corresponds to the intercept, β represents the parameters linked to the explanatory variables, u is associated with the vector of residuals with spatial dependence $\varepsilon N(0, \sigma^2 I)$, and I is an identity matrix" (Rodríguez-Albarracín et al., 2023).

Additionally, the multivariate linear regression model (MRA) was tested considering also the X and Y coordinates as explanatory variables, as described by Böhner and Bechtel (2018). This type of analysis is considered as an extension of the linear spatial trend function and if the assumptions of independence of residuals are satisfied, it can be applied as a function of prediction and spatialization of unknown values.

The choice of the best fit models was based on the lowest value of the Akaike information criterion (AIC). The assumption of independence of residuals through the Moran index test (MIT) with the matrix of weights of all neighbors (Liu and Chen, 2021) is satisfied where independence is fulfilled when the p-value of the test is greater than 0.05. And the highest correlation value (r) between the observed values of the response of the dependent variables Y and the values estimated by the model is considered as an additional criterion (Hoge et al., 2018).

It is important to interpret the impact of the explanatory variables of the selected spatial regression model, taking into account that in some models such as autoregressive models, such interpretation cannot be done directly with the coefficients of the model. In the case of classical regression models the impacts of the coefficients must be estimated and interpreted, as described by Elhorst (2014) through the determination of direct, indirect, and total impacts obtained from the `impacts` function of the `spatialreg` library of R (Méndez, 2020). For this study, direct impacts were analyzed and the relative importance of each explanatory variable was calculated as a function of the total impacts.

3.2.3. Predictive models for C sequestration potential

The potential deficit of SOC saturation or C sequestration potential (*Sat-def*) relates to the additional amount of OC that can stabilize a soil, considering that there is a potential saturation (*sat-pot*) (referring to a theoretical maximum of SOC that can be stabilized in fine soil particles), the C sequestration potential will correspond to the difference between the potential saturation and the actual C (Figure 3). The latter is related to the C contained in the fine particles that is the MAOM fraction (Hassink, 1997; Fujisaki et al., 2018). The potential saturation of C for the study area was calculated using Feller and Beare (1997); and the actual C was calculated with Rodríguez-Albarracín et al. (2023). The methodological details are described in that study.

Rodríguez-Albarracín et al. (2023) determined the individual contribution of soil minerals on the C sequestration potential. However, the activity of microorganisms must be considered due to their effects on C dynamics (Follet et al., 2001). In this sense, spatial regression models were developed as a dependent variable of the C sequestration potential (*Sat-def*) determined by Rodríguez-Albarracín et al. (2023) for depths of 0 cm to 20 cm, maintaining the explanatory variables proposed by the authors associated with the relative abundance of soil minerals (goethite, hematite, gibbsite, kaolinite). These variables were represented by their spectral amplitudes in the Vis-NIR-SWIR range, adding as explanatory variables the MBC, beta-glucosidase, phosphatase and urease activity, POM, and MAOM obtained with method 2 in addition to the general abundance of bacteria, fungi, actinomycetes, acidobacteria, and vesicular arbuscular mycorrhizal fungi obtained with method 1 (Figure 3). The models described in the previous section were tested, and the best model was selected considering the fulfillment of the assumptions described in that section.

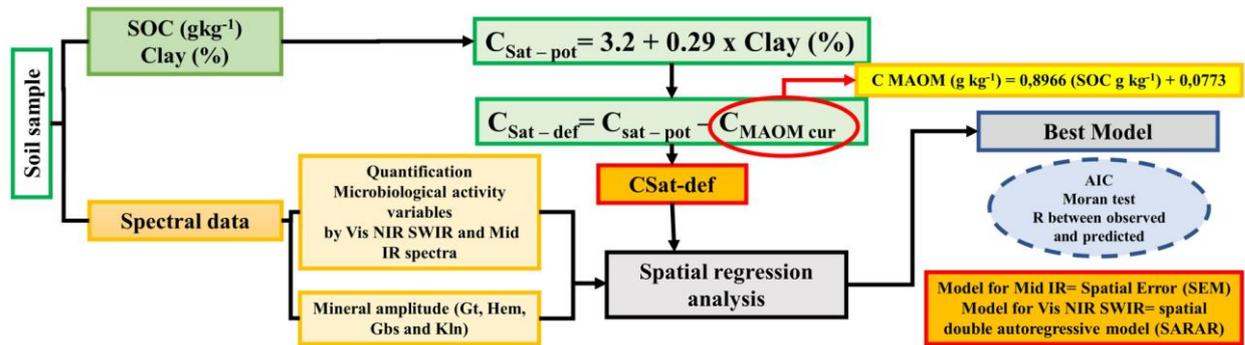


Figure 3. Flow chart of the carbon sequestration potential adjustment through the inclusion of variables associated with microbiological activity. These are sat-pot: potential C saturation; sat-def: C saturation deficit; CMAOM cur: actual concentration of C in mineral-associated organic matter; Gt: goethite, Hem: hematite; Gbs: gibbsite; Kln: kaolinite; AIC: Akaike information criterion.

3.3. Results

3.3.1. Method 1

The spatial prediction of bacterial abundance, fungal abundance, actinomycetes, and acidobacteria were obtained with an R^2 of 0.67, 0.63, 0.62 and 0.69 using Mid-IR reflectance (Figure 4) and making use of the RF algorithm. An improvement in R^2 was observed with the Vis-NIR-SWIR dataset, reaching 0.77 for bacterial abundance, 0.80 for fungi abundance, 0.85 for actinomycetes, and 0.78 for the additional map of vesicular-arbuscular mycorrhizal fungi with an R^2 of 0.78. It is evident that a similar distribution is maintained with both ranges of the electromagnetic spectrum. This distribution of organisms responds to a greater amount of iron oxides (Figure 8). We observed potential values of microorganism abundance or an index of microorganism activity based on reflectance (Figure 4). The lower reflectance values were associated with a higher activity rate of the organisms. The areas corresponding to the higher abundance (green ones) were related to higher iron oxide content, especially of gibbsite. Towards the western zone, some green areas can also be observed, marked in the Vis-NIR-SWIR and responding to medium contents of kaolinite. Similarly, in the central-western zone, green zones associated with high contents of kaolinite, goethite, and gibbsite can be seen for both spectral ranges.

The enzyme activity index maps obtained with the first method (Figure 5 a, c, e, g, i, k) shows some variations, especially in the distribution of beta-glucosidase and urease for both ranges of the electromagnetic spectrum. These had a higher potential content of enzymes (dark and light green areas) in the Mid-IR spectrum (Figure 5 a, c, e). Some areas with lower contents (red color) can be seen in the western, central-eastern, and north-eastern zones. This

zone (north-eastern) seems to coincide with low gibbsite contents, while the other zones are due to their urban location. This differentiation is not observed with the Vis-NIR-SWIR spectra (Figure 5 g, i, k), where the zones with potentially high contents are lower. Some zones with low contents (red color) are observed towards the southeast that seem to be related to low hematite and goethite contents. The enzymatic activity of acid phosphatase was only elaborated for the Vis-NIR-SWIR electromagnetic spectrum, according to the spectral range (Zornoza et al., 2008). As with the other enzymes, the highest potential contents are in the zones with the highest iron oxide content (Figure 8). However, unlike the other enzymes, the areas with low iron oxide contents and in general of minerals associated with the clay fraction are observed with low potential contents.

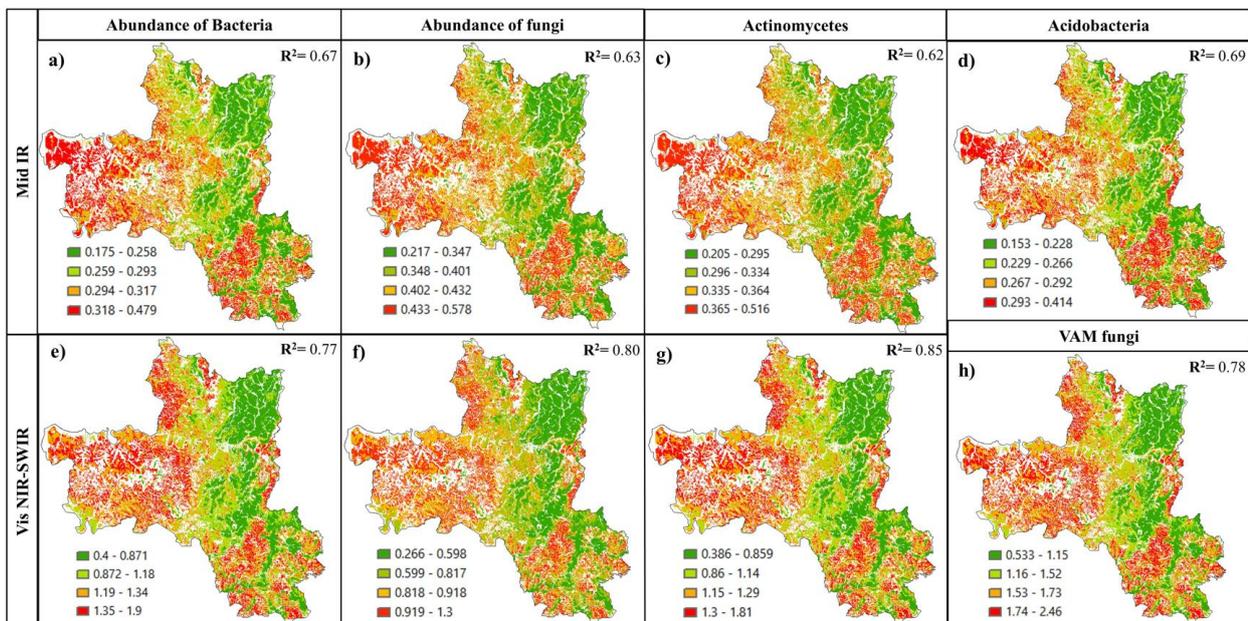


Figure 4. Spatial prediction of mycorrhizal activity index based on reflectance of Mid-IR and Vis-NIR-SWIR spectral ranges for bacterial abundance (a, e), fungal abundance (b, f), actinomycetes (c, g), acidobacteria (d) and vesicular arbuscular mycorrhizal fungi (h). R^2 is coefficient of determination of the spatial prediction

Spearman's correlation analysis between laboratory measurements and the microbiological activity index maps (Figure 5 and 6 a, c, e, g, i, k) showed high correlations for the two ranges of the electromagnetic spectrum evaluated. For Mid-IR, correlations of -0.66 for beta-glucosidase and -0.57 for urease were obtained (Figure 5 a, c). For Vis NIR-SWIR we calculated correlations of -0.55 for beta-glucosidase, -0.70 urease, and -0.65 acid phosphatase (Figure 5 g, i, k). The correlations are negative because there is an inverse relationship between reflectance values and microbiological activity; higher reflectance translates into lower activity. For MBC maps there was a similar distribution in both

electromagnetic spectrums. However, with Mid-IR (Figure 6 a) there were larger areas in red that represented lower potential MBC as occur in urban areas. In contrast for Vis-NIR-SWIR (Figure 6g) this differentiation was not seen, but areas with average potential values were detailed towards the western zone, related to average contents of gibbsite. Correlations between map values and laboratory data were obtained for MBC of -0.63 for Mid IR and -0.59 for Vis NIR-SWIR.

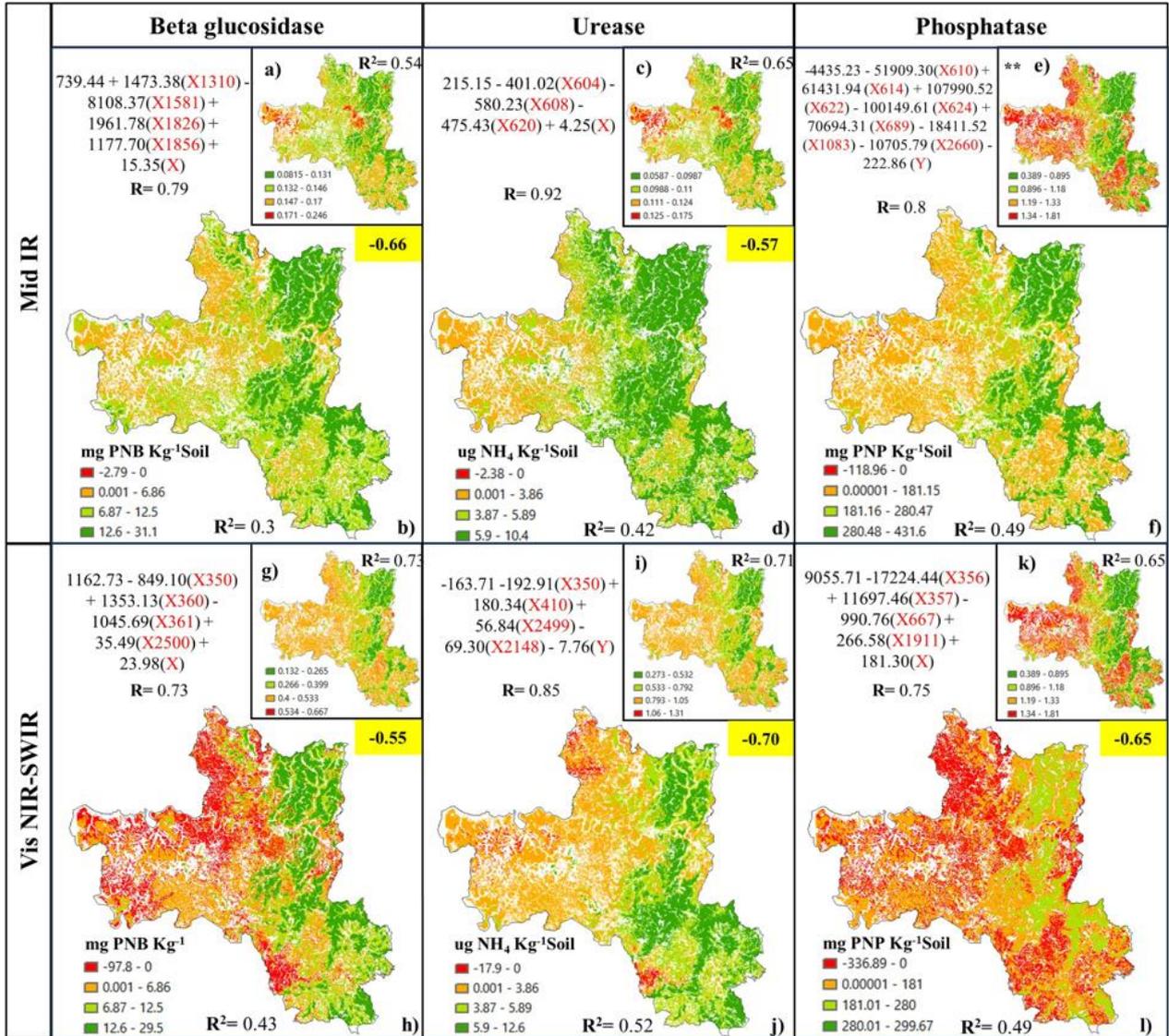


Figure 5. Spatial prediction based on Mid-IR and Vis-NIR-SWIR reflectance of the enzyme activity index determined in method 1 for beta-glucosidase (a, g), urease (c, i), and phosphatase (e, k). Values inside yellow boxes correspond to Spearman correlations between laboratory analyses and activity indexes. Spatial prediction of enzyme quantification results of method 2 for beta-glucosidase (b, h), urease (d, j), phosphatase (f, l). The quantification equations for each variable are shown; the numbers in red indicate the specific bands X and Y correspond to the longitude and latitude. R is the correlation value between the observed value and the one predicted by the quantification equation. R^2 is coefficient of determination of the spatial prediction.

We also elaborated POM and MAOM maps (Figure 6 c and e), with the Mid-IR, including the ranges reported by Parikh et al. (2014) that follow a similar distribution to the MBC, especially MAOM. The POM showed small differentiations towards the northeastern zone with potentially lower values. The correlations obtained with the POM and MAOM laboratory values were -0.28 and -0.74. These results allow validation of these

microbiological activity maps elaborated from soil spectral data. They allowed us to validate the methodology for those variables that could not be analyzed in the laboratory, such as the abundance of bacteria, fungi, actinobacteria, acidobacteria, and vesicular arbuscular mycorrhizal fungi that contributed to the development of exploratory analysis of potential values of these variables.

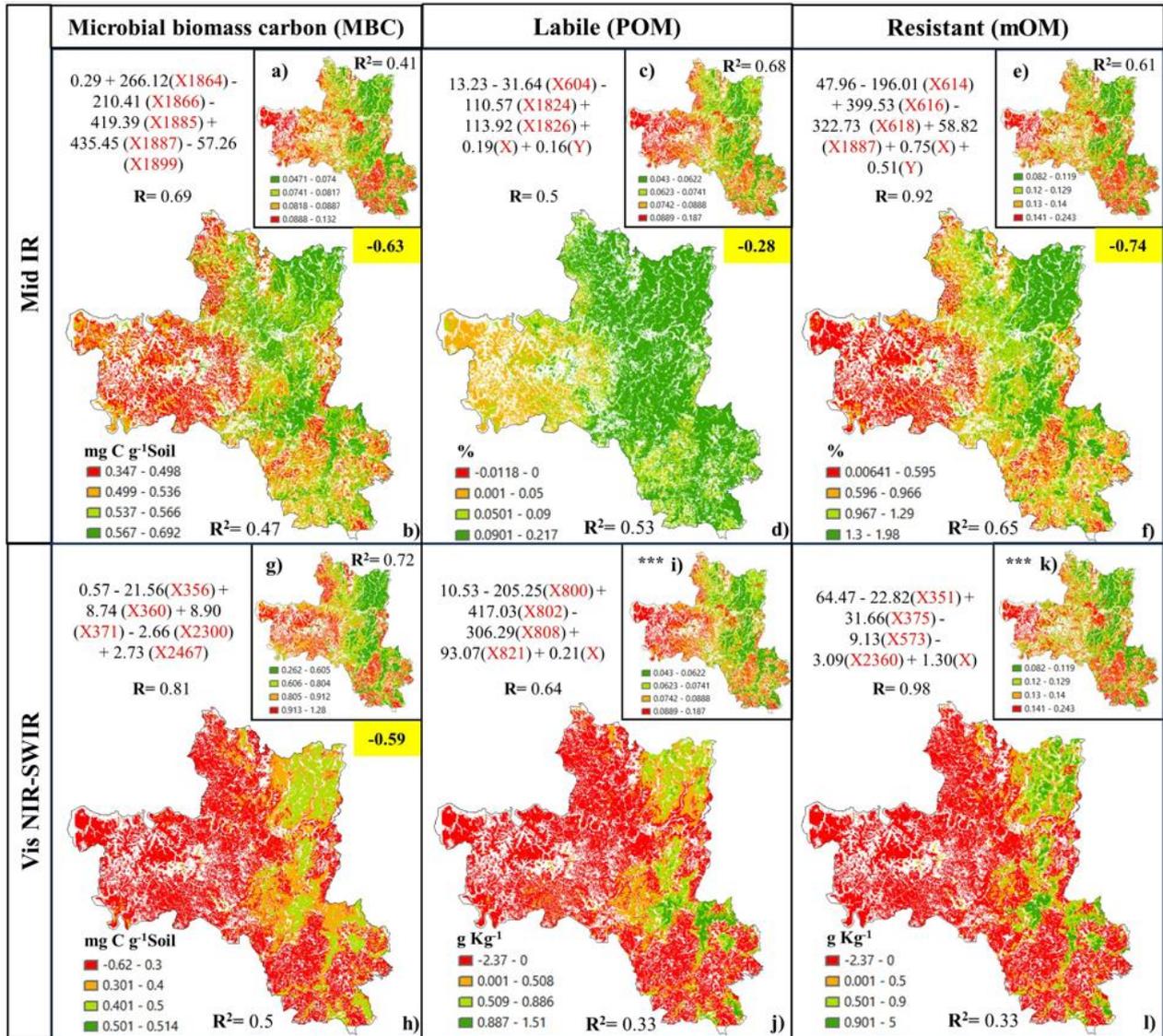


Figure 6. Spatial predictions based on Mid-IR and Vis-NIR-SWIR reflectance of the activity index determined in method 1 for microbial biomass carbon (MBC) (a, g), labile C compounds (c, i), and resistant C compounds (e, k). Values inside the yellow boxes correspond to Spearman correlations between laboratory analyses and the activity index. Labile compounds were compared to particulate organic matter (POM) and resistant compounds to mineral-associated organic matter (MAOM). Spatial predictions of method 2 quantification results for MBC (b, h), POM (d, j), MAOM (f, l). The quantification equations for each variable can be seen, the numbers in red indicate the specific band, X and Y corresponding to longitude and latitude. R is the correlation value between the observed value and the one predicted by the quantification equation. R^2 is coefficient of determination of the spatial prediction.

3.3.2. Method 2

The results of the different parameters analyzed in the various spatial regression models tested are shown in the appendix tables. The data were fitted to different models with high R

values that were not far from the fits associated with the multiple linear regression model considering location. Therefore, due to the ease of replicating the proposed methodology, we decided to work with this model for the prediction of microbiological activity. The specific Mid-IR and Vis-NIR-SWIR bands obtained with the RFE technique and those of greatest importance in the prediction of each of the variables analyzed are shown as predictors in red in the regression equations in Figures 5 and 6.

For beta-glucosidase most of the bands (1310, 1581, 1826, 1856 cm^{-1}) were in the ranges reported by Rasche et al. (2013) for the Mid-IR spectrum (Figure 5b), and an R of 0.79 was obtained. The adjusted map shows a similar distribution to that of the initial method, with more high and medium values. In contrast, for Vis-NIR-SWIR (Figure 5h) the bands obtained (350, 360, 361 and 2500 nm) were not within the range reported by Zornosa et al. (2008) (1372-2272 nm). This is reflected by the difference in the distribution of the adjusted map that also varies considerably in relation to the Mid-IR map. In this map there are greater areas with low activities of this enzyme; however, the ranges related to the highest values maintain a similar distribution. Additionally, the R^2 of the map validation for Vis-NIR-SWIR was higher than that obtained for Mid-IR (0.43 and 0.3).

For urease, the bands used in the model (604, 608 and 620 cm^{-1}) were not within the ranges reported by Rasche et al. (2013) for Mid-IR (Figure 5d), although a strong R of 0.92 was obtained. Despite not showing coincidence between the bands reported in the literature and those obtained in the model, a similar distribution was observed with the adjusted map, with few low values and smaller areas with moderated values. Similarly, the bands associated with the prediction model for Vis-NIR-SWIR (350, 410, 2499 and 2148 nm) were not in the ranges reported by Zornoza et al. (2008); however, the distribution of the adjusted map (Figure 5j) was similar to the one elaborated with the first methodology, showing some areas with low values in the northwestern and southwestern localities and larger areas in green color, associated with high values of this enzymatic activity.

For acid phosphatase, an initial map with the Mid-IR spectrum was not constructed, because no studies were found relating this enzymatic activity to the indicated range (Table 1). We observed that the bands with greater importance in the prediction of this enzyme corresponded to 610, 614, 622, 624, 689, 1083 and 2660 cm^{-1} , from which we elaborated a prediction equation with an R of 0.8. This map showed an R^2 of 0.49 in the validation (Figure 5f), and it is clear that the high values of this activity are concentrated in the areas with high iron oxide contents (Figure 8). The most important bands associated with the Vis-NIR-SWIR spectrum correspond to 356, 357, 667 and 1911 nm, of which only the last one was within the

ranges reported by Zornoza et al. (2008). We obtained an R^2 for validation equal to that obtained with the Mid-IR spectrum; however, we saw strong differences in the maps and the predicted values, because higher data were obtained with Mid-IR, reflecting wider areas with low or no enzyme activity (areas in red) with the prediction based on the Vis-NIR-SWIR spectrum (Figure 5l). However, if the maps of the initial methodology are compared, a similar distribution was observed.

For the MBC, the most important bands of the Mid-IR spectrum for the prediction of this variable (1864, 1866, 1885, 1887, 1899 cm^{-1}) are within the ranges reported by Rasche et al. (2013), from which we obtained a prediction equation with an R correlation of 0.69. The quantification map of this variable had an R^2 of 0.47 (Figure 6b) and a similar distribution was obtained for the map with the first method, with a reduction of the areas with low contents, especially in the southern zone. In relation to the Vis-NIR-SWIR spectrum, the most important bands in the prediction correspond to 356, 360, 371, 2300 and 2467 nm, of which only the last two were within the ranges reported by Zornoza et al. (2008). We obtained an R correlation of 0.81 for the prediction equation and a map with a validation R^2 of 0.5, with a distribution like that obtained in the first methodology (Figure 6h). However, few areas with high contents were observed that were strongly related to the areas with higher iron oxide contents (Figure 8).

The most important bands of the Mid-IR spectrum for POM prediction did not match to those reported by Parikh et al. (2014). Although, they obtained a prediction equation with an R correlation of 0.5 and a map with an R^2 in the validation of 0.53 that presented strong differences in relation to the map obtained with the first method (Figure 6d), given that areas with low contents were not observed and areas with high percentage of particulate matter were mostly observed. For the Vis-NIR-SWIR spectrum, a map of this variable was not constructed using the first method, because no specific bands for these forms of organic matter were reported in this spectrum. The most important bands corresponded to 800, 802, 808 and 821nm, from which we adjusted a prediction equation with an R correlation of 0.64 and obtained a map with R^2 of 0.33. This showed a similar distribution on the map obtained with the Mid-IR spectrum with the first methodology, only with a reduction of the zones with high values (Figure 6j). This was clearly one of the most difficult variables to map.

The MAOM prediction was related to the Mid-IR spectrum bands 614, 616, 618 and 1887cm^{-1} that did not coincide with those reported by Parikh et al. (2014), but from these, a prediction equation was obtained with an R correlation of 0.92 and a map with R^2 of 0.65, whose distribution was similar to that of the map developed with the first method (Figure 6f).

No map was made with the first method using the Vis-NIR-SWIR, because there were no studies that report specific ranges or bands for this variable (Table 1). The most important bands in the prediction correspond to 351, 375, 573 and 2360 nm, from which a prediction equation was adjusted with an R correlation of 0.98 and a map with an R^2 of 0.33 that had a distribution similar to that of the Mid-IR map of the first methodology, while the higher values were restricted to a smaller area (Figure 61).

3.3.3. Predictive models for C sequestration potential

The prediction equation for C sequestration potential proposed by Rodríguez-Albarracín et al. (2023) for the 0 to 20 cm depth was adjusted by including microbiological variables. Thus, we obtained an equation for Mid-IR and a Vis-NIR-SWIR (Table 2).

Table 2. Fitted spatial regression models and their related statistics for C sequestration potential (Sat-def). λ = autoregressive parameters, ρ = spatial autocorrelation coefficient, r = correlation coefficient, MIT = Moran index test.

Model	λ	R	AIC	r	NT	MIT	variables
Mid-IR Depth 0 to 20 cm							
PAR	0.837 (1.78E-06)		2024.22	0.33	1.81E-08	5.39E-02	Sat-def
SEM	0.816 (4.95E-06)		1931.259	0.59	8.35E-12	1.80E-01	B-g + U + AKt + AGb + AHm + F + Acti
SLM		0.751 (3.20E-05)	1930.69	0.69	1.86E-12	8.61E-02	B-g + U + AKt + AGb + AHm + F + Acti
SARAR	0.526 (3.43E-01)	0.519 (3.26E-01)	1929.71	0.66	1.39E-12	0.463	B-g + U + AKt + AGt + AHm + F + Acti
SLMA		0.792 (1.86E-05)	1941.47	0.66	2.84E-12	1.64E-01	B-g + MAOM + AKt + AGb + AHm + F + Acti
Multiple linear regression				0.59	1.70E-11	1.30E-05	B-g + U + AKt + AGt + AHm + F + Acti + X
Vis NIR SWIR Depth 0 to 20 cm							
PAR	0.986 ($< 2.220e-16$)		6775.45	0.50	6.80E-10	8.78E-12	Sat-def
SEM	0.976 ($< 2.220e-16$)		6533.76	0.61	6.63E-12	3.07E-06	B + F + AKt + AGt + AHm + B-g + P + POM
SLM		0.928 ($< 2.220e-16$)	6550.45	0.59	3.54E-13	4.08E-09	B + F + AKt + AGt + AHm + B-g + P + POM
SARAR	0.892 ($< 2.220e-16$)	0.585 (8.35E-04)	6520.93	0.61	1.45E-12	0.591	F + VAM + AKt + AGt + AHm + B-g + P + POM
SLMA		0.940 ($< 2.220e-16$)	6517.21	0.62	1.02E-12	3.44E-03	B + AKt + AGt + AHm + B-g + P + POM
Multiple linear regression				0.53	5.08E-12	0	B + F + VAM + AGb + AGt + AHm + U + POM + MAOM + X + Y

Where A = the amplitude of the different minerals AKln (kaolinite), AGt (goethite), AHem (hematite), and AGbs (gibbsite) from Mendes et al. (2020), cited by Rodriguez-Albarracin et al. (2023) and enzymatic activity of beta-glucosidase (B-g), urease (U), and phosphatase (P) and activity indexes of bacteria (B), fungi (F), actinomycetes (Acti), and vesicular arbuscular mycorrhizal fungi (VAM). The models selected as the best ones are shown in gray.

According to Table 2 we observe that all models for Mid-IR meet the assumptions of independence of residuals (MIT > 0.05); therefore, the criterion for the selection of the best model was based on the highest value of r , associated with SLM ($r=0.69$). We observed that

for the SLM model of MIR (Table 3) the C sequestration potential (Sat-def) was explained by the enzymatic activity of beta-glucosidase and urease, the activity index of fungi and actinomycetes, and the abundance of kaolinite, gibbsite, and hematite. Minerals had the highest impact and were also the most important variables, followed by the activity index of actinomycetes. Enzymatic activity had low impact and low importance (Figure 7).

Table 3. Parameters of the SLM spatial regression model for Mid-IR and SARAR model for Vis-NIR-SWIR for C sequestration potential (Sat-def)

Coefficients	Estimate	Asymptotic Std. Error	z value	Pr(> z)	Impact Direct	Impact p-value
Mid-IR Depth 0 to 20 cm						
(Intercept)	2.21	1.676	1.321	0.186		
Beta-glucosidase	-0.21	0.026	-8.110	4.44E-16	-0.217	4.15E-14
Urease	0.27	0.089	2.963	0.003	0.270	0.00162
AKt	642.11	298.187	2.153	0.031	654.972	3.66E-02
AGb	-2150.77	1748.846	-1.230	0.219	-2193.856	2.25E-01
AHm	340.73	156.824	2.173	0.030	347.552	0.02764
Fungi	-81.88	32.360	-2.530	0.011	-83.523	0.01771
Actinomycetes	90.20	38.759	2.327	0.020	92.002	0.02908
Vis NIR SWIR Depth 0 to 20 cm						
(Intercept)	1.42	1.427	0.995	0.3196		
Fungi	27.82	8.436	3.297	0.0010	27.956	0.00077
VAM Fungi	-16.28	4.609	-3.532	0.0004	-16.362	0.00031
AKt	511.14	169.000	3.025	0.0025	513.690	0.00192
AGt	-169.49	41.327	-4.101	4.11E-05	-170.337	0.00013
AHm	740.10	89.606	8.259	2.22E-16	743.783	4.44E-16
Beta-glucosidase	0.05	0.021	2.427	0.015	0.051	0.01185
Phosphatase	0.00	0.002	-2.013	0.044	-0.004	0.03827
POM	-1.37	0.183	-7.464	8.42E-14	-1.374	2.41E-13

In relation to the Vis-NIR-SWIR spectra, we observed that only the SARAR model satisfied the assumption of independence of the residuals (MIT= 0.59) (Table 2), presenting the lowest AIC value (6520) and a high r (0.61). Because of that, it was chosen as the best fitted model. The enzymatic activity of beta-glucosidase and acid phosphatase, activity index of fungi, VAM fungi, POM and the abundance of kaolinite, hematite and goethite explained the Sat-def. The minerals had the highest impacts, followed by the activity index of fungi (Table 3). The most important variables corresponded to hematite and kaolinite, followed by the activity index of fungi, goethite, and VAM fungi, while the lowest impacts and the least importance were related to enzymatic activity and POM (Figure 7).

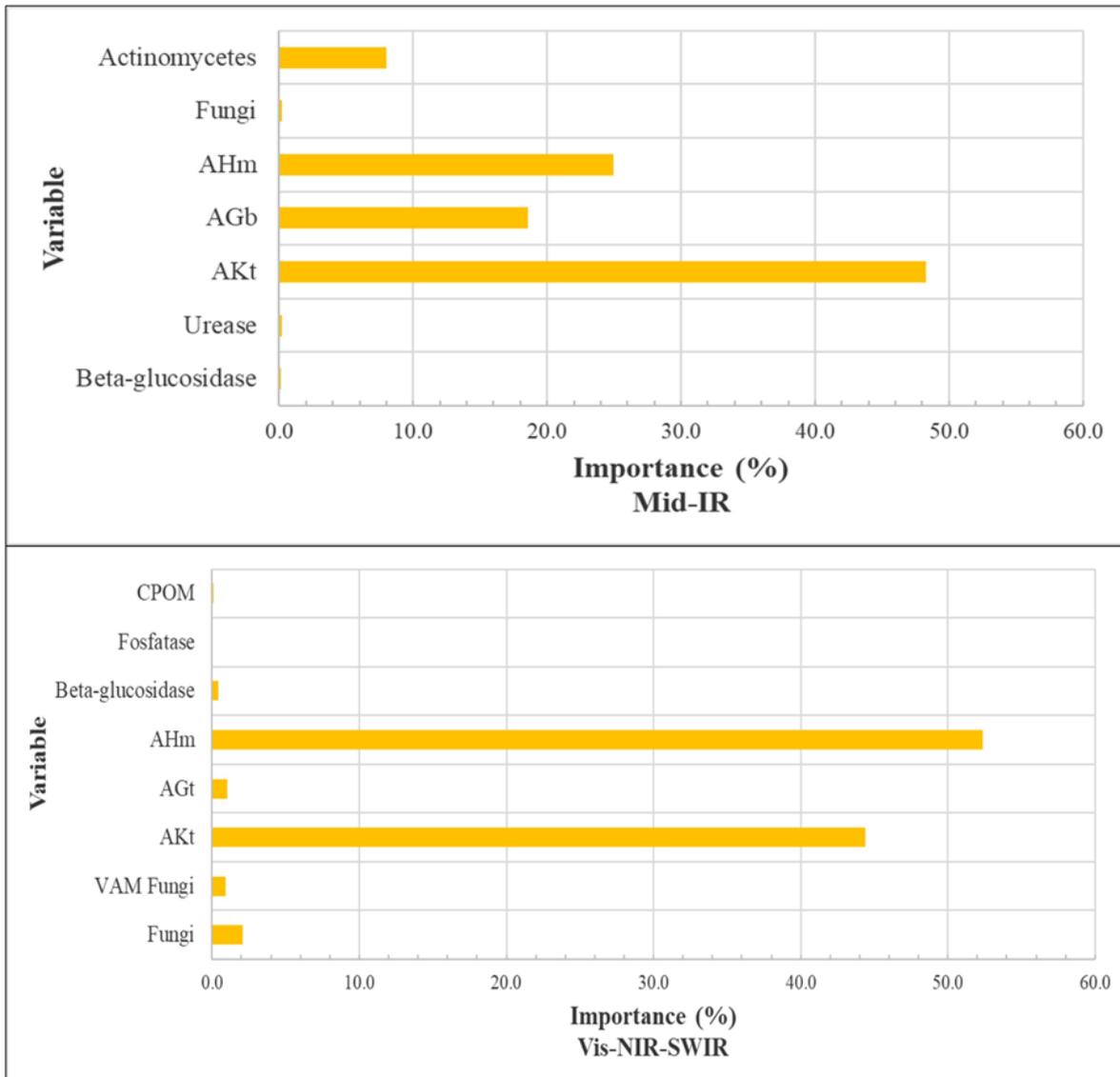


Figure 7. Importance of the explanatory variables of the spatial regression models SLM for Mid-IR and SARAR for Vis-NIR-SWIR for the prediction of C sequestration potential (sat-def) at a depth of 0 cm to 20 cm.

3.4. Discussion

3.4.1. Method 1

Our study performed the prediction and spatialization of the activity index for biological variables such as MBC, soil enzymes, and different microbial groups by RF algorithm, based on Mid-IR reflectance and using environmental covariates such as SYSI, topographical relief, and climate. We obtained an R^2 for MBC of 0.41 (Figure 6a), which is lower than that reported by other studies (Soriano-Disla et al., 2014 [$R^2=0.82$]; Ludwig et al., 2008 [$R^2=0.84$]; Rasche et al., 2013 [$R^2=0.92$]; Nath et al., 2021 [$R^2=0.84$]), when we made predictions using PLS models. The R^2 obtained from the activity index of microbial groups

(Figure 4) such as bacteria ($R^2=0.67$), fungi ($R^2=0.63$), actinomycetes ($R^2=0.62$) and acidobacteria ($R^2=0.69$) are close to Nath et al. (2021) for identification of microbial groups ($R^2=0.75$). For beta-glucosidase and urease activity, we obtained an R^2 of 0.54 and 0.65 (Figure 5 a, c). These values were lower than Rasche et al. (2013) for beta-glucosidase ($R^2=0.89$) and urease ($R^2=0.91$) and lower than the R^2 value of Mimmo et al. (2002) for the prediction of enzymatic activity through the Mid-IR spectrum ($R^2=0.7$).

For the predictions based on Vis-NIR-SWIR spectra we obtained an R^2 of 0.72 for MBC (Figure 6g), 0.73 for beta-glucosidase (Figure 5g), 0.71 for urease (Figure 5i), 0.65 for acid phosphatase (Figure 5k), 0.77 for the activity index of bacteria (Figure 4e), 0.80 for fungi (Figure 4f), 0.85 for actinomycetes, and 0.78 for VAM fungi (Figure 4g and h). These values were similar to those obtained by Zornoza et al. (2008), whose study was the basis for the selection of spectral ranges. They report an R^2 of 0.76 for MBC, 0.88 for phosphatase, 0.55 for beta-glucosidase, and 0.46 for urease. Additionally, they report R^2 based on PLS regression of 0.93 for bacteria, 0.77 for fungi, 0.92 for actinobacteria, and 0.91 for VAM. Rinnan and Rinnan (2007) perform NIR-based predictions using PLS. They obtain R^2 values of 0.78 for fungi and 0.72 for bacterial activity, while Soriano-Disla et al. (2014) report R^2 of 0.93 for MBC.

In general, the spatial predictions of the different microbiological activity indices obtained R^2 values mostly higher than those of Rodriguez-Albarracín et al. (2023) in the mapping of variables such as clay ($R^2= 0.7$) and SOC ($R^2=0.7$), in the study area of this research, using the DSM methodology for the mapping.

3.4.2. Methodology 2

The second method allowed identifying specific bands related to microbiological activity for Vis-NIR-SWIR and Mid-IR ranges through the RFE algorithm, from which we obtained MLR equations for the prediction of each variable (Figures 5 and 6). The R^2 of the spatial prediction determined by RF for some variables predicted with the Mid-IR spectrum were low, as observed for beta-glucosidase (0.3) and urease (0.42) (Figure 5 b and d); however, the R of the prediction equations were 0.79 and 0.92, showing a considerable improvement of these metrics compared to the first method, approaching the values reported by Rasche et al. (2013).

For beta-glucosidase the bands involved in the MLR prediction model for the Mid-IR spectral range corresponded to 1310, 1581, 1826 and 1856 cm^{-1} (Figure 5b). These are within the range reported by Rasche et al. (2013) (2301-1279 cm^{-1}). For the 1310 cm^{-1} band there are

no studies reporting a specific activity. Parikh et al. (2014) cite the range 1330 to 1315 cm^{-1} in which the influence of the C=O ester group is highlighted. In the 1315 cm^{-1} sulfone band, in 1300-1340 diamond C=C and in 1300 cm^{-1} the CH overtone, is also reported from 1379-1384 cm^{-1} by Liu et al. (2015). Gao and Chorover (2009) report the influence of Amide III (NH, CN, CH, NH) at 1337-1313 cm^{-1} , and Nicolas et al. (2017) report the general activity of aliphatic groups in the range 1350-1450 cm^{-1} . In relation to the 1581 cm^{-1} band (Figure 5b), aromatic groups are reported at 1510 cm^{-1} (Nicolás et al 2017); and specifically aromatic C=C at 1630-1580 cm^{-1} (Ladd et al., 1993), who also point out the influence of amide II at 1550 and lignin at 1500 cm^{-1} . Parikh et al. (2014) point out the influence of amide II (NH C=N) from 1590-1500 cm^{-1} , aromatic CH deformation at 1570 cm^{-1} , aromatic C=C stretching from 1550-1500 cm^{-1} , and 1600-1580 cm^{-1} . The influence of amide II is also reported by Cao et al. (2011) from 1540-1549 cm^{-1} and Omoike and Chorover (2006) from 1544-1516 cm^{-1} . According to Cao et al. (2011) and Mc Whirter et al. (2002), the amide II band shift is due to the interaction between iron oxides, such as goethite and the abundance of surface adsorbed bacteria observed in the 1550 cm^{-1} band. The bands 1826 and 1856 cm^{-1} (Figure 5b) can be related to the carbonyl influence reported at 1620-1850 cm^{-1} by Liu et al. (2015). Demyan et al. (2012), note that the peaks 1980 to 1870 cm^{-1} and 1792 cm^{-1} are related to SiO bonds of quartz minerals and carbonyl stretching.

For urease, the bands involved in the prediction model 604, 608 and 620 cm^{-1} (Figure 5d), are within the CH₂ stretching vibrational range of bacterial cellular compounds of 650-450 cm^{-1} (Parikh et al., 2014). Ludwig et al. (2008) indicate that the broad range of 1057-397 cm^{-1} is related to nitrogenous compounds. The range 690-670 cm^{-1} is associated with clay and quartz minerals (Demyan et al., 2012). In the 696 cm^{-1} peak there is an association of the interaction of amide IV from bacterial cells with the AlO groups of clay minerals (Deo et al., 2001).

The bands involved for acid phosphatase prediction 610, 614, 622, 624 and 689 cm^{-1} (Figure 5f) can be related to some bacterial compounds (Parikh et al., 2014) and the interaction of these with clay minerals (Deo et al., 2001). The 1083 cm^{-1} band can be associated with the stretching of P=O groups of polyphosphate products, nucleic acid phosphodiester, and phosphorylated proteins recognized in their study in the 1084 cm^{-1} band (Ojeda et al., 2008). Gao and Chorover (2009) report COC, C-C, and PO₃²⁻ vibrations at 1080 cm^{-1} . Liu et al. (2015) mention the influence of P-O=R features in the range 1000-1100 cm^{-1} and C-O stretching in the range 1006-1122 cm^{-1} . Lad et al. (1993) indicate the influence of carbohydrates in the 1050 cm^{-1} band. Parikh et al. (2014) point out that the range 1000-1080

cm^{-1} is an indicative of POM cellulose, the range 1160-1020 cm^{-1} reflects the influence of polysaccharides such as cellulose lignin or pectin, and CO and CN stretches in the band 1085 cm^{-1} . Also, the influence of POFe bonds in the bands 1037, 1045 and 1027 cm^{-1} are a result of the interaction between the PO groups of the bacterial walls and the surfaces of metal oxides, specifically the PO reported at 1075-1028 cm^{-1} and the phosphorus of humic fractions at 1100-1000 cm^{-1} (Parihn et al. 2014). Cao et al. (2011) report on the 1031 cm^{-1} band the symmetric stretching of PO and at 1051 cm^{-1} of nucleic acids and the C-OH stretch of phosphorylated proteins, which are displaced at 1049 and 1092 cm^{-1} when bound to goethite. The 2660 cm^{-1} band is also involved in the prediction of acid phosphatase (Figure 5f) and can be related to the 2600-2500 cm^{-1} range associated with the H-bridges of OH bonds of carboxylic acids (Parikh et al., 2014). A close range of 2640-2399 cm^{-1} is reported for enzymatic activity in general (Rasche et al. 2013).

The Mid-IR spectra involved in the prediction of MBC corresponded to 1864, 1866, 1885, 1887 and 1899 cm^{-1} (Figure 6b) and are contained within the broad 1959-939 cm^{-1} range reported by Rasche et al. (2013). This range is also related to the interaction of SiO bonds of quartz minerals and carbonyl stretching identified in the 1980-1870 cm^{-1} range (Demyan et al. 2012). Ammann & Brandl (2011) note that cellular carbohydrate and protein compounds associated with microorganisms are distinguished in the 650 cm^{-1} and 1800 cm^{-1} bands.

The prediction of POM through Mid-IR involved the bands 604, 1824 and 1826 cm^{-1} (Figure 6d) reflects the interaction between bacterial compounds and clay minerals, quartz minerals, and organic compounds with the presence of carbonyl groups (Deo et al., 2001; Demyan et al., 2012). They also relate to nitrogenous compounds (Ludwig et al., 2008) and these compounds are required by fungi in different forms such as nitrates, amines, amides (Nicolás et al., 2017). Furthermore, these bands are related to carbohydrates and proteins from microorganisms (Ammann and Brandl, 2011). Resistant organic matter MAOM also involves bands within these ranges associated with 614, 616, 618 and 1887 cm^{-1} (Figure 6f). The differentiation of POM and MAOM in the Mid-IR spectrum is difficult. There are multiple molecular bonds in the same Mid-IR frequency range that may limit specific assignments; however, the prediction of resistant materials has higher statistics metrics (Figure 6f) (Parikh et al., 2014).

In contrast, in the Vis-NIR-SWIR range the differentiation of the spectrum bands for POM and MAOM was clearer (Figure 6 j and l). The POM involved the 800, 802, 808 and 821 nm bands. NH, CH, and CO ratios have been reported in the 820 peak (Stenberg et al., 2010) and in a 825 nm response of aromatic compounds (Viscarra-Rossel and Behrens, 2010). The

MAOM involved the bands 351, 375, 573 and 2360 nm (Figure 6l). Sharma et al. (2021) report the interaction of fulvic acids in the peaks 355 nm and 395 nm and strongly humified material in the range 600-670 nm. Viscarra-Rossel and Behrens (2010) report the interaction of carbonates in 2336 nm, carbohydrates in 2381 nm and methyl functional groups in the range 2307-2469 nm.

The Vis NIR-SWIR spectral range for beta-glucosidase prediction in the bands 350, 360, 361 and 2500 nm were involved. Zornoza et al. (2008) point out an adequate range of 1372-2272 nm. Yang et al. (2022) report the interaction of the methyl CH group at 2450 nm and 2440 nm, and the interaction of fungi with iron oxides at 390 nm.

The prediction of urease involved the 350, 410, 2148 and 2499 nm bands (Figure 5j). Viscarra-Rossel and Behrens (2010) indicate at 2137 nm as the response of polysaccharides, and Fidêncio et al. (2002) and Cozzolino and Morón (2003) report NH of amine and phenolic COH at 2200 nm. Yang et al. (2022) report the interaction of clay minerals with fungi in the 2140 and 2150 nm bands and the interaction of fungi and iron oxides at 410 nm.

For acid phosphatase there were the bands 356, 357, 667 and 1911 nm (Figure 5l) that also involved the response of strongly humified material (600-670 nm) reported by Sharma et al. (2021). Viscarra-Rossel and Behrens (2010) indicate the influence of carboxylic acid at 1930 nm and polysaccharides at 1961 nm. The 191 nm band is in the range 1732-2092 nm (Zornoza et al., 2008).

MBC related bands 356, 360, 371, 2300 and 2467 nm (Figure 6h) are involved in enzymatic activity and are associated with fungal interaction with clay minerals and iron oxides (Yang et al. 2022). CH bonds and fulvic acids are also recognized in these bands (Viscarra-Rossel and Behrens, 2010). Additionally, a relationship is known between the ranges for MBC (2270-2630 nm) and the ranges reported for bacteria, fungi, actinobacteria and VAM fungi (Zornoza et al., 2008).

Soriano-Disla et al. (2014) point out a clear differentiation of N and C in the Mid-IR spectrum that favors strong correlations of this spectral region with soil enzymatic activity. We verified in Vis-NIR-SWIR and Mid-IR enzymatic activity and differentiation of C forms (Figures 5 and 6). The quality of organic matter responds to the extracellular enzymatic activity of fungi and bacteria (De Beeck et al., 2021).

Few works try to predict microbial composition through reflectance spectroscopy (Zornoza et al., 2008). These studies are based on the prediction of other soil parameters related to biological indicators (Soriano-Disla et al., 2014), such as Fe oxides, that modify microbial community composition providing specific niche conditions product of redox

processes (Whitman et al. 2018, Chen et al. 2020). Goethite and hematite are the most common iron oxides in Brazilian soils (Schaefer et al., 2008; Carvalho Filho et al., 2015). In general, the areas with higher amounts of kaolinite, goethite, hematite, and gibbsite were those with higher relative concentrations based on the activity index of fungi, actinomycetes, acidobacteria, MBC beta-glucosidase, urease, and acid phosphatase activity (Figures 4, 5 and 6). POM and MAOM show a close relationship; however, differences in predictions with each sensor are clear (Figure 6). In Vis-NIR-SWIR these higher abundances seem to be restricted to higher contents of iron oxides. On the other hand, for Mid-IR the highest concentrations of POM were found in locations with higher contents of clay minerals and Fe and Al oxides.

Several organic particles exhibit a net negative charge at ambient pH values (Rijnaarts et al., 1995), that varies in magnitude with pH due to ionization of surface functional groups and extracellular polymeric substances (EPS) (Omoike and Chorover 2004). EPS functional groups are mainly protonated at $\text{pH} < 2.0$ and become negatively charged as pH increases due to the dissociation of protons in their functional groups (Cao et al., 2011). Evidently, the pH of the study area favors negative charges on EPS functional groups (Figure 8) that will bind with positively charged mineral surfaces, like Fe and Al oxides (Parikh et al., 2014). Carboxyl groups can form inner and outer sphere complexes with oxide surfaces, and the pH of the study area favors inner sphere complexes (Gao et al., 2009; Parikh et al., 2014), resulting in the stabilization of these C forms (Figure 6).

According to Puissant et al. (2019), the acid phosphatase is an isoenzyme that can report higher activity at pH 5 that does not correspond to what was found in the present study, since the areas where the $\text{pH} > 5$ had lower activity of this enzyme (Figure 5 f and l). Nitrogen compounds associated with EPS are preferentially adsorbed on clay minerals and phosphate compounds preferentially on goethite (Cao et al., 2011). In our study, the highest phosphatase activity was concentrated in areas with higher goethite and gibbsite content (Figure 5 and 8). Urease includes more areas with higher activity that seem to accompany the high and intermediate concentrations of kaolinite (Figure 5 d and j, Figure 8).

The measurement of enzyme activity allows predicting global microbial activity (Arias et al., 2005). Soil microbial properties are sensitive indicators of soil quality and are important in the evaluation of management practices (Zornoza et al., 2008). Microbiological variables such as enzyme activity, MBC and microbial groups are significantly related to soil safety (Nath et al., 2021). Therefore, prediction equations based on Vis-NIR-SWIR and Mid-IR spectral information (Figures 5 and 6) can contribute to soil health monitoring.

3.4.3. Prediction of C sequestration potential

Mid-IR and Vis-NIR-SWIR prediction models of C sequestration potential (Sat-def) show that the main contribution comes from soil mineralogy (Table 3). The crystallinity of iron oxides in Brazilian soils is low (Schaefer et al., 2008), and this stabilizes SOM more effectively (Duiker et al., 2003). The importance of minerals in the C sequestration potential, highlights the major contribution of kaolinite and hematite without ignoring the importance of goethite and gibbsite in the stabilization of this new C (Rodríguez-Albarracín et al., 2023). In our research, the high positive impacts of kaolinite and hematite and the high negative impacts of gibbsite and goethite were confirmed (Table 3). Clearly, mineralogy not only conditions OC stabilization but it also conditions microbiological activity. For example, Fe oxides such as goethite can modify the composition of the microbial community through oxido-reduction processes that provide specific niche conditions (Jeevani et al., 2021). The above could explain the relatively lower contribution of microbial activity compared to that of clay minerals (Table 3, Figure 7) in C sequestration potential. For Rinnan and Rinnan (2007) the general concentrations of microbial compounds in soil are low but are directly related to the quantity and quality of OM, given the relationship of the products of microbial activity with organic functional groups. Such functional groups interact with soil oxides through electron transfer (Chen et al., 2020) from which metabolic energy is generated for microorganisms (Omoike et al., 2004).

Parikh et al. (2014) report different studies that have used spectroscopy to evaluate the interaction between bacteria and biomolecules with mineral surfaces in which the importance of hematite, kaolinite, and goethite are evident. For goethite, the interactions with EPS were highlighted with kaolinite and hematite having a greater relationship with bacteria. The contribution of bacteria in C stabilization is reported by Rong et al. (2010), while, in our study no influence of bacteria on new C sequestration potential was seen (Table 3). This occurred possibly due to a higher dominance of fungi that may reduce bacterial diversity (Davinic et al., 2012). Therefore, it is possible that there were dominant contributions of fungal exoenzymes, since as Puissant et al. (2019) report, beta-glucosidase activity in soils with pH > 5 (Figure 8) is related to a higher abundance of actinomycetes that may explain the positive impacts that this variable showed in the Mid-IR model (Table 3). Soils with high Fe contents are dominated by actinomycetes (Jeevani et al., 2021).

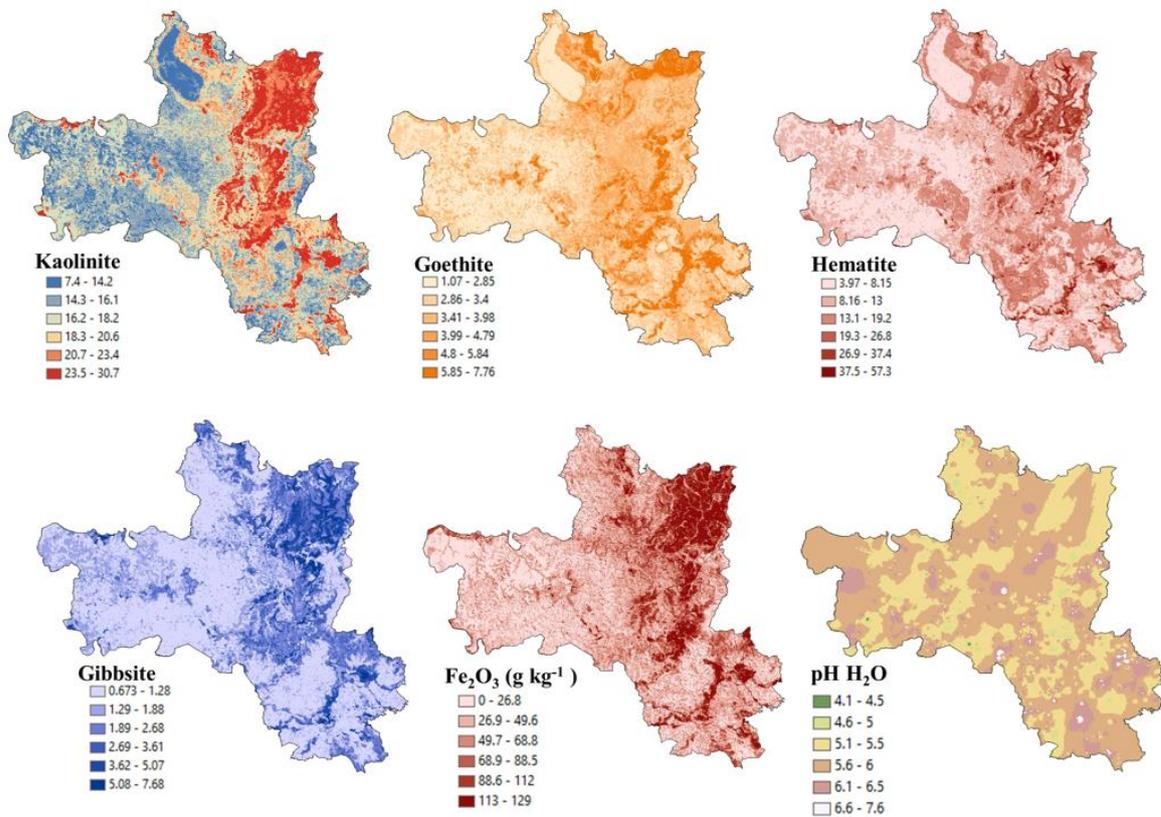


Figure 8. Additional variables: maps of kaolinite, goethite, hematite and gibbsite taken from Mendes et al. (2021), maps of iron oxide (Fe_2O_3) and pH in water taken from Rodríguez-Albarracín et al. (2023)

In general, fungi represent an important component in both prediction models (Table 3, Figure 7). They are the main decomposers of lignin and lignocellulose into SOM, through the secretion of oxidative and hydrolytic enzymes that generate extracellular metabolites that facilitate the decomposition of SOM (Yang et al., 2022). This highlights the importance of enzymatic activity because net and charged organic molecules including enzymes from bacteria and fungi accumulate in these EPSs (De Beeck et al., 2021). In general, enzymatic activity influences SOC depletion and sequestration, and the formation of labile forms of C, especially hydrolytic and oxidative enzymes (Zhang et al., 2020).

The enzymatic activity of beta-glucosidase was important in both models (Table 3). It is found in the matrix of EPS of fungi and bacteria (De Beeck et al., 2021), and acts in the last phase of cellulose degradation (Gil-Sotres et al., 2005), obtaining glucose as a final product. Glucose is an important source of energy C for the growth and activity of soil microorganisms (Merino et al., 2016). In addition, the prediction models variables for this enzyme allowed corroboration that spectroscopically beta-glucosidase is related to MAOM (Figure 5 b and h).

Urease enzymatic activity is part of the nitrogen cycle (Meyer et al., 2015) and spectroscopically presents strong relationships with functional groups with nitrogenous compounds (Figure 5 d and j) that are required by fungi (Nicolás et al., 2017). These microorganisms contribute significantly to the sequestration of new C, according to our research (Table 3).

Acid phosphatase is also involved in C sequestration (Table 3 and Figure 5fl). This enzyme is related to phosphorus compounds like nucleic acids as present in bacterial walls, phosphate groups that interact with the surfaces of metal oxides, and humic fractions (Omoike and Chorover, 2006; Parickh et al., 2014). It is possible that this enzymatic activity involves bacterial contributions to the C sequestration potential that was not observed in the mathematical models.

The influence of VAM fungi on C sequestration potential was also seen (Table 3), with a negative impact indicating that a reduction in the abundance of these organisms results in an increase of sequestration potential. The hyphae of these fungi form a microhabitat (mycosphere) that serves as an important reservoir of carbonaceous compounds and photosynthates accessible to other microorganisms (Halsey et al., 2016). A similar influence is present with POM (Table 3), as this fraction includes forms of C that microorganisms readily decompose (Zhang et al., 2020) and that are important for energy and sources of C and oxygen for respiration (Follett et al., 2001).

Some of the C forms bound to clay fractions may be relatively inert. They do not respond to land use, while others may be more active and respond to land use changes (Demyan et al., 2012). The zones with higher sequestration potential reported by Rodríguez-Albarracín et al. (2023) respond to zones with lower microbial activity and medium enzymatic activity (Figures 4 and 5) that can be activated with land use change. The authors discussed the historical land use map of Tayebi et al. (2021) and underlined that areas with more than 15 years of agriculture have both the lowest and highest C sequestration potential, given that in some areas minerals such as goethite are completely saturated with C and in others, kaolinite and hematite show high C sequestration potential (Figure 8). Land use change causes a variation of organic compounds to reach the soil (Kirsten et al., 2021). These zones with higher sequestration potential (Figure 4, 5 and 6) are key for promoting crop renewal to take advantage of the potential that minerals such as kaolinite and hematite have to retain new C and thus reactivate microbial activity in these zones, given that an increase in SOC stocks would enhance the activity of microorganisms.

3.5. Conclusions

Maps of activity indices related to the relative abundance of bacteria, fungi, actinomycetes, acidobacteria, and vesicular-arbuscular mycorrhizal fungi were obtained, based on information from the Vis-NIR-SWIR and Mid-IR spectral ranges by the digital soil mapping framework. The best results were found for Vis-NIR-SWIR spectrum, which R^2 varied from 0.77 to 0.85.

The method based on literature for the relative determination of the enzyme activity index and MBC from Vis-NIR-SWIR and Mid-IR spectral information was validated, obtaining correlations from -0.55 to -0.7 with the laboratory data. The mapping based on this methodology obtained R^2 values between 0.41 and 0.73.

Equations were obtained for the quantification of enzymatic activity, MBC, and POM and MAOM, for each spectral range, based on multiple linear regression models including as explanatory variables the specific bands for each variable and the X and Y coordinates that presented correlations from 0.64 to 0.98. POM was the variable with the lowest adjustment (0.50). The spatial modeling showed lower R^2 than those obtained with the first methodology; however, for variables such as MBC, an improvement was verified.

The bands involved in the multiple linear prediction models, and therefore the most important for the Mid-IR spectral range for beta-glucosidase corresponded to 1310, 1581, 1826, 1856 cm^{-1} ; for urease they corresponded to 604, 608, 620 cm^{-1} , for Phosphatase 610, 614, 622, 624, 689, 1083, 2660 cm^{-1} , for MBC 1864, 1866, 1885, 1887, 1899 cm^{-1} , for POM 604, 1824, 1826 cm^{-1} , for MAOM 614, 616, 618 and 1887 cm^{-1} . For the Vis NIR-SWIR spectral range for beta-glucosidase they corresponded to 350, 360, 361, 2500 nm, urease 350, 410, 2499, 2148 nm, phosphatase 356, 357, 667, 1911 nm, MBC 356, 360, 371, 2300, 2467 nm, POM 800, 802, 808, 821 nm, and MAOM 351, 375, 573 and 2360 nm.

The C sequestration potential was fitted to a double SARAR autoregressive model, considering the Vis-NIR-SWIR electromagnetic spectrum bands with an r of 0.61 and a spatial error model (SEM) for the Mid-IR spectral range with r of 0.7. The explanatory variables include mineralogy (kaolinite, hematite, goethite, and gibbsite) and microbiology (activity index of fungi, actinomycetes, vesico-arbuscular mycorrhizal fungi, enzymatic activity of beta-glucosidase, urease and acid phosphatase and POM). The greatest importance corresponded to minerals with positive impacts for kaolinite and hematite, and negative impacts for goethite and gibbsite. The microbiological variables with the greatest importance were related to the general abundance of fungi, those of lesser importance were of enzymatic activity, especially that of acid phosphatase.

The influence of land use was identified in C sequestration potential by its relation to the organic molecules that reach the soil and interact with microbiota and minerals. The areas with higher C sequestration potential could be key to climate change mitigation strategies.

References

- ADAMCZYK, B., SIETIÖ, O.M., STRAKOVÁ, P., PROMMER, J., WILD, B., HAGNER, M., PIHLATIE, M., FRITZE, H., RICHTER, A. & J. HEINONSALO. 2019. Plant roots increase both decomposition and stable organic matter formation in boreal forest soil. *NATURE COMMUNICATIONS* (2019) 10:3982. <https://doi.org/10.1038/s41467-019-11993-1>
- ALVARES, C.A., STAPE, J.L., SENTELHAS, P.C., DE MORAES GONÇALVES, J.L. & G. SPAROVEK. 2013. Koppen's climate classification map for Brazil. *Meteorol. Z.* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- AMMANN, A. & H. BRANDL. 2011. Detection and differentiation of bacterial spores in a mineral matrix by Fourier transform infrared spectroscopy (FTIR) and chemometrical data treatment. *BMC Biophysics* 2011, 4:14. <http://www.biomedcentral.com/2046-1682/4/14> (14 July 2011)
- ARIAS, M., GONZÁLEZ-PÉREZ, J., GONZÁLEZ-VILA, F. & A. BALL. 2005. Soil health—a new challenge for microbiologists and chemists. *INTERNATIONAL MICROBIOLOGY* (2005) 8:13-21.
- BREIMAN, L. 2001. Random forests. *Machine Learning*, 45 (2001), pp. 5-32
- BÖHNER, J. & B. BECHTEL. 2018. GIS in climatology and meteorology. In Huang, B. *Comprehensive Geographic Information Systems. Reference Module in Earth Systems and Environmental Sciences*. Elsevier. Pages 196-235.
- BURNS, DA. & E.W. CIURCZAK. 2001. *Handbook of near-infrared analysis*. Second edition. Marcel Dekker; New York: 2001. revised and expanded
- BUCKERIDGE, K., LA ROSA, A., MASON, K., WHITAKER, J., MCNAMARA, N., GRANT, H. & N. OSTLE. 2020. Sticky dead microbes: Rapid abiotic retention of microbial necromass in soil *Soil Biology and Biochemistry* (2020).
- CAO, Y., WEI, X., CAI, P., HUANG, Q., RONG, X. & W. LIANG. 2011. Preferential adsorption of extracellular polymeric substances from bacteria on clay minerals and iron oxide. *Colloids and Surfaces B: Biointerfaces* 83 (2011) 122–127
- CARVALHO FILHO, A., INDA, A.V., FINK, J.R. & N. CURI. 2015. Iron oxides in soils of different lithological origins in ferriferous quadrilateral (Minas Gerais, Brazil) *Appl Clay Sci*, 118 (2015), pp. 1-7, [10.1016/j.clay.2015.08.037](https://doi.org/10.1016/j.clay.2015.08.037)
- CHEN, Q., MENG, Z., LIU, X., JIN, Q. & R. SU. 2018. Decision variants for the automatic determination of optimal feature subset in RF-RFE. *Genes* 9. <https://doi.org/10.3390/genes9060301>
- CHEN, C., HALL, S.J., COWARD, E. & A. THOMPSON. 2020. Iron-mediated organic matter decomposition in humid soils can counteract protection. *Nature Communications* 11, 1–13.

- CHODAK, M. 2011. Near-infrared spectroscopy for rapid estimation of microbial properties in reclaimed mine soils. *Journal of Plant Nutrition and Soil Science* 174, 702-709.
- COHEN, M.J., PRENGER, J.P. & W.F. DEBUSK. 2005. Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. *Journal of Environmental Quality* 34, 1422e1434.
- COTRUFO, M.F., RANALLI, M.G., HADDIX, M.L., SIX, J. & E. LUGATO. 2019. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nature Geoscience*, 12(12), pp.989-994. <https://doi.org/10.1038/s41561-019-0484-6>
- COZZOLINO, D. & A. MORÓN. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *Journal of Agricultural Science*. 2003; 140:65–71.
- DAVINIC, M., FULTZ, L.M., ACOSTA-MARTINEZ, V., CALDERON, F.J., COX, S.B., DOWD, S.E., ALLEN, V.G., ZAK, J.C. & J. MOORE-KUCERA. 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. *Soil Biol. Biochem.* 46, 63–72. <https://doi.org/10.1016/j.soilbio.2011.11.012>
- DE BEECK, M., PERSSON, P. & A. TUNLID. 2021. Fungal extracellular polymeric substance matrices – Highly specialized microenvironments that allow fungi to control soil organic matter decomposition reactions. *Soil Biology and Biochemistry* 159 (2021) 108304. <https://doi.org/10.1016/j.soilbio.2021.108304>
- DE MASTRO, F., TRAVERSA, A., COCOZZA, C., PALLARA, M. & G. BRUNETTI. 2020. Soil Organic Carbon Stabilization: Influence of Tillage on Mineralogical and Chemical Parameters. *Soil Syst.* 2020, 4(3), 58
- DEMATTÊ, J., FONGARO, C., RIZZO, C. & J.L. SAFANELLI. 2018. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.*, 212 (2018), pp. 161-175.
- 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., BELLINASSO, H., FRANCELINO, M.R., SCHAEFER, C.E.G.R., VICENTE, L.E., DOS SANTOS, U.J., DE SÁ BARRETTO SAMPAIO, E. V., MENEZES, R.S.C., DE SOUZA, J.J.L.L., ABRAHÃO, W.A.P., COELHO, R.M., GREGO, C.R., LANI, J.L., FERNANDES, A.R., GONÇALVES, D.A.M., SILVA, S.H.G., DE MENEZES, M.D., CURI, N., COUTO, E.G., DOS ANJOS, L.H.C., CEDDIA, M.B., PINHEIRO, É.F.M., GRUNWALD, S., VASQUES, G.M., MARQUES JÚNIOR, J., DA SILVA, A.J., BARRETO, M.C. D. V., NÓBREGA, G.N., DA SILVA, M.Z., DE SOUZA, S.F., VALLADARES, G.S., VIANA, J.H.M., DA SILVA TERRA, F., HORÁK-TERRA, I., FIORIO, P.R., DA SILVA, R.C., FRADE JÚNIOR, E.F., LIMA, R.H.C., ALBA, J.M.F., DE SOUZA JUNIOR, V.S., BREFIN, M.D.L.M.S., RUIVO, M.D.L.P., FERREIRA, T.O., BRAIT, M.A., CAETANO, N.R., BRINGHENTI, I., DE SOUSA MENDES, W., SAFANELLI, J.L., GUIMARÃES, C.C.B., POPPIEL, R.R., E SOUZA, A.B., QUESADA, C.A. & DO COUTO, H.T.Z. 2019. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma* 354, 113793. <https://doi.org/10.1016/j.geoderma.2019.05.043>
- DEMATTÊ, J.A.M., NOVAIS, J.J., ROSIN, N.A., ROSAS, J.T.F., POPPIEL, R.R., DOTTO, A.C., & A.F.S. PAIVA. 2023. The Brazilian Soil Spectral Library (VIS-NIR-SWIR-MIR) Database: Open Access (V.001) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8092774>

- DEMYAN, M., RASCHE, F., SCHULZ, E., BREULMANN, M., MULLER, T. & G. CADISCH. 2012. Use of specific peaks obtained by diffuse reflectance Fourier transform mid-infrared spectroscopy to study the composition of organic matter in a Haplic Chernozem. *European Journal of Soil Science*, April 2012, 63, 189 – 199. doi: 10.1111/j.1365-2389.2011.01420.x
- DEO, N., NATARAJAN, K. & P. SOMASUNDARAN. 2001. Mechanisms of adhesion of *Paenibacillus polymyxa* onto hematite, corundum and quartz. *Int. J. Miner. Process.* 62 2001 27–39
- DICK, R.P.; BREAKWELL, D.P. & R.F. TURCO. 1996. Soil enzymes activities and biodiversity measurements as integrative microbiological indicators. In: DORAN, J.W.; & JONES, A.J., eds. *Methods for assessing soil quality*. Soil Soc. Am., 1996, p. 247-271. Special publication.
- DI IORIO, E., CIRCELLI, L., LORENZETTI, R., COSTANTINI, E., PERL, S. & C. COLOMBO. 2019. Estimation of andic properties from Vis-NIR diffuse reflectance spectroscopy for volcanic soil classification. *Catena* 182 (2019) 104109
- DOS REIS, C., PINHEIRO, D., CALDAS, J. & C. BAYER. 2014. Carbon sequestration in clay and silt fractions of Brazilian soils under conventional and no-tillage systems. *Soils and Plant Nutrition. Sci. agric. (Piracicaba, Braz.)* 71 (4).
- DUIKER, S. W., RHOTON, F. E., TORRENT, J., SMECK, N. E. & R. LAL. 2003. Iron (hydr)oxide crystallinity effects on soil aggregation. *Soil Sci. Soc. Am. J.* 67, 606–611.
- ELHORST, J.P. 2014. *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer, Heidelberg New York Dordrecht London, <http://www.springer.com/economics/regional+science/book/978-3-642-40339-2>
- ERMIDA, S.L., SOARES, P., MANTAS, V., GÖTTSCHE, F.M., TRIGO, I.F., 2020. Google earth engine open-source code for land surface temperature estimation from the Landsat series. *Remote Sens.* 12. <https://doi.org/10.3390/RS12091471>
- FICK, S.E. & R.J. HIJMANS. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>.
- FIDÊNCIO, PH., POPPI, RJ., ANDRADE, JC. & H. CANTARELLA. 2002. Determination of organic matter in soil using nearinfrared spectroscopy and partial least squares regression. *Communications in Soil Science and Plant Analysis*. 2002; 33:1607–1615.
- FOLLETT, R., KIMBLE, J. & R. LAL. 2001. The potential of US grazing lands to sequester soil carbon. Lewis Publishers, Boca Raton, Florida (2001), pp. 401-430
- GAO, X., METGE, D.W., RAY, C., HARVEY, R.W. & J. CHOROVER. 2009. Surface complexation of carboxylate adheres *Cryptosporidium parvum* oocysts to the hematite-water interface. *Environ. Sci. Technol.* 43, 7423–7429.
- GAO, X. & J. CHOROVER. 2009. In-situ monitoring of *Cryptosporidium parvum* oocyst surface adhesion using ATR-FTIR spectroscopy. *Colloids and Surfaces B: Biointerfaces* 71 (2009) 169–176. doi:10.1016/j.colsurfb.2009.02.003
- GIL-SOTRES, F., TRASAR-CEPEDA, C., LEIRÓS, M. & S. SEOANE. 2005. Different approaches to evaluating soil quality using biochemical properties. *Soil Biology and Biochemistry.* 37, 877-887.

- GRESCHUK, L., ARAÚJO, M., RODRÍGUEZ-ALBARRACÍN, H., BELLINASO, H., SILVERO, N., PAIVA, A., POPPIEL, R., ROSIN, N., CAMPOS, L., DALMOLIN, R., BALLESTER, M. & J. DEMATTÊ. 2022. Combining spectral ranges for soil discrimination: A case study in the State of Maranhao – Brazil. *Geoderma Regional* 29 (2022) e00507. <https://doi.org/10.1016/j.geodrs.2022.e00507>
- GOUDA, S., KERRY, R., DAS, G., PARAMITHIOTIS, S., SHIN, H. & J. PATRA. 2018. Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture. *Microbiological Research* 206 (2018) 131–140. <http://dx.doi.org/10.1016/j.micres.2017.08.016>
- GÓMEZ, A., VAN LIER, Q., SILVERO, N., INFORSATO, L., DE MELO, M., RODRÍGUEZ-ALBARRACÍN, H., ROSIN, N., FIM ROSAS, J., RIZZO, R. & J.A.M DEMATTÊ. 2023. Digital mapping of the soil available water capacity: tool for the resilience of agricultural systems to climate change. *Science of the Total Environment* 882 (2023) 163572
- GUYON, I., WESTON, J., BARNHILL, S. & V. VAPNIK. 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. <https://doi.org/10.1023/A:1012487302797>
- HALSEY, J., SILVA, M & F. ANDREOTE. 2016. Bacterial selection by mycospheres of Atlantic Rainforest mushrooms. *Antonie van Leeuwenhoek* (2016) 109:1353–1365. doi:10.1007/s10482-016-0734-1)
- HART, M. M., CROSS, A. T., D'AGUI, H. M., DIXON, K. W., VAN DER HEYDE, M., MICKAN, B., HORST, C., GREZ, B. M., VALLIERE, J. M., VISCARRA ROSSEL, R. A., WHITELEY, A., WONG, W. S., ZHONG, H. & P. NEVILL. 2020. Examining assumptions of soil microbial ecology in the monitoring of ecological restoration, *Ecol. Solut. Evid.*, 1, e12031, <https://doi.org/10.1002/2688-8319.12031>, 2020
- HASSINK, J. 1994. Effect of soil texture on the size of the microbial biomass and on the amount of c and n mineralized per unit of microbial biomass in dutch grassland soils. *Soil Biol Biochem* 26: 1573-1581
- HEGLER, F., POSTH, N., JIANG, J. & A. KAPPLER. 2008. Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. *FEMS Microbiol Ecol* 66 (2008) 250–260. DOI:10.1111/j.1574-6941.2008.00592.x
- HOGUE, M., WOHLING, T. & W. NOWAK. 2018. A primer for Model Selection: The Decisive Role of Model Complexity. *Water Resources Research*, 54, 1688-1715. <https://doi.org/10.1002/2017WR021902>
- IUSS WORKING GROUP WRB. 2015. World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome, Italy (2015)
- JEEWANI, P., ZWIETEN, L., ZHU, Z., GE, T., GUGGENBERGER, G., LUO, Y. & J. XU. 2021. Abiotic and biotic regulation on carbon mineralization and stabilization in paddy soils along iron oxide gradients. *Soil Biology and Biochemistry* 160 (2021) 108312. <https://doi.org/10.1016/j.soilbio.2021.108312>
- JINDALUANG, W., KHEORUENROMNE, I., SUDDHIPRAKARN, A., SINGH, B. & B. SINGH. 2013. Influence of soil texture and mineralogy on organic matter content and composition in physically separated fractions soils of Thailand. *Geoderma* 195–196 (2013) 207–219. <http://dx.doi.org/10.1016/j.geoderma.2012.12.003>

- KLEBER, M., EUSTERHUES, K., KEILUWEIT, M., MIKUTTA, C., MIKUTTA, R. & P.S. NICO. 2015. Mineral-organic associations: formation, properties, and relevance in soil environments. *Advances in Agronomy* 1–140.
- KIRSTEN, M., MIKUTTA, R., VOGEL, C., THOMPSON, A., MUELLER, C., KIMARO, D., BERGSMA, H., FEGER, K. & K. KALBITZ. 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. *Scientific Reports* | (2021) 11:5076
- KUHN, M. 2021. caret: Classification and Regression Training. R package version 6.0-88. <https://CRAN.R-project.org/package=car>
- LADD, J., FOSTER, R.C. & J.O. SKJEMSTAD. 1993. Soil structure: carbon and nitrogen metabolism. *Geoderma*, 56 (1993) 401-434. Elsevier Science Publishers B.V., Amsterdam. <https://doi.org/10.1016/B978-0-444-81490-6.50034-0>
- LIANG, C., AMELUNG, W., LEHMANN, J. & M. KASTNER. 2019. Quantitative assessment of microbial necromass contribution to soil organic matter. *Glob Change Biol.* 25, 3578–3590.
- LIU, X. & J. CHEN. 2021. Variable Selection for the Spatial Autoregressive Model with Autoregressive Disturbances. *Mathematics* 2021, 9, 1448. <https://doi.org/10.3390/math9121448>
- LIU, Y., LI, O., LI, Y., BAO, J., HU, Z., HAO, D., SONG, D., WANG, Y. & M. YANG. 2015. Detection of Nanoscale Soil Organic Matter by Middle Infrared Spectrum for Forensic Science. *Hindawi Publishing Corporation Journal of Chemistry* Volume 2015, Article ID 189421. <http://dx.doi.org/10.1155/2015/189421>
- LUDWIG, B., NITSCHKE, R., TERHOEVEN-URSELMANS, T., MICHEL, K. & H. FLESSA. 2008. Use of mid-infrared spectroscopy in the diffuse reflectance mode for the prediction of the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science* 171, 384e391.
- MENDEZ, C. 2020. Spatial regression analysis in R. R Studio/RPubs. Available at <https://rpubs.com/quarcs-lab/tutorial-spatial-regression>.
- MENDES, W., DEMATTE, J.A.M., ROSIN, N., TERRA, F., POPPIEL, R., URBINA-SALAZAR, D., BOECHAT, C., SILVA, E., CURI, N., SILVA, S., SANTOS, U. & G. VALLADARES. 2022. The Brazilian soil Mid-infrared Spectral Library: The Power of the Fundamental Range. *Geoderma* 415 (2022) 115776. <https://doi.org/10.1016/j.geoderma.2022.115776>
- MENDES, W., DEMATTÊ, J.A.M., BONFATTI, B., RESENDE, M., CAMPOS, L. & A. DA COSTA, A. 2021. A novel framework to estimate soil mineralogy using soil spectroscopy. *Applied Geochemistry*, Volume 127, 2021, 104909, ISSN 0883-2927, <https://doi.org/10.1016/j.apgeochem.2021.104909>.
- MERINO, C., GODOY, R. & F. MATUS. 2016. Soil enzymes and biological activity at different levels of organic matter stability. *Journal of Soil Science and Plant Nutrition*. 16, 14-30.
- MEYER, A.H., WOOLDRIDGE, J. & J.F. DAMES. 2015. Variation in urease and β -glucosidase activities with soil depth and root density in a, cripp's pink ,/m7 apple orchard under conventional and organic management. *South African Journal of Plant and Soil*. 32, 227-234.

- MINASNY, B. & A. MCBRATNEY. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* 32 (2006) 1378–1388. doi:10.1016/j.cageo.2005.12.009
- MCBRATNEY, A., SANTOS, M. & B. MINASNY. 2003. On digital soil mapping. *Geoderma*, 117 (2003), pp. 3-52
- MCWHIRTER, M.J., MCQUILLAN, A.J. & P.J. BREMER. 2002. Influence of ionic strength and pH on the first 60 min of *Pseudomonas aeruginosa* attachment to ZnSe and to TiO₂ monitored by ATR-IR spectroscopy. *Colloids Surf. B Biointerfaces* 26, 365–372.
- MIMMO, T., REEVES, J.B., MCCARTY, G.W. & G. GALLETTI. 2002. Determination of biological measures by mid-infrared diffuse reflectance spectroscopy in soils within a landscape. *Soil Science* 167, 281e287
- MÜLLER, B., 2015. Experimental interactions between clay minerals and bacteria: a review. *Pedosphere* 25, 799–810.
- NATH, D., LAIK, R., MEENA, V., PRAMANICK, B. & SINGH. 2021. Can mid-infrared (mid-IR) spectroscopy evaluate soil conditions by predicting soil biological properties?. *Soil Security* 4 (2021) 100008. <https://doi.org/10.1016/j.soisec.2021.100008>
- NICOLÁS, C., ALMEIDA, J., ELLSTRÖM, M., BAHR, A., BONE, S., ROSENSTOCK, N., BARGAR, J., TUNLID, A., PERSSON, P. & H. WALLANDER. 2017. Chemical changes in organic matter after fungal colonization in a nitrogen fertilized and unfertilized Norway spruce forest. *Plant Soil* (2017) 419:113–126. DOI 10.1007/s11104-017-3324-8
- NICOLAS, C., MARTINBERTELSEN, T., FLOUDAS, D., BENTZER, J., SMITS, M.M., JOHANSSON, T., TROEIN, C., PERSSON, P. & A. TUNLID. 2019. The soil organic matter decomposition mechanisms in ectomycorrhizal fungi are tuned for liberating soil organic nitrogen. *The ISME Journal* 13, 977–988.
- OJEDA, J.J., ROMERO-GONZALEZ, M.E., POURAN, H.M. & S.A. BANWART. 2008. In situ monitoring of the biofilm formation of *Pseudomonas putida* on hematite using flow-cell ATR-FTIR spectroscopy to investigate the formation of inner-sphere bonds between the bacteria and the mineral. *Mineral. Mag.* 72, 101–106.
- OLIVEIRA, J.B. & PRADO, H. 1989. Carta pedológica de Piracicaba: Escala 1:100.000.
- OMOIKE, A., CHOROVER, J., KWON, K.D. & J.D. KUBICKI. 2004. Adhesion of bacterial exopolymers to α -FeOOH: inner-sphere complexation of phosphodiester groups. *Langmuir* 20, 11108–11114.
- OMOIKE, A. & J. CHOROVER. 2004. Spectroscopic study of extracellular polymeric substances from *Bacillus subtilis*: aqueous chemistry and adsorption effects. *Biomacromolecules* 5, 1219–1230.
- OMOIKE, A. & J. CHOROVER. 2006. Adsorption to goethite of extracellular polymeric substances from *Bacillus subtilis*. *Geochim. Cosmochim. Acta* 70, 827–838.
- PADARIAN, J., MINASNY, B. & A.B. MCBRATNEY. 2020. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* 6, 35–52. <https://doi.org/10.5194/soil-6-35-2020>

- PADARIAN, J., MINASNY, B., MCBRATNEY, A. & P. SMITH. 2022. Soil carbon sequestration potential in global croplands. *PeerJ*, 10, p.e13740. <https://doi.org/10.7717/peerj.13740>
- PARIKH, S., GOYNE, K., MARGENOT, A., MUKOME, F. & F. CALDE-RON. 2014. Soil Chemical Insights Provided through Vibrational Spectros-copy. Publications from USDA-ARS / UNL Faculty. Paper 1471.
- PUISSANT, J., JONES, B., GOODALL, T., MANG, D., BLAUD, A., GWEON, H., MALIK, A., JONES, D., CLARK, I., HIRSCH, P. & R. GRIFFITHS. 2019. The pH optimum of soil exoenzymes adapt to long term changes in soil pH. *Soil Biology and Biochemistry* 138 (2019) 107601. <https://doi.org/10.1016/j.soilbio.2019.107601>
- RASCHE, F., KNAPP, D., KAISER, C., KORANDA, M., KITZLER, B., ZECHMEISTER-BOLTENSTERN, S., RICHTER, A. & A. SESSITSCH. 2011. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *The ISME Journal* 5, 389-402.
- RASCHE, F., MARHAN, S., BERNER, D., KEIL, D., KANDELER, E. & G. CADISCH. 2013. midDRIFTS-based partial least square regression analysis allows predicting microbial biomass, enzyme activities and 16S rRNA gene abundance in soils of temperate grasslands. *Soil Biology & Biochemistry* 57 (2013) 504-512
- RIJNAARTS, H.H.M., NORDE, W., BOUWER, E.J., LYKLEMA, J. & A.J.B. ZEHNDER. 1995. Reversibility and mechanism of bacterial adhesion. *Colloids Surf. B Biointerfaces* 4, 5–22.
- RINNAN, R. & A. RINNAN. 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. *Soil Biol. Biochem.* 39 (7), 1664–1673. <https://doi.org/10.1016/j.soilbio.2007.01.022>.
- RODRÍGUEZ-ALBARRACÍN, H., DEMATTÊ, J., ROSIN, N., DARGHAN, E., SILVERO, N., CERRI, C., MENDES, W. & M. TAYEBI. 2023. Potential of soil minerals to sequester soil organic carbon. *Geoderma* 436 (2023) 116549. <https://doi.org/10.1016/j.geoderma.2023.116549>
- RONG, X., CHEN, W., HUANG, Q., CAI, P. & W. LIANG. 2010. *Pseudomonas putida* adhesion to goethite: studied by equilibrium adsorption, SEM, FTIR and ITC. *Colloids Surf. B Biointerfaces* 80, 79–85.
- SAFANELLI, J.L., POPPIEL, R.R., RUIZ, L.F.C., BONFATTI, B.R., MELLO, F.A.O., RIZZO, R. & J.A.M. DEMATTÊ. 2020. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. *ISPRS Int. J. Geo-Inf.* 2020, 9, 400. DOI: <https://doi.org/10.3390/ijgi9060400>
- SCHAEFER, C., FABRIS, J. D. & J.C. KER. (2008). Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. *Clay Minerals*, 43:137-154, 2008
- SHARMA, V., CHAUHAN, R. & R. KUMAR. 2021. Spectral characteristics of organic soil matter: A comprehensive review. *Microchemical Journal* 171 (2021) 106836. <https://doi.org/10.1016/j.microc.2021.106836>
- SORIANO-DISLA, J.M., JANIK, L.J., VISCARRA-ROSSEL, R.A., MACDONALD, L.M. & M.J. MCLAUGHLIN. 2014. The performance of visible, near, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.* 49, 139–186. doi:10.1080/05704928.2013.811081

- SOUZA, A.B., DEMATTE, J.A.M., MELLO, F.A.O., SALAZAR, D.F.U., MENDES, W.S. & J.L. SAFANELLI. 2020. Ratio of clay spectroscopic indices and its approach on soil morphometry. *Geoderma* 357, 113963.
- STENBERG, B., VISCARRA ROSSEL, R.A., MOUAZEN, A.M. & J. WETTERLIND. 2010. Chapter five— visible and near infrared spectroscopy in soil science. In: Donald, L.S. (Ed.), *Advances in Agronomy*, vol. 107. Academic Press, pp. 163–215.
- TABATABAI, M.A. 1994. Soil enzymes. In: WEAVER, R.W.; SCOTT, A. & BOTTOMELEY, P.J., eds. *Methods of soil analysis: microbiological and biochemical properties*. Madison, Soil Sci. Soc. Am., 2:778-835, 1994. (Special Publication, 5)
- TAYEBI, M., FIM ROSAS, J.T., MENDES, W.D.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. & J. A.M. DEMATTÊ. 2021. Drivers of Organic Carbon Stocks in Different LULC History and along Soil Depth for a 30 Years Image Time Series. *Remote Sens.* 2021, 13, 2223. <https://doi.org/10.3390/rs13112223>
- USGS. 2018. USGS EROS archive - digital elevation - Shuttle radar Topography mission (SRTM) 1 arc-second global [WWW document] *Earth Resour. Obs. Sci. Cent* <https://doi-org.ezproxy.unal.edu.co/10.5066/F7PR7TFT> (2018)
- VANCE, E.D., BROOKES, P.C., JENKINSON, D.S., 1987. An extraction method for measuring soil microbial biomass C. *Soil Biol. Biochem.* 19 (6), 703–707. [https://doi.org/10.1016/0038-0717\(87\)90052-6](https://doi.org/10.1016/0038-0717(87)90052-6).
- VISCARRA-ROSSEL, R. & T. BEHRENS. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158 (2010) 46–54. doi:10.1016/j.geoderma.2009.12.025
- VISCARRA ROSSEL, R.A. & W.S. HICKS. 2015. Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. *European Journal of Soil Science* 66, 438–450. <https://doi.org/10.1111/ejss.12237>.
- VISCARRA-ROSSEL, R.A., YANG, Y., BISSETT, A., BEHRENS, T., DIXON, K., NEVIL, P. & S. LI. 2022. Environmental controls of soil fungal abundance and diversity in Australia's diverse ecosystems. *Soil Biology and Biochemistry* 170 (2022) 108694. <https://doi.org/10.1016/j.soilbio.2022.108694>
- WANG, Y. & Y. LI. 2023. Mapping the ratoon rice suitability region in China using random forest and recursive feature elimination modeling. *Field Crops Research* 301 (2023) 109016.
- WEBER, K. A., ACHENBACH, L. A. & J. COATES. 2006. Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat. Rev. Microbiol.* 4, 752–764. doi: 10.1038/nrmicro1490
- WEBSTER, R. & M. OLIVER. 2007. *Geostatistics for Environmental Scientists*. Editorial Jhon Wiley & Sons, Ltd. 333 p.
- WHITMAN, T., NEURATH, R., PERERA, A., CHUJACOBY, I., NING, D., ZHOU, J., NICO, P., PETTRIDGE, J. & M. FIRESTONE. 2018. Microbial community assembly differs across minerals in a rhizosphere microcosm. *Environmental Microbiology* 20, 4444–4460.

- YANG, L., LI, X., SHIB, J., SHEN, F., QI, F., GAO, B., CHEN, Z., ZHU, A. & C. ZHOU. 2020. Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. *Geoderma* 369 (2020) 114337.
- YANG, Y., SHEN, Z., BISSET, A. & R. VISCARRA ROSSEL. 2021. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. *SOIL Discuss.* [preprint], <https://doi.org/10.5194/soil-2021-79>, in review, 2021.
- YANG, Y., SHEN, Z., BISSETT, A. & R. VISCARRA-ROSSEL. 2022. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. *SOIL*, 8, 223–235, 2022. <https://doi.org/10.5194/soil-8-223-2022>
- ZERAATPISHEH, M., JAFARI, A., BODAGHABADI, M.B., AYOUBI, S., TAGHIZADEH-MEHRJARDI, R., TOOMANIAN, N., KERRY, R. & M. XU. 2020. Conventional and digital soil mapping in Iran: past, present, and future *Catena*, 188 (2020), p. 104424
- ZHANG, L., CHEN, X., XU, Y., JIN, M., YE, X., GAO, H., CHU, W., MAO, J. & M. THOMPSON. 2020. Soil labile organic carbon fractions and soil enzyme activities after 10 years of continuous fertilization and wheat residue incorporation. *Scientific Reports* (2020) 10:11318. <https://doi.org/10.1038/s41598-020-68163-3>.
- ZHANG, Y., FREEDMAN, Z., HARTEMINK, A., WHITMAN, T. & J. HUANG. 2022. Characterizing soil microbial properties using MIR spectra across 12 ecoclimatic zones (NEON sites). *Geoderma* 409 (2022) 115647.
- ZHANG, L. & L. WANG. 2023. Optimization of site investigation program for reliability assessment of undrained slope using Spearman rank correlation coefficient. *Computers and Geotechnics* 155 (2023) 105208.
- ZHANG, X., CHEN, S., XUE, J., WANG, N., XIAO, Y., CHEN, Q., HONG, Y., ZHOU, Y., TENG, H., HU, B., ZHUO, Z., JI, W., HUANG, Y., GOU, Y., RICHER-DE-FORGES, A., ARROUAYS, D. & Z. SHI. 2023. Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping. *Geoderma* 432 (2023) 116383
- ZHAO, G., DING, W., TIAN, J., LIU, J., GU, Y., SHI, S., WANG, R. & N. SUN. 2022. Spearman rank correlations analysis of the elemental, mineral concentrations, and mechanical parameters of the Lower Cambrian Niutitang shale: A case study in the Fenggang block, Northeast Guizhou Province, South China. *Journal of Petroleum Science and Engineering* 208 (2022) 109550
- ZORNOZA, R., GUERRERO, C., MATAIX-SOLERA, J., SCOW, K., ARCENEGUI, V. & J. MATAIX-BENEYTO. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem.* 2008 July; 40(7): 1923–1930. doi:10.1016/j.soilbio.2008.04.003.

Appendix

Regression models using Mid-IR spectra. λ = autoregressive parameters, ρ = spatial autocorrelation coefficient, r = correlation coefficient, MIT = Moran index test.

Model	λ	ρ	AIC	r	NT	MIT	variables
Betaglucosidase							
PAR	0.2052 0.72152817 2		269.52	0.114515	0.00015087635 5	0.28960972930 1	B glucosidase
SEM	- 0.70765326 0.35795617 3		248.1838	0.7900588 3		0.18969030985	
SLM		0.0980222 1 8.53E-01	248.9943	0.779016	6.40E-01	7.81E-01	
SARAR	- 0.90623758 3.86E-01	0.3358844 6 5.59E-01	248.213	0.79459	6.81E-01	0.666415	
SLMA		- 0.9674810 9 1.49E-01	244.0075	0.876088	2.34E-01	4.78E-01	
Multiple linear regression				0.7863	0.40679382	0.62178553	X1310 + X1581 + X1826 + X1856 + X
MBC							
PAR	0.06365281 9.17E-01		-36.1966	0.03381	4.45E-02	3.95E-01	MBC
SEM	-1.656003 3.71E-02		-53.0489	0.765684	1.25E-01	7.91E-01	
SLM		-0.449537 4.71E-01	-46.7547	0.71352	2.74E-01	5.90E-01	
SARAR	-1.948006 5.91E-03	-0.37682 5.67E-01	-49.2735	0.78073	9.89E-02	0.594808	
SLMA		-1.541499 4.58E-02	-47.7809	0.7684	3.25E-01	8.85E-01	
Multiple linear regression				0.694275	0.0758433	0.26690437	X1864 + X1866 + X1885 + X1887 + X1899
Urease							
PAR	-0.1227791 8.39E-01		187.4721	0.057155	8.06E-02	4.72E-01	Urease
SEM	-0.2254 7.70E-01		177.336	0.662996	5.42E-01	5.42E-01	
SLM		-0.1394 8.09E-01	177.3627	0.66201	6.02E-01	5.47E-01	
SARAR	-0.2295 8.62E-01	-0.02123 9.84E-01	177.8583	0.65665	6.29E-01	0.036841	
SLMA		-0.64557	178.903	0.7171	7.18E-01	9.94E-01	

		4.11E-01					
Multiple linear regression				0.9164282 9	0.5450442	0.953685	X604 + X608 + X620 + X
Phosphatase							
PAR	-0.695538 3.23E-01		436.5051	0.294914	4.66E-02	6.98E-01	Phosphatase
SEM	-1.708131 3.04E-02		419.5673	0.744989	6.90E-01	8.64E-01	
SLM		-0.848109 1.83E-01	424.1408	0.688341	4.81E-02	8.62E-01	
SARAR	-1.6005 7.61E-02	- 0.3892522 6.07E-01	421.1563	0.74545	5.21E-01	0.90999	
SLMA		-1.55095 3.02E-02	423.0452	0.759904	3.81E-02	5.84E-01	
Multiple linear regression				0.8004089 5	0.95430983	0.44943272	X610 + X614 + X622 + X624 + X689 + X1083 + X2660 + Y
POM							
PAR	-0.32003 6.31E-01		-54.386	0.14853	1.05E-06	5.35E-01	POM
SEM	-1.70472 4.10E-02		-55.6317	0.603126	1.34E-03	8.57E-01	
SLM		0.7736157 2.68E-01	-54.0022	0.47296	3.05E-04		
SARAR	-1.70821 6.27E-02	0.0075366 9.92E-01	-53.6318	0.603425	1.33E-03	0.854312	
SLMA		-1.398681 6.11E-02	-51.0468	0.66046	1.18E-04	6.12E-01	
Multiple linear regression				0.5005439 5	0.00146353	0.23588697	X604 + X1824 + X1826 + X + Y
mOM							
PAR	0.51104233 2.76E-01		66.36513	0.316946	4.60E-02	1.37E-01	mOM
SEM	-2.243194 3.15E-03		17.68908	0.93247	7.48E-02	5.90E-01	
SLM		0.417768 2.66E-01	22.1285	0.90499	6.64E-02	3.31E-01	
SARAR	-2.2696241 4.71E-07	0.5093 4.17E-02	16.11034	0.936927	2.86E-02	0.558712	
SLMA		-1.06488 1.73E-01	17.23342	0.95357	2.87E-02	3.85E-01	
Multiple linear regression				0.9193920 6	0.46953504	0.39331561	X614 + X616 + X618 + X1887 + X + Y

Regression models using Vis NIR-SWIR spectra. λ = autoregressive parameters, ρ = spatial autocorrelation coefficient, r = correlation coefficient, MIT = Moran index test.

Model	λ	ρ	AIC	r	NT	MIT	variables
Betaglucosidase							
PAR	0.205243 7.22E-01		269.5247	0.114515	0.00015087635 5	0.28960972930 1	B glucosidase
SEM	0.0485971 9.43E-01		264.745	0.5889	3.10E-04	3.98E-01	
SLM		0.350019 5.03E-01	263.635	0.579	5.40E-04	2.79E-01	
SARAR	- 0.2996447 2 8.52E-01	0.3865110 5 7.00E-01	266.4204	0.601578	3.09E-04	0.475324	
SLMA		-0.638987 3.38E-01	258.7513	0.798672	2.56E-02	7.61E-01	
Multiple linear regression				0.733414	0.010611	0.841555	X350 + X360 + X361 + X2500 + X
MBC							
PAR	0.0636528 1 9.17E-01		-48.1165	0.03381	4.45E-02	3.95E-01	MBC
SEM	-1.787408 4.59E-02		-53.0489	0.762504	4.82E-02	9.29E-01	
SLM		-0.014175 9.82E-01	-47.6278	0.676818	7.32E-02	5.69E-01	
SARAR	-1.415988 1.26E-01	0.2957249 6.04E-01	-48.2873	0.734847	2.53E-02	0.781315	
SLMA		-0.239454 7.13E-01	-47.5104	0.769772	9.01E-02	7.61E-01	
Multiple linear regression				0.775849	0.213436	0.296102	X356 + X360 + X371 + X2300 + X2467
Urease							
PAR	- 0.1227791 8.39E-01		187.4721	0.057155	8.06E-02	4.72E-01	Urease
SEM	-0.2254 7.70E-01		177.336	0.662996	5.42E-01	5.42E-01	
SLM		-0.1394 8.09E-01	177.3627	0.66201	6.02E-01	5.47E-01	
SARAR	-0.2295 8.62E-01	-0.02123 9.84E-01	177.8583	0.65665	6.29E-01	0.036841	
SLMA		-0.64557 4.11E-01	178.903	0.7171	7.18E-01	9.94E-01	
Multiple linear regression				0.853596	0.67725	0.466732	X350 + X410 +

regression						X2499 + X2148 + Y
Phosphatase						
PAR	-0.695538 3.23E-01		436.5051	0.294914	4.66E-02	6.98E-01 Phosphatase
SEM	-1.708131 3.04E-02		419.5673	0.744989	6.90E-01	8.64E-01
SLM		-0.848109 1.83E-01	424.1408	0.688341	4.81E-02	8.62E-01
SARAR	-1.6005 7.61E-02	- 0.3892522 6.07E-01	421.1563	0.74545	5.21E-01	0.90999
SLMA		-1.55095 3.02E-02	423.0452	0.759904	3.81E-02	5.84E-01
Multiple linear regression						X356 + X357 + X667 + X1911 + X
POM						
PAR	-0.32003 6.31E-01		-54.386	0.14853	1.05E-06	5.35E-01 POM
SEM	-1.70472 4.10E-02		-55.6317	0.603126	1.34E-03	8.57E-01
SLM		- 0.7736157 2.68E-01	-54.0022	0.47296	3.05E-04	
SARAR	-1.70821 6.27E-02	0.0075366 9.92E-01	-53.6318	0.603425	1.33E-03	0.854312
SLMA		-1.398681 6.11E-02	-51.0468	0.66046	1.18E-04	6.12E-01
Multiple linear regression						X800 + X802 + X808 + X821 + X
mOM						
PAR	0.5110423 3 2.76E-01		66.36513	0.316946	4.60E-02	1.37E-01 mOM
SEM	-2.243194 3.15E-03		17.68908	0.93247	7.48E-02	5.90E-01
SLM		0.417768 2.66E-01	22.1285	0.90499	6.64E-02	3.31E-01
SARAR	- 2.2696241 4.71E-07	0.5093 4.17E-02	16.11034	0.936927	2.86E-02	0.558712
SLMA		-1.06488 1.73E-01	17.23342	0.95357	2.87E-02	3.85E-01
Multiple linear regression						X351 + X375 + X573 + X2360 + X

4. A PROGRESSIVE KNOWLEDGE STRATEGY TO DETECT MICROBIOLOGICAL ACTIVITY IN SOIL THROUGH VIS NIR-SWIR AND MID IR SPECTRAL CHARACTERIZATION AND MACHINE LEARNING TECHNIQUES

Abstract

The activity of microorganisms is related to the degradation of soil organic matter (SOM) and involves intra- and extracellular enzymatic activity, which interferes with its final forms. Therefore, the fractionation of SOM is key to understanding the results of such activity and the destiny of sequestered and stabilized carbon (C). The fine mineral fraction contributes more to the preservation of SOM compared to particulate OM (POM), since the SOM associated with the mineral fraction (MAOM) contains the highest amount of transformed microbial OM, including microbial biomass and necromass. However, POM constitutes the raw material for microbial activation considering that it includes many of the oxygen-containing functional groups that are preferentially utilized by microorganisms. This type of analysis can be assisted by new techniques, such as spectroscopy, as Vis NIR SWIR and Mid IR wavelengths respond to functional groups of different types of soil organic compounds, iron oxides and clay minerals that are related to microorganisms, so their interpretation is considered a key tool for the characterization of this microbial activity. Therefore, the objective of this work was to develop a strategy to analyze microbiological activity at microscale using Vis NIR SWIR and Mid-IR spectroscopic techniques. Thirty-five samples were taken in the region of Piracicaba, São Paulo, and analyzed for quantification of microbial biomass carbon (MBC) and enzymatic activity of beta-glucosidase, urease and phosphatase, and fractionation of the SOM into POM and MAOM was performed. Subsequently, an interpretation of the Vis NIR SWIR and Mid IR spectra of the different fractions was performed, supported with information reported in the literature and with machine learning techniques that allowed the extraction of specific bands related to microbiological activity. Additionally, based on the results of the spectral interpretation, the C forms of the POM and MAOM fractions and their interaction with the particles that compose the clay fraction were characterized using Imbuia infrared beamline of the Sirius synchrotron at the Brazilian Synchrotron Light Laboratory's (LNLS). It was observed that the spectral peaks corresponding to functional groups CH, NH, CO, COH, CO and PO are related to the enzymatic activity of beta-glucosidase, urease and phosphatase. β -glucosidase was related to labile compounds of the MOS so pronounced peaks were identified in the POM fraction, whose amplitude is considerably reduced when analyzed in the MAOM fraction. Urease activity was related to NH functional groups, and its detection was also characteristic in the POM fraction. Phosphatase activity is strongly related to phosphate groups (PO) and prevails in the different fractions, due to the quality of the organo-mineral bonds. It was corroborated that the soil particles present Al and Fe coatings that favor the interaction with the extracellular enzymes. Additionally, it is concluded that the POM fraction includes the organic compounds that activate the enzymatic degradation, and the spectral analysis of this fraction favors the discrimination of the different enzymes, without discarding that a joint analysis of the fractions helps in the identification of the activity of the microorganisms that together with regression techniques and machine learning would favor the quantification and characterization of the microbial diversity, understanding of the destiny of the C forms, C sequestration potential and consequently in the evaluation of soil quality.

Keywords: Soil spectroscopy, Enzyme activity, Functional groups, Microbiological activity

4.1. Introduction

Soil organic matter (SOM) degradation, especially its degree of decomposition, is related to intra- and extracellular enzymatic activity of fungi and bacteria (De Beeck et al., 2021). Enzymes act as regulators in litter decomposition thus influencing labile forms of SOM (Zhang et al., 2020). Stable forms of SOM are also a product of this microbiological activity, as microorganisms produce biomass through metabolic processing of plant residues, and microbial residues are converted into stable forms through interaction with soil minerals (Adamczyk et al., 2019). It is important to study and relate in the predictions of carbon (C) the activity and diversity of exoenzymes and their interactions with minerals, because this microbial and extracellular enzymatic activity directly affects the effectiveness of mineral protection of soil organic carbon (SOC), and promotes its release (Yang et al., 2021), that is, the final forms of SOC are the product of microbial activity and mineralogical interaction.

Additionally, it is important to include the fractionation analysis of SOM because it facilitates the understanding of the fate of the sequestered and stabilized C, helping to discriminate the labile forms of C and the more stable ones (Rasche et al. 2013, Demyan et al. 2012). According to Parikh et al. (2014), the fractionation by density of SOM responds to its degrees of stability, considering that the fine size fraction, linked to the mineral, contributes more to the conservation of SOM compared to particulate OM (POM) (Kleber et al. 2015, Torn et al. 2013). The SOM associated with the mineral fraction (MAOM) involves the largest amount of transformed microbial OM, including microbial biomass and necromass (microbial residues) (Buckeridge et al. 2020, Liang et al. 2019). However, it is the POM that is the prime material for microbial metabolic activation considering that it includes many of the oxygen-containing functional groups that are preferentially utilized by microorganisms (Parikh et al., 2014).

Reflectance spectroscopy is related to the study of the interaction of light with matter, where an exclusive response of a particular element or molecule is presented, thus serving as a fingerprint that facilitates identification (Viscarra Rossel et al. 2006, Madejová et al. 2017). According to Silvero et al. (2020), these interactions occur throughout the electromagnetic spectrum and are classified as electronic transitions in the visible (Vis: 400 - 700 nm), non-fundamental vibrations in the near infrared (NIR: 700 - 1100 nm) and shortwave infrared (SWIR: 1100 - 2500 nm) and fundamental vibrations in the mid-infrared portions (mid MIR: 2500 - 25000 nm, 4000 - 400 cm^{-1}). Considering that microbiological activity presents this close relationship with the types and amounts of SOM, that is, it responds to the quantity and quality of organic carbon (OC) (Rasche et al., 2013), Vis NIR SWIR and Mid IR reflectance

spectroscopy analysis is key for the characterization of this activity, because it facilitates the differentiation of the functional groups of the different soil organic compounds (Ojeda et al. 2008, Viscarra Rossel and Hicks, 2015) and allows the interpretation of clay mineralogy and iron oxides that are related to soil microorganisms (Viscarra Rossel et al. 2022, Yang et al. 2021). There are no specific visible and infrared spectroscopic absorption bands assigned to microbial communities, however, given that the soil physicochemical environment conditions the abundance and function of microorganisms (Rasche et al., 2011), and the spectral response of the soil favors the prediction of fundamental components such as minerals, OM and water content (Yang et al. 2021, Viscarra Rossel et al. 2022) that are required by fungal and bacterial communities for their growth and obtaining energy (Müller, 2015), it is possible to relate wavelengths to microbiological characteristics.

According to Rinnan & Rinnan (2007), the greatest difficulty in relating soil microbiological properties directly to spectral frequencies is due to the fact that microbial biomass and enzymatic activity have low contents compared to other edaphic compounds, so they hardly induce changes in spectra. However, it is possible to predict these microbiological properties due to their strong relationship with different C-containing soil organic compounds (Cohen et al., 2005). This correlation of microbiological activity with OC is indeed reflected in spectral patterns (Chodak, 2011). For example, extracellular enzymes produced by fungi for lignin and lignocellulose degradation release C into the soil solution, so it is possible to interpret spectral information as a bioindicator of soil microbiological properties (Rasche et al., 2013). Additionally, the interaction with soil mineralogy can also be interpreted spectrally, Fe, for example, is related to microbial energy generation for Fe-reducing microorganisms (Weber et al., 2006), as well as for phototrophic bacteria (Hegler et al., 2008), and the spectral response of iron oxides is recognized in the 540, 640 and 900 nm wavelengths, oxides associated with hematite in the 550 nm range, goethite around 440 to 470nm (Dematte et al., 2014). Fe oxides have also been recognized in the SWIR 1400 and 1900nm ranges (Dalmolin et al., 2005). Now, some authors highlight the modification of these wave coverages upon interaction with microbiological activity, for example, Fe oxides associated with fungal diversity in the bands 390, 410, 460 nm (Vis NIR) (Yang et al., 2022), CH-goethite links ratio at 1725 nm (Sharma et al., 2021). Parikh et al. (2014) indicates that the interaction of phosphate groups presents in bacterial cell walls and goethite surfaces, which favors POFe bonds, are distinguished in the bands 1027, 1037 and 1045 cm^{-1} , in the Mid IR spectrum. Additionally, Rong et al. (2010) recognized the interaction of water and polymer bridges favoring bacteria-goethite bonds in the band 1085 cm^{-1} .

In general, as mentioned by Nath et al. (2021) the soil matrix provides and regulates the habitat of different microbial communities, therefore, reflectance spectroscopy analyses on soil samples contribute to the prediction of compounds used by microbes, and to the prediction of products of microbiological activity, because these analyses provide integrated measures of the mineral-organic composition of the soil (Viscarra Rossel et al., 2016). Rasche et al. (2013) point out some wave coverages of the Mid IR spectrum that favor the prediction of glucosidase, xylosidase and urease, as well as bacterial abundance. In relation to Vis-Nir spectra, there have been reported some predictions associated with bacterial abundance (Viscarra-Rossel et al. 2022, Zornoza et al. 2008), microbial biomass (Coûteaux et al. 2003, Zornoza et al. 2008, Chodak 2011), enzymatic activity in soil (Cohen et al. 2005, Zornoza et al. 2008, Chodak 2011), respiration (Zornoza et al. 2008, Chodak 2011), fungal diversity and abundance (Yang et al., 2022). In general, the soil matrix provides and regulates the habitat of different microbial communities (Nath et al., 2021), therefore, reflectance spectroscopy analyses on soil samples contribute to the prediction of compounds used by microbes, and to the prediction of products of microbiological activity, because these analyses provide integrated measures of the mineral-organic composition of the soil (Viscarra Rossel et al., 2016). Rasche et al. (2013) point out some wave coverages of the Mid IR spectrum that favor the prediction of glucosidase, xylosidase and urease, as well as bacterial abundance. In relation to Vis-Nir spectra, there have been reported some predictions associated with bacterial abundance (Viscarra-Rossel et al. 2022, Zornoza et al. 2008), microbial biomass (Coûteaux et al. 2003, Zornoza et al. 2008, Chodak 2011), enzymatic activity in soil (Cohen et al. 2005, Zornoza et al. 2008, Chodak 2011), respiration (Zornoza et al. 2008, Chodak 2011), fungal diversity and abundance (Yang et al., 2022).

NIR and Mid IR reflectance spectra respond to the concentration of compounds related to C-H, N-H, S-H, C=O and O-H chemical bonds, thus allowing differentiation of the organic composition of a soil sample (Zornoza et al. 2008, Parikh et al. 2014, Viscarra Rossel et al. 2022). It is these compounds with which microbial properties (for example, soil microbial biomass or enzymatic activities) are closely related (Parikh et al., 2014). Since these compounds are integrated with fungal and bacterial products. Ammann & Brandl (2011) indicates typical fingerprints for microorganisms by referring to the Mid IR spectrum wave coverages associated with cellular carbohydrate and protein compounds. According to Jiang et al. (2004) hydroxyl, carboxyl, phosphoryl and amide groups are common in bacterial cell walls. In general, the organic nature of bacteria and fungi makes the peak locations similar between SOM and microbial samples. For this reason, many Vis NIR SWIR and Mid IR band

assignments are common to the response of different forms of SOC (Rasche et al 2013, Parikh et al 2014).

Specific wave coverages related to the specific property to be predicted can be extracted from the spectra by multivariate regression analysis and machine learning techniques using different algorithms (Zornoza et al. 2008, Shen and Viscarra Rossel 2021). For example, Rasche et al. (2013) developed a partial least squares regression (PLSR) analysis based on Mid IR reflectance spectroscopy to predict soil microbial biomass and enzymatic activities.

Evaluating the interactions of organic compounds with mineral surfaces is fundamental to understanding SOM stabilization. The charge of biomolecular functional groups on the surfaces of microorganisms and on minerals determines the adhesion process, being useful reflectance spectroscopy because it allows to evaluate such organo-mineral interactions facilitating the analysis of the binding mechanisms (Parik et al., 2014). Moreover, if this type of analysis is performed on the different fractions of the SOM, it favors the understanding of SOC dynamics (Six et al., 2004). The biogeochemical interaction between mineral particles and OM is a fundamental factor in the preservation of SOC, and the study of the selective contribution of clay minerals and iron and aluminum oxides to this stabilization of organic compounds is favored by individual analyses of SOM fractions (Kirsten et al., 2021).

Synchrotron radiation (SR) is basically electromagnetic radiation that is emitted when charged particles are accelerated radially (Hota, 2021). When spectroscopy is combined with a microscope and this type of radiation, it allows the identification of the molecular chemistry of soil particles, because it favors the capture of images that help to differentiate the composition, structure and distribution of chemical constituents with a high signal-to-noise ratio and fine spatial resolution, which helps to improve the understanding of organo-mineral interactions, SOC sequestration, and the spatial distribution of these interactions (Lehmann and Salomon, 2010). Since synchrotron radiation provides spectroscopic analysis with high spatial resolution, it facilitates the mapping of C contents and forms and the interaction with minerals in small soil fractions (Lehmann et al., 2007).

In general, traditional methods for analyzing soil microbiological activity are expensive and time consuming, and given their importance in understanding C dynamics, soil quality and soil health, there is a need to develop techniques through simple and cost-effective methods (Rasche et al. 2013). Spectroscopy has made great progress in detecting specific bands in the Mid IR region. On the other hand, we evolved towards microscale analysis and at this level no work has been done.

Therefore, the objective of this study was to develop a strategy to analyze micro-scale microbiological activity by spectroscopic detection. For these, we performed the interpretation of Vis NIR SWIR and Mid IR spectra of the SOM fractions (POM and MAOM) based on the relationship of the spectra with the quantification of microbial biomass carbon (MBC) and of the enzymatic activity of beta-glucosidase, urease and phosphatase (which interfere in the carbon, nitrogen and phosphorus cycles). Machine learning techniques were used to determine specific bands related to microbiological activity. Additionally, from the results of the spectral interpretation, we characterized the C-forms of the POM and MAOM fractions and their interaction with the particles that compose the clay fraction using the National Imbuia Beamline Synchrotron Light (μ -MIR) laboratory.

4.2. Methodology

4.2.1. Study area

The study area is located in the State of São Paulo, Brazil, with approximately 2,598 km², comprising eight municipalities (Figure 1). According to Alvares et al. (2013), this region presents an average annual temperature ranging between 20 and 22.5 ° C and annual rainfall between 1200 and 1400 mm, with a climate defined as subtropical Cwa, which presents a dry winter and a rainy summer according to the Köppen classification. The terrain includes undulating highlands and rolling hills with altitudes ranging from 450 to 950 m. The area is dominated by agricultural land uses, such as sugarcane and pasture, under no-till and conventional tillage management systems. The main soil types according to the World Reference Base (IUSS Working Group WRB, 2015) are Cambisol, Gleysol, Ferralsol, Nitosol, Lixisol, Leptosol, Arenosol and Planosol.

4.2.2. Selection of sampling points

Thirty-five representative sites of the study area were selected for soil sampling through the conditional Latin hypercube sampling method (Figure 1), which corresponds to a stratified random sampling procedure in which the selected samples respond to multivariate characteristics according to the indicated covariates (Yang et al., 2020). In this case, the covariates considered were the Synthetic Soil Image (SYSI), the terrain elevation model (DEM), soil type, variability in clay content and C, and the maps of Rodríguez-Albarracín et al. (2023) corresponding to microbial biomass carbon (MBC), general abundance of fungi, Actinomycetes and Acidobacteria, enzymatic activity of Beta glucosidase, Urease and phosphatase, and maps of labile and resistant carbon forms.

4.2.3. Laboratory analysis

The 35 selected samples were 2mm sieved and the laboratory analyses were performed (Figure 1). The methodology of Vance et al. (1987) was used to quantify the carbon in the microbial biomass. The enzymatic activity of beta-glucosidase, phosphatase and urease were determined following the methodology of Tabatabai (1994) and Dick et al. (1996). The soil texture (Teixeira et al., 2017) and SOC (Walkley and Black, 1934) analyses were also performed. Finally, the physical fractionation of the samples to quantify C in particulate organic matter (POM) and C in OM associated with the mineral fraction (mOM) (Cotrufo et al., 2019) making use of the methodology described by Jindaluang et al. (2013), which starts from the dispersion of the soil with a 5% sodium hexametaphosphate solution; subsequently, the separation of the sand and clay/silt fraction is performed by 53 μm sieving. The microbiological analyses were performed with wet samples sieved at 2 mm.

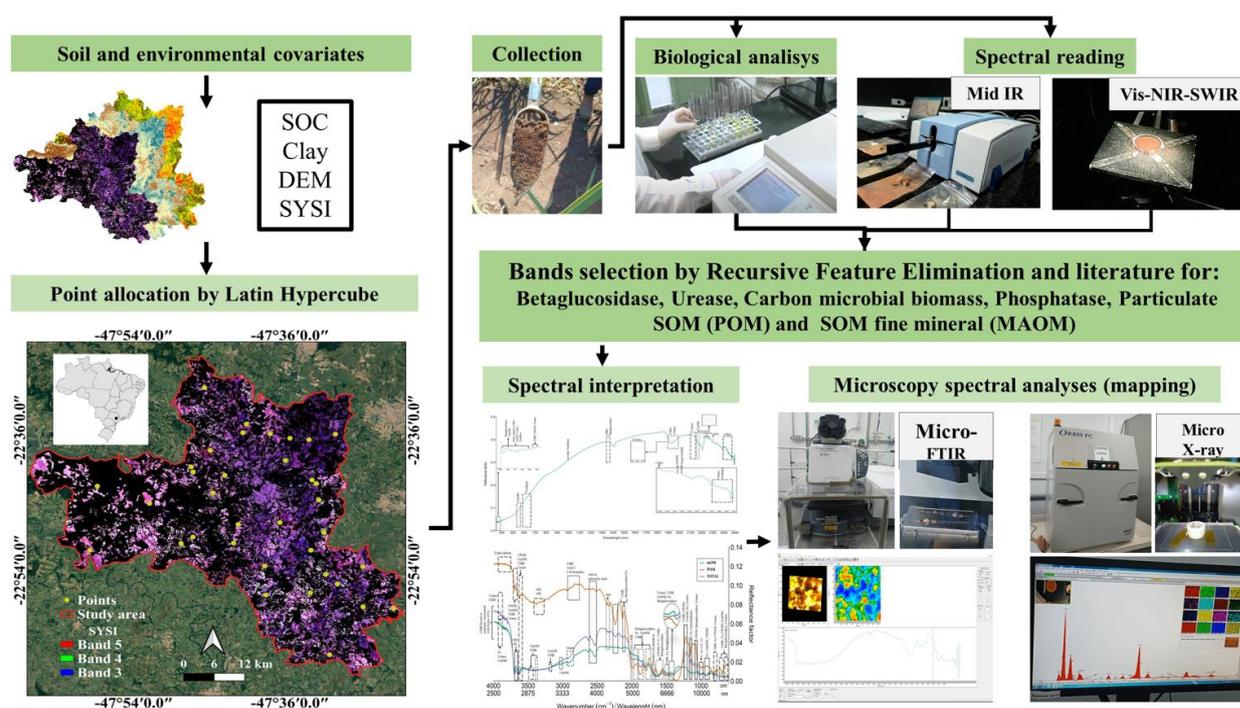


Figure 1. Study area and flowchart of methodology

4.2.4. Spectral data acquisition

Spectral data from the Vis-NIR-SWIR and Mid IR ranges, were acquired through the use of two sensors, the FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, Col., USA) for obtaining reflectance data from 350 to 2500 nm (Vis-NIR-SWIR) (Dematté et al., 2019) and the Alpha Sample Compartment RT-DLaTGS ZnSe sensor (Bruker Optik

GmbH), equipped with a drift attachment (Souza et al., 2020) for obtaining spectral information in the mid-infrared range (4000 to 600 cm^{-1}), with a resolution of 1.2 nm and 64 scans per second as described by Terra et al. (2015). For sensor calibration, a gold plate was used as a standard. FieldSpec 3, has a spectral resolution of 3 nm for the range between 400 and 700 nm and 10 nm for the range from 700 to 2500 nm, therefore, resampling was performed at 1 nm, obtaining 2151 spectral bands (Greschuk et al., 2022). FieldSpec 3 readings require air-dried, ground and sieved soil samples at 2 mm. For Mid IR range readings, soil samples were ground and sieved at 100 mesh.

4.2.5. Statistical analysis and spectral behavior

Due to the high dimensionality presented by the data set related to spectral information, it has redundant and irrelevant information (Chen et al., 2018), so it is necessary to select the most important data set (Wang & Li, 2023). To perform this selection, the Recursive Feature Elimination (RFE) algorithm developed by Guyon et al. (2002) and implemented in the "Caret" package of R is used, applying the Random Forest (RF) machine learning method as an internal model (Kuhn, 2021).

The RFE algorithm is a method with a wide applicability that allows handling nonlinear and more complex relationships, considering comprehensively the relationships between the analyzed factors (Wang & Li, 2023). RFE is based on backward selection, where initially a model is created using all n predictors, the performance is calculated by k -fold cross-validation (RMSE) and the importance of the variable; subsequently the least important predictor is removed from the set, and the model is refitted, again the performance is evaluated and the least important predictor is removed; this is done repeatedly until the optimal number of predictors is determined by taking the model with the best performance (RMSE) (Zhang et al., 2023).

For the present work, the RFE algorithm was applied for each variable (enzymatic activity of betaglucosidase, phosphatase and urease, POM and mOM) and for each range of the electromagnetic spectrum, to obtain the most important Vis NIR SWIR and Mid IR bands of each variable (Figure 1), which were subsequently analyzed directly in the spectra, through a visual interpretation of the distribution patterns.

4.2.6. Mid-IR and XRF Spectroscopy

Mid IR spectral images were recorded at the IMBUA-micro station of the Brazilian Synchrotron Light Laboratory (LNLS), composed with a Focal Plane Array (FPA) IR detector

and a mercury-cadmium-telluride (MCT) detector. The FPA 128x128 detector allows the capture of images at 25-fold magnification, in a spectral range from 4000 to 700 cm^{-1} . The images correspond to 350 x 350 μm^2 field of views with a pixel size of 3.3 μm . In order to avoid particle overlapping, suspensions of soil in water (1:100) were prepared, then a drop of this suspension was placed on gold substrate plates. The advantage of this sample preparation is that it avoids physicochemical alterations, and the particle chemistry remains practically intact (Wan et al., 2007).

Multivariate data analysis is useful to interpret the information generated by spectroscopy (Lehmann and Solomon, 2010), therefore, the interpretation and speciation analysis was performed from the bands identified with RFE and using as support the visual interpretation of the Mid-IR spectrum.

To map the chemical composition of the minerals, the samples were scanned using X-ray fluorescence spectrometer (μ -XRF) (Orbis PC EDAX, USA). The X-rays were generated by a Rh anode operating at 45 kV and 900 μA (Gomes et al., 2019) and focused to 30 μm by polycapillary lens system. X-ray spectra are detected by a silicon drift detector (Junior et al. 2019, Macedo et al. 2021).

As pointed out by Lehmann and Solomon (2010), the combination of these two techniques (Mid-IR Spectro microscopy and XRF) helps to understand the biochemical interaction of organo-mineral complexes, as it favors the understanding of the spatial arrangement of organic and mineral particles.

4.3. Results

4.3.1. Mid IR spectral interpretation

Table 2 shows the Mid IR bands obtained by the RFE, and Table 3 shows the most important bands for each of the evaluated variables, according to the RFE results. The Mid IR spectra readings were taken for each of the SOM fractions (Table 2 and 3) and considering these bands and the reports of different authors, the spectra were interpreted. In order to facilitate the description of the interpretation, 3 contrasting samples are presented, in relation to C and clay content (Table 1). Sample P16, is a more sandy sample with the lowest C content, compared to samples P 30 and P35.

Table 1. Physicochemical analysis of samples P16, P30 and P35. MBC= microbial biomass carbon, Ctotal= total carbon, CPOM= carbon of particulate organic matter, CmOM= carbon of organic matter associated with minerals, Fe= iron, P= phosphorus, Ca= calcium, Mg= magnesium, SB= base saturation.

Laboratory analysis		Samples		
		P16	P30	P35
Microbiological	MBC (mg C g ⁻¹ Soil)	0.4	0.43	0.7
	Ureasea (ug NH ₄ Kg ⁻¹ Soil)	4.8	8.12	12.6
	Phosphatasea (mg PNF Kg ⁻¹ Soil)	137.91	91.75	310.16
	Betaglucosidasea (mg PNF Kg ⁻¹ Soil)	6.51	16.2	32.67
Fractionation MOS	C total (%)	0.451	1.119	2.656
	C POM (%)	0.062	0.092	0.300
	CmOM (%)	0.389	1.026	2.356
	CPOM_gkg	0.618	0.923	3.005
	CmOM_gkg	3.887	10.265	23.558
Chemical	Fe mgkg	37.4	7.8	10.5
	P mgkg	32.3	24.2	29.5
	pH H ₂ O	5.3	6.9	6.1
	Ca mmolkg	4	23.7	73.4
	Mg mmolkg	2.1	12.3	28.6
	SB mmolkg	6.7	37	115.1
Physical	Sand gkg	941	867	149
	Silt gkg	22	33	103
	Clay gkg	37	100	748

It is observed that for the MBC of the unfractionated sample, which we will call "total", the most important bands of the Mid IR spectrum are in the range 1880, 864 and 602 cm⁻¹ (Table 2). Phosphatase and Urease present similar ranges in the "total" samples, and in general the different variables of this fraction present high importance in the bands 600 - 620, 1600-1900 cm⁻¹ with the exception of Fe. If we observe the Mid IR spectra (Figure 2, 3 and 4), in the ranges below 1000 cm⁻¹ is where different peaks and valleys are presented related to enzymatic activity, MBC and labile and resistant forms of C. Analyzing the spectra of sample P16, the spectral response of Fe also follows these microbiological variables, due to the fact that microorganisms take energy from Fe to carry out their metabolic processes (Weber et al. 2006, Müller 2015). Between 1000 and 1500 cm⁻¹ high activity is also present, and a high ratio is evident in the enzymatic activity of Betaglucosidasea, phosphatasea and Ureasea, as well as POM.

Interaction of betaglucosidasea, phosphatasea and MBC activity is evidenced between 2000 and 1500 cm⁻¹, together with Fe (Figure 2). On the contrary, the activity of Urease,

phosphatase and MBC is concentrated from 4000 to 3500 cm^{-1} , also maintaining the interaction with Fe. It is also observed that the highest reflectance is present in the POM fraction, because it is related to the coarsest particles ($>53 \mu\text{m}$) and it is in this fraction where the functional groups with the highest presence of O are present (Parikh et al., 2014), and it is the fraction that contains the highest food contribution of the microorganisms, because it presents the most easily available forms of C (Follet et al. 2001, Zhang et al. 2020). The high reflectance in each of the most important band ranges is reduced in the total and mOM fractions, especially in the latter, where the most stable carbon forms and smaller particles associated with the clay fraction are found. The unfractionated sample behaves similarly to POM between 4000 to 2000 cm^{-1} , with lower reflectance in the range 1200 to 600 cm^{-1} .

Table 2. Bands of the Mid IR spectrum related to the microbiological variables (MBC (microbial biomass carbon), enzymatic activity of ureasea, phosphatasea and betaglucosidasea, POM (particulate organic matter), mOM (organic matter associated with minerals), total C, Fe (iron)), according to the results of the Recursive Feature Elimination (RFE) algorithm.

Variables	Fractionation MOS Mid IR (cm^{-1})		
	Total	POM	mOM
MBC	600, 602, 604, 606, 614, 616, 618, 620, 622, 626, 640, 651, 661, 667, 669, 671, 679, 681, 816, 822, 1222, 1226, 1230, 1234, 1236, 1240, 1242, 1244, 1246, 1250, 1267, 1269, 1275, 1287, 1289, 1291, 1293, 1540, 1542, 1556, 1558, 1560, 1824, 1850, 1852, 1854, 1856, 1858, 1860, 1862, 1864, 1866, 1868, 1870, 1872, 1875, 1877, 1879, 1881, 1883, 1885, 1887, 1889, 1891, 1893, 1895, 1897, 1899, 1901, 1903, 1905, 1907, 1942, 2001, 2005, 2044,	777, 779, 781, 785, 828, 830, 832, 1289, 1312, 1322, 1324, 1326, 1332, 1344, 1346, 1348, 1354, 1358, 1361, 1363, 1365, 1367, 2727, 2729, 2748, 2758, 2760, 2762, 2764, 2766, 2768, 2770, 2772, 2778, 2782, 2784, 2786, 2790, 2792, 2805, 2829, 2870, 2872, 2888, 2925, 2945, 2956, 2972, 2980, 3678, 3680,	618-916, 922, 930-948, 957-965, 979-1259, 1271-1532, 1540-1544, 1556-1605, 1613-1685, 1693-1703, 1711, 1715, 1724-2056, 2064-2072, 2072, 2079, 2087-2109, 2121-2127, 2140-2150, 2162-2231, 2240-2244, 2250, 2254, 2262-2297, 2309-2319, 2331-2358, 2364-2382, 2393, 2403, 2413, 2425, 2429, 2433, 2444, 2452, 2454, 2470, 2482, 2489, 2491, 2501-2507, 2513, 2519, 2533, 2558, 2560, 2566, 2576, 2586, 2605, 2617, 2646, 2660, 2664, 2678, 2680, 2686, 2690, 2697, 2709, 2713, 2721, 2733, 2743, 2746, 2748, 2756, 2760, 2780, 2788, 2799, 2801, 2811, 2813, 2821, 2823, 2852, 2854, 2862, 2884, 2894, 2929, 2933, 2945, 2949, 2954, 2962, 2964, 2970, 2984, 2992, 3005, 3023, 3027, 3037, 3188, 3196, 3204, 3443, 3594, 3614, 3616, 3619, 3621, 3633, 3637, 3643, 3661-3676, 3690-3698
Ureasea	600-620, 632-643, 659-704, 722, 730-732, 755, 761-777, 781-820, 824, 832, 842, 873, 895, 902, 912, 918, 922, 926-928 932, 967, 987-989, 993-997, 1006-1008, 1012-1018, 1024-1026, 1032-1057, 1063-1101, 1108, 1116, 1122-1497, 1503-1805, 1815 1824-1838, 1846-1848 1854-1901, 1909-2072, 2095 2130 2138 2162 2170, 2201, 2207, 2221-2256, 2282 2295-2297 2305, 2327, 2348, 2352, 2358, 2360, 2366, 2401, 2440, 2452, 2456, 2458, 2464, 2476-2480, 2489-2501, 2519-2554, 2590, 2609, 2898, 2919, 2925, 3164, 3264, 3272, 3335, 3404, 3425, 3443-3445, 3453-3455, 3472-	614, 616, 618, 691, 694, 696, 765, 787, 1018, 1020, 1022, 1024, 1026, 1028, 1030, 1032, 1034, 1036, 1293, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1336, 1346, 1348, 1354, 1356, 1358, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1383, 1399, 1401, 1403, 1409, 1412, 1414, 1416, 1418, 1420, 3698, 3700, 3702,	600-628, 632, 634, 640, 659, 665-669, 687, 694, 698, 700, 708, 730, 740, 767, 783, 791, 793, 804-818, 828-834, 840, 999, 1020, 1024, 1034, 1038-1046, 1050-1059, 1063-1067, 1085, 1101, 1110-1120, 1126, 1128, 1136-1140, 1150, 1161, 1163, 1167, 1169, 1183, 1185, 1189-1212, 1230-1371, 1383, 1385, 1401, 1403, 1409, 1416, 1424, 1426, 1428, 1440, 1444, 1458, 1475, 1483, 1487, 1540-1550, 1558-1569, 1575, 1577, 1601, 1607, 1622-1640, 1671, 1683, 1766, 1777, 1787, 1791, 1803, 1811, 1815, 1819-1824, 1832-1840, 1848-1858, 1870, 1883, 1891-1911, 1928, 1979-2034, 3780, 3845, 3888, 3890-3896,

	3476,3504, 3598, 3600, 3629, 3633, 3645-3649,3657-3663,3682-3700 3714,3833-3839 3855, 3900,3904-3910, 3933-3941, 3951, 3959, 3963-3969, 3975, 3980-3998		3904, 3906, 3949, 3951, 3965-3971, 3980-3998
Phosphatasea	600-630, 634, 640-651, 659-665, 673-704, 726-736, 742, 769-779, 785, 789-895, 910, 934-936, 957, 961, 969, 979-993, 1018-1026, 1042, 1069-1097, 1104-1152, 1159-1191, 1201, 1212-1283, 1295-1299, 1320-1328, 1342, 1354, 1358, 1399, 1403, 1426-1428, 1438, 1454-1458, 1465, 1509, 1513, 1520, 1524, 1556-1560, 1601, 1628-1632, 1648-1658, 1683, 1687, 1693-1701, 1720, 1748-1752, 1768, 1781, 1875-1883, 1911, 1942, 1958, 2038-2056, 2231, 2242, 2250, 2358, 2482, 2499, 2523, 2527, 2533-2535, 2593, 2605, 2611-2621, 2637, 2652-2662, 2670-2688, 2695-2727, 2737-2741, 2750-2770, 2778-2780, 2847-2854, 2911, 2923, 3098, 3231, 3249, 3280-3284, 3298, 3376, 3435-3437, 3451-3466, 3523-3525, 3610, 3619-3629, 3635, 3643-3674, 3682-3727, 3733, 3751, 3808-3810, 3835-3839, 3853-3857, 3873, 3900-3908, 3939, 3951-3955, 3961-3969, 3980-3998,	600-612, 622-630, 643-647, 669-671, 679-704, 712-724, 749-753, 767-775, 781, 789-818, 834-840, 895, 987-999, 1006-1026, 1034, 1042, 1053-1057, 1101, 1138-1140, 1152-1159, 1230-1248, 1254, 1261, 1273-1287, 1297-1416, 1426, 1432-1436, 1446, 1448, 1534, 1556, 1569, 1583-1607, 1622, 1652, 1736, 1856, 1885, 1895, 2025, 2030-2036, 2525, 2533, 2544, 2692, 2741, 2829, 2839, 2847, 2850-2862, 2870, 2876, 2911-2958, 2988, 2990, 2996, 3000, 3003, 3213, 3441, 3606-3733, 3741, 3873, 3929-3933, 3955-3957, 3992-3998,	620-647, 653-708, 716, 722-736, 753, 771, 793, 800-816, 826-853, 879, 906, 912, 914, 948-953, 969-1079, 1087-1097, 1108, 1114-1218, 1248, 1256, 1261-1275, 1289-1422, 1436, 1450-1460, 1473, 1483-1550, 1558-1597, 1611-1634, 1646, 1656, 1669, 1679, 1681, 1683, 1691, 1693, 1711, 1713, 1724, 1726, 1736, 1748, 1750, 1758, 1781-1809, 1815, 1822, 1838, 1846, 1852, 1868, 1870, 1872, 1889-1903, 1913-1917, 1926, 1950, 1952, 1962, 1989-2003, 2030, 2484, 2491-2495, 2670-2674, 2954, 2958, 3007, 3188, 3196, 3204, 3251, 3364, 3374, 3417, 3508, 3510, 3551-3557, 3614-3621, 3659, 3670-3708, 3718, 3778, 3802, 3833, 3839, 3849, 3851, 3865, 3900-3924, 3931, 3937, 3943, 3947, 3951-3998,
POM	602, 604, 820, 1299, 1301, 1303, 1752, 1824, 1826, 1832, 1834, 1842, 1919, 1921, 1923, 2042,	600-640, 647-716, 726-734, 740-855, 865-995, 1004, 1014-1034, 1044-1169, 1175-1226, 1244, 1246, 1252-1254, 1265-1450, 1460, 1467, 1471, 1475, 1483-1513, 1520, 1524, 1540-1593, 1601-1697, 1705-1715, 1720-1734, 1742-1758, 1777, 1779, 1787, 1840, 1852, 1862, 1864, 1870-1875, 1887-1905, 1921-1926, 1932, 1934, 1962, 1972, 1995, 2032, 2038-2062, 2089-2121, 2127-2132, 2156, 2160, 2180, 2187, 2197, 2201-2248, 2256, 2260-2285, 2293, 2317-2319, 2331-2348, 2368-2378, 2384, 2389, 2405, 2415, 2429, 2507, 2517, 2566, 2584, 2686, 2733, 2780, 2782, 2913, 2952, 3037, 3056, 3109, 3235, 3257, 3315, 3374, 3425, 3464-3472, 3484-3488, 3500-3514, 3563, 3565, 3580-3586, 3600, 3608, 3614, 3619, 3623, 3665, 3674, 3696-3704, 3727, 3733, 3739, 3751, 3792, 3816, 3818, 3851-3855, 3878, 3882, 3898-3902, 3945-3947, 3961-3975, 3990-3998,	624-638, 649, 651, 665-669, 677, 679, 687, 689, 694-702, 718, 783, 808-812, 914, 944, 997, 1010, 1030, 1034, 1040, 1046, 1244, 1250-1336, 1342-1356, 1369, 1371, 1381-1454, 1460-1509, 1520-1530, 1542, 1564-1615, 1622-1634, 1640, 1709, 1752, 1771, 1777, 1785, 1789-1970, 1983, 1997, 2005-2015, 2042-2050, 2056, 2064-2074, 2784-2792, 2803, 2815, 2817, 2935, 2956, 3619, 3649, 3663, 3665, 3667, 3670, 3690, 3692,

C total	602, 606, 614, 616, 618, 1862, 1864, 1881, 1883, 1885, 1887,	602, 694, 696, 822, 1022, 1024, 1026, 1028, 1030, 1091, 1093, 1095, 1097, 1099, 1307, 1310, 1312, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1381, 1383, 1385, 1387, 1389, 1391, 1393, 1395, 1397, 1399, 1401, 1403, 1405, 1409, 1412, 1418, 1420, 1650, 1893, 2727, 2748, 2756, 2762, 2766, 2772, 2774, 2778, 2784, 2790, 2792, 2892, 3614, 3616, 3619, 3621, 3623, 3625, 3627, 3629, 3631, 3633, 3635, 3637, 3641, 3661, 3692, 3694, 3696, 3698, 3700, 3955,	622-634, 645-657, 665-669, 677, 679, 685-689, 696-702, 718-726, 734-742, 749-759, 769, 783-810, 822, 826, 832-849, 857, 871-875, 910, 922, 928, 979, 991-1152, 1159-1187, 1197, 1203, 1228, 1246-1354, 1361-1369, 1393, 1395, 1412, 1416, 1424, 1428, 1448-1463, 1473-1477, 1493-1507, 1528-1534, 1540, 1542, 1558, 1560, 1571-1585, 1597-1624, 1634, 1636, 1689, 1760, 1768, 1777, 1781-1819, 1838-1946, 1962, 1981-1987, 1997-2025, 2030-2048, 2187, 2236, 2240-2244, 2258-2262, 2403, 2425, 2905, 2915, 2921-2925, 2949, 3188, 3196, 3204, 3251, 3433, 3443-3449, 3457, 3459, 3690-3700
mOM	600, 606, 614, 616, 618, 1862, 1864, 1881, 1883, 1885, 1887,	1312, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1381, 1383, 1385, 1387, 1389, 1391, 1399, 1401, 1403, 3692, 3694, 3696, 3698, 3700,	622-632, 640, 645-649, 655-669, 677, 679, 687, 698, 700, 702, 710, 749, 751, 757, 769, 791-806, 820, 836, 840-844, 871-875, 993, 995, 1010-1020, 1028-1155, 1163-1187, 1201, 1261-1269, 1279, 1285-1289, 1297-1354, 1361, 1389, 1428, 1509, 1756, 1775, 1779, 1801, 1805, 1830, 1850-1917, 1928, 1932, 1974, 1977, 1981, 1987, 1995-2011, 2023, 2028, 2030, 2905, 2915-2925, 3188, 3196, 3251, 3443, 3445, 3459, 3698
Fe	604, 612, 618, 1279-1314, 1320-1332, 1340-1344, 1352-1356, 1365, 1416-1471, 1540-1542, 1562-1607, 1622-1632, 1677-1679, 1707-1713, 2050, 3692, 3716, 3933-3998	1318, 1320, 1322, 1324, 1326, 1328, 1395, 1412, 1414, 1420, 1877	640, 643, 645, 647, 649, 812, 814, 816, 936, 938, 942, 944, 946, 1371, 1442, 1444, 3616, 3619, 3661, 3663, 3665, 3667, 3670, 3672, 3674, 3676
Betaglucosidasea	606, 608, 1310, 1581, 1826, 1834, 1856, 1870, 2040, 2042, 2046	606, 1310, 1312, 1326, 1346, 1365, 1367, 1399, 1401, 1426, 1872	632, 1034, 1036, 1044, 1046, 1050, 1053, 1587, 1644, 1781, 1850

Table 3. Bands of the Mid IR and Vis NIR SWIR spectrum of major importance for the prediction of microbiological variables (MBC (microbial biomass carbon), enzymatic activity of ureasea, phosphatasea and betaglucosidasea, POM (particulate organic matter), mOM (organic matter associated with minerals), total C, Fe (iron)), according to the results of the Recursive Feature Elimination (RFE) algorithm.

Variables	Fractionation MOS Mid IR (cm ⁻¹)			Vis NIR SWIR (nm)
	Total	POM	mOM	
MBC	864, 1883, 1887, 1885, 602	1363, 830, 1365, 1361, 1324	1316, 1312, 802, 804, 2003	350, 351, 360, 361, 356, 371
Ureasea	606, 604, 618, 608, 620	1399, 1326, 1324, 1401, 1418, 1356	667, 612, 665, 634, 622	350, 2499, 366, 351, 379
Phosphatasea	610, 612, 624, 614, 606, 622	1338, 1340, 1354, 3704, 3661	1057, 1055, 1046, 1067, 1044	356, 357, 361, 360, 2482, 2476
POM	604, 602, 1834, 1832, 1919, 1921	1312, 689, 1310, 614, 1664, 1577	1477, 1467, 1583, 1293, 1581	802, 790, 353, 813, 812, 821
C total	1883, 1887, 618, 614, 1862	1326, 1399, 1324, 1346, 1369	1316, 1320, 1318, 1322, 873	1883, 1887, 618, 614, 1862
mOM	614, 618, 1881, 1883, 1864, 1862	1326, 1399, 1324, 1322, 1401	1320, 1318, 1316, 873, 1322	351, 360, 350, 375, 2444, 2360
Fe	3998, 1436, 1434, 1422, 1589	1324, 1320, 1322, 1414, 1395	936, 812, 643, 814, 640	-
Betaglucosidasea	2042, 1856, 1834, 1581, 1862, 606, 608	1326, 1328, 1399, 1324, 1346	600, 602, 1852, 1850, 1589	350, 351, 361, 360, 2500

Peaks of the functional groups -NH (3400 cm⁻¹), NH⁺ (2400 cm⁻¹), PO₂ (960, 1200 cm⁻¹), P-O=R (1000-1100 cm⁻¹), amide II and Amide I (1600 cm⁻¹) proposed by Parikh et al. (2014), Liu et al. (2015) and Ladd et al. (1993) were identified, which are lost in the mOM fraction (Figure 2). The spectral response of amide I and II is not recognized in the "total" sample, possibly due to the interaction of other compounds. The Urease behavior identified in the spectrum of the total sample between 3000 and 3100 cm⁻¹, is transferred to the POM fraction but highlighting the response of MBC and aromatic CH bonds and in the mOM fraction this response is associated with the more stable C forms. The interaction of phosphatasea activity and the functional groups -NH, OH of carboxylic acids and NH⁺ (3400, 2550, 2400 cm⁻¹ respectively) is also highlighted.

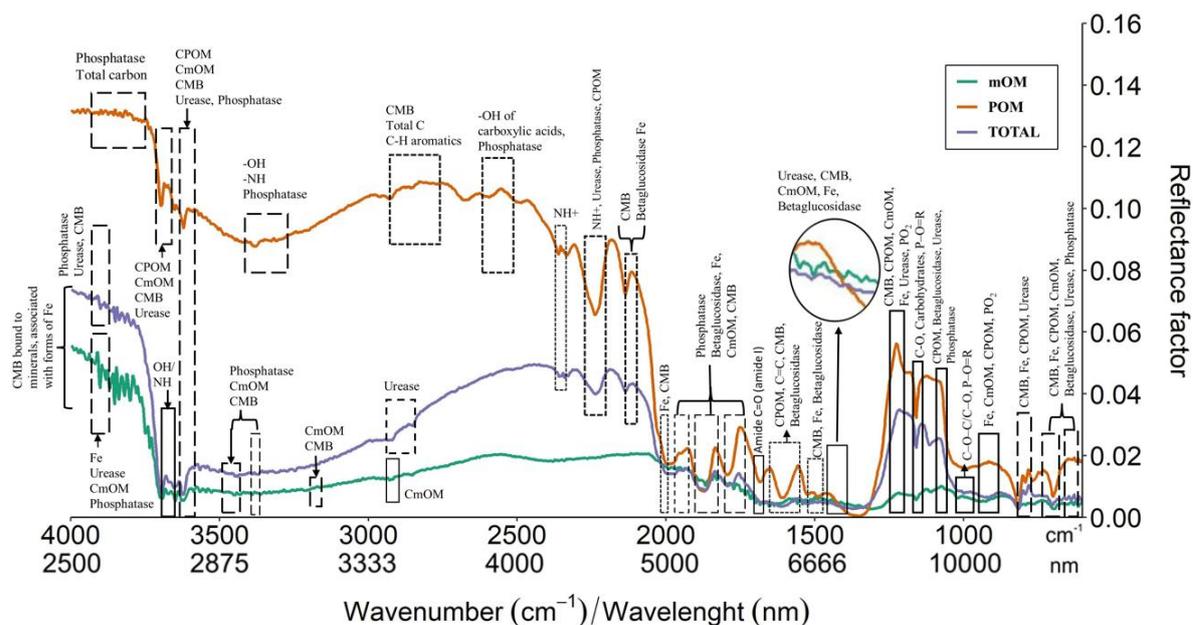


Figure 2. Characterization of the microbiological activity of the Mid IR spectra of sample 16, focused on the interpretation of the enzymatic activity of Betaglucoisidase, Urease and Phosphatase, MBC (microbial biomass carbon), POM (particulate organic matter) and mOM (organic matter associated with minerals).

When analyzing the spectra of sample P30 (Figure 3), which presents a similar textural behavior to P16 (Figure 2) but with higher C and clay content (Table 1), most of the peaks and valleys are maintained, however, the spectra of the mOM and total fractions present greater similarity between 4000 and 3000 cm^{-1} . In these fractions the response of MBC and CmOM at 3500 cm^{-1} is not observed clearly anymore, on the contrary, in the POM fraction it seems that the peaks and valleys are more pronounced, as it is also observed in the response of -NH. Additionally, the OH response of carboxylic acids and phosphatase activity observed at 2500 cm^{-1} , especially in the POM fraction in sample P16 (Figure 2), are observed less pronounced at P30 (Figure 3), but this behavior is carried over to the total and mOM fractions. Similarly, the expression of NH^+ at 2400 cm^{-1} can be clearly observed in all three fractions unlike P16 which is only observed in the total and POM fractions. Another difference is identified between 2000 and 1500 cm^{-1} , where the lowest reflectance is observed in POM and similar and differentiable behaviors of the three fractions are observed, being the mOM fraction the one that presents higher reflectance especially in the peaks where the Fe-MBC interaction is detected. The MBC, Fe and Betaglucoisidase interaction observed in the P16 spectra in the range 1500-1400 cm^{-1} is not observed to be intertwined at P30, and a higher reflectance of mOM is identified. Around the range 1200 to 1000 cm^{-1} the behavior is similar to P16, however, the peak of the C-O-C/C-O and P-O=R bonds that is present at 1000 cm^{-1} , in

agreement with that proposed by Liu et al. (2015) and Parikh et al. (2014), is not observed anymore in the P30 sample and overlapping of mOM and total is present, this is also identified from 800-600 cm^{-1} , in the response of MBC, Fe, POM, Urease, phosphatase and Betaglucosidase.

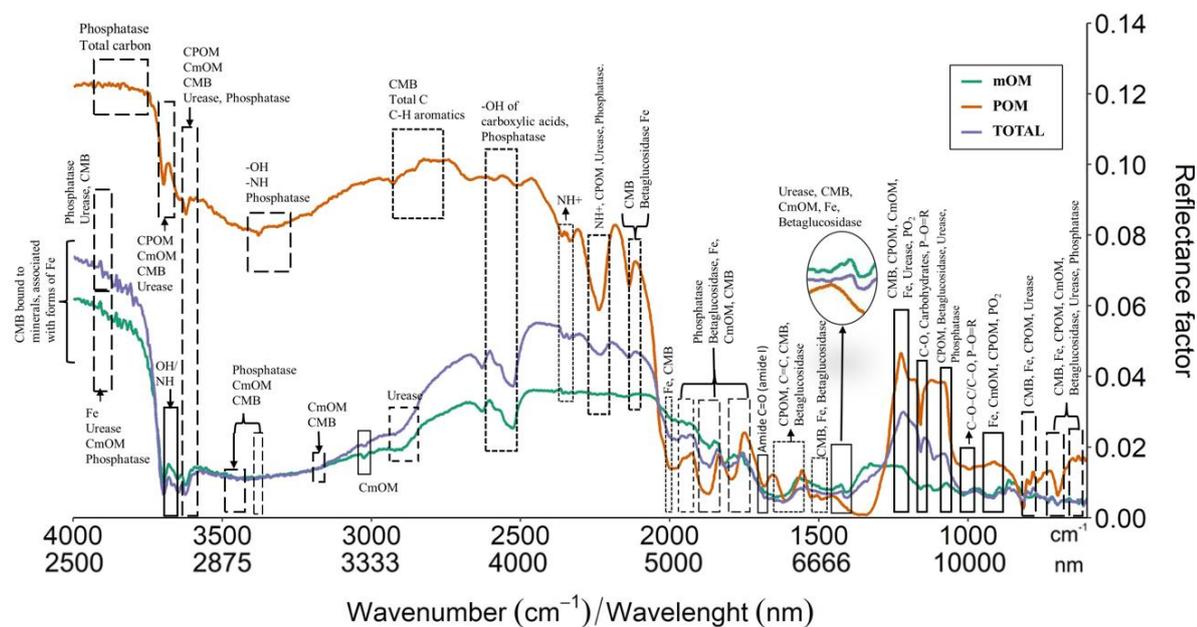


Figure 3. Characterization of the microbiological activity of the Mid IR spectra of sample 30, focused on the interpretation of the enzymatic activity of Betaglucosidase, Urease and Phosphatase, MBC (microbial biomass carbon), POM (particulate organic matter) and mOM (organic matter associated with minerals).

Sample P35 has a higher clay and C content than samples P30 and P16 (Table 1), which is reflected in the groups of spectra of the different fractions. The differences in the spectral behavior are not as marked as in the other two samples, due to the fact that it is more clayey, therefore, the concentration of particles larger than 53 μm is lower. The peaks and valleys along the spectrum are maintained with less difference in reflectance, and between the ranges 4000 to 3500 cm^{-1} the peaks are much more pronounced. The -NH response is clearly seen in the mOM and total fractions, which coincides with the POM peak associated with the labile and resistant forms of C, MBC and Urease. Similarly, the interaction response of these forms of carbon, urease, phosphatase and MBC at the 3550 cm^{-1} peak is clearly seen in all three fractions. The peaks and valleys near 3500 cm^{-1} associated with phosphatase, MBC and CmOM are observed more pronounced in the total and mOM fractions, as is the -NH and phosphatase peak of the POM fraction. In contrast, Urease activity at 2900 cm^{-1} is not observed as clearly in the mOM and total fractions, but the behavior of the aromatic CHs in

the POM fraction is more pronounced. The OH of the carboxylic acids and the phosphatase activity in the 2500 cm^{-1} range are not as clearly observed as in the other two samples. In contrast, the NH^+ response at 2200 cm^{-1} is clearer in all three fractions. The NH^+ (Liu et al., 2015), Urease, phosphatase and POM valleys observed at 2250 cm^{-1} are not observed anymore in the total fraction, however, they are maintained in the POM fraction.

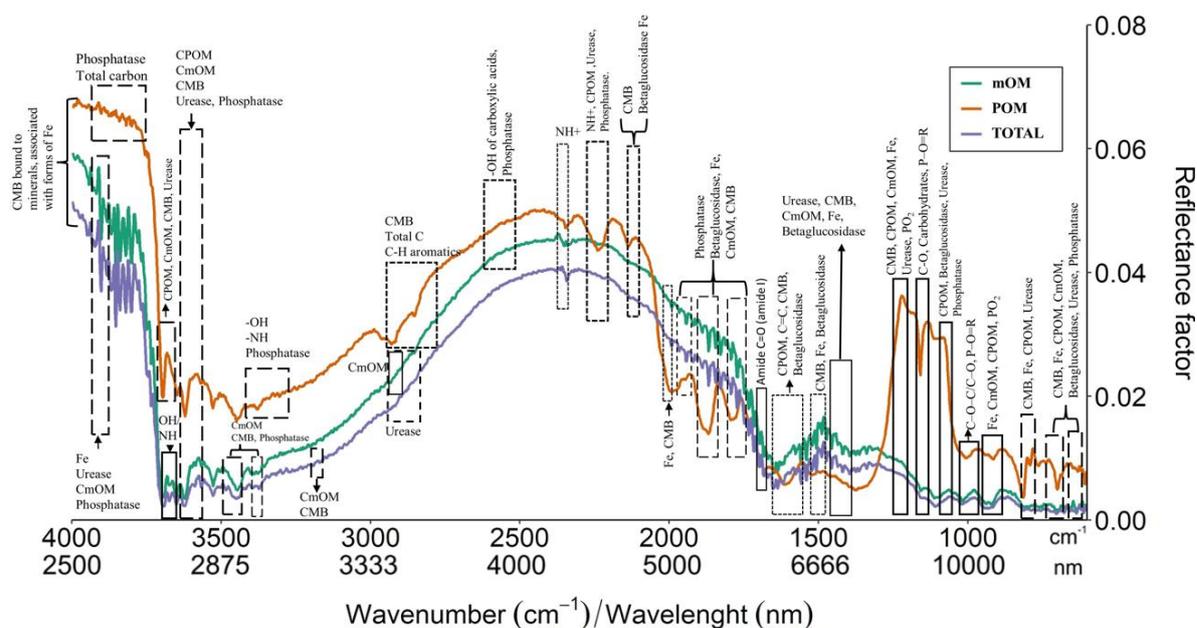


Figure 4. Characterization of the microbiological activity of the Mid IR spectra of sample 35, focused on the interpretation of the enzymatic activity of Betaglucohidase, Urease and Phosphatase, MBC (microbial biomass carbon), POM (particulate organic matter) and mOM (organic matter associated with minerals).

The peak relating MBC and beta-glucosidase at 2050 cm^{-1} in the POM and total fractions disappears in the total fraction of sample 35. The behavior between 2000 and 1700 cm^{-1} is very contrasting and a higher reflectance of the mOM fraction is maintained, as observed in sample 30 (Figure 3), however, multiple pronounced peaks and valleys are observed, highlighting the interaction of betaglucohidase, cmOM, MBC and Fe (Figure 4). The response of amide I and II is lost in the POM spectrum. Between 1550 and 1450 cm^{-1} the behavior is very different from the other two samples, the reflectance of the mOM fraction is higher and the peaks are very pronounced where the activity of betaglucohidase, MBC is highlighted. On the other hand, between 1200 and 600 cm^{-1} , the spectra behave similarly to the two previous samples, with a higher reflectance of the POM fraction and in the C-O-C/C-O and P-O=R response at 1000 cm^{-1} , there is overlapping and an increase in the peak of the mOM fraction. The reflectance below 800 cm^{-1} is very attenuated in the mOM and total fractions.

4.3.2. Spectral interpretation Vis NIR SWIR

In contrast to the Mid IR spectroscopy, the Vis NIR SWIR spectra of samples P16, P30 and P35 show similar behaviors, where only differences in reflectance values are observed, due to the differential texture of the three samples (Figure 5). The region 350 to 400 nm presents differential peaks of the activity of the three enzymes, MBC and CmOM. At 395 to 400 nm the interaction of Ureasea, MBC and CmOM is observed.

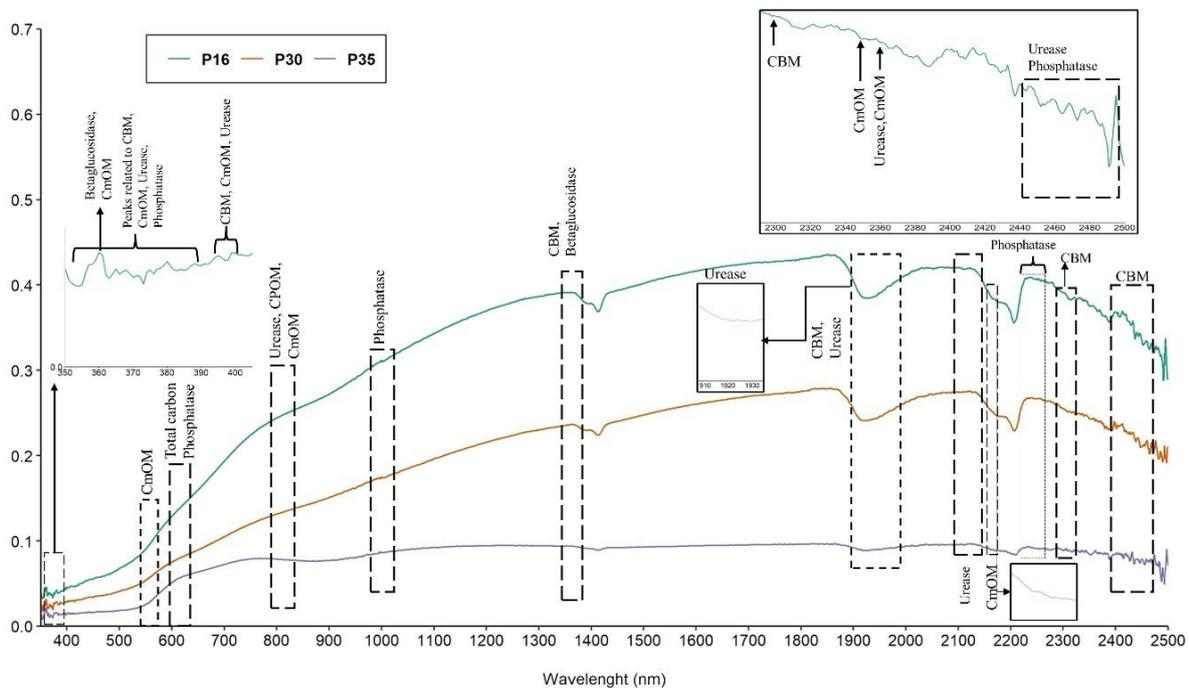


Figure 5. Characterization of microbiological activity from Vis NIR SWIR spectra of samples 16, 30 and 35, focused on the interpretation of enzymatic activity of Betaglucosidasea, Ureasea and Phosphatasea, MBC (microbial biomass carbon), POM (particulate organic matter) and mOM (mineral-associated organic matter).

An elbow is observed at 550 nm associated with CmOM, more pronounced in the more clayey sample (P35), this absorbance is gradually lost in the samples with higher sand content (Figure 5). At 1000 nm, a small peak associated with phosphatase activity is identified, which is more pronounced in the less clayey sample (P16). At 1350-1360 nm, an elbow related to MBC and Betaglucosidasea is observed, which is lost in the more clayey sample (P35).

A valley is observed in 1900 -1920 nm related to ureasea activity, which is clearly distinguishable in samples P16 and P30. Another small valley is observed at 2100 nm, associated with the activity of this enzyme (Figure 5, Table 4).

Phosphatase activity at 2250 nm is clearly observed in the three samples, identifying a series of peaks. At 2300 nm the presence of MBC is identified, which is slightly displaced in sample P30. Another region associated with this variable is observed between 2400 and 2500 nm, and between 2440 and 2500 nm valleys and peaks associated with Ureasea and phosphatasea are identified (Figure 5, Table 4).

Table 4. Bands of the Vis NIR SWIR spectrum related to the microbiological variables (MBC (microbial biomass carbon), enzymatic activity of ureasea, phosphatasea and beta-glucosidasea, POM (particulate organic matter), mOM (organic matter associated with minerals), total C), according to the results of the Recursive Feature Elimination (RFE) algorithm.

Variables	Vis NIR SWIR (nm)
MBC	350, 351, 352, 355, 356, 357, 358, 360, 361, 362, 363, 364, 365, 367, 368, 369, 370, 371, 372, 375, 376, 378, 380, 381, 382, 386, 387, 395, 398, 402, 403, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 2300, 2436, 2467, 2469, 2474, 2475, 2476, 2482, 2485, 2486, 2489, 2490, 2492
Ureasea	350, 351, 352, 355, 359, 360, 363, 366, 367, 368, 369, 370, 373, 374, 376, 379, 380, 381, 382, 383, 384, 386, 387, 393, 395, 397, 398, 400, 403, 406, 410, 555, 797, 804, 805, 814, 815, 2148, 2149, 2151, 2360, 2444, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2479, 2481, 2482, 2483, 2485, 2486, 2489, 2490, 2492, 2494, 2495, 2496, 2499
Phosphatasea	350, 351, 352, 355, 356, 357, 360, 361, 365, 367, 368, 371, 381, 576, 577, 578, 579, 580, 581, 582, 583, 667, 1911, 1912, 1913, 1914, 1915, 1916, 1918, 1921, 1922, 2426, 2469, 2470, 2474, 2475, 2476, 2477, 2482, 2483, 2484, 2485, 2486, 2490, 2491, 2493
POM	353, 790, 802, 803, 810, 812, 813, 816, 817, 819, 821
C total	1883, 1887, 618, 614, 1862
mOM	350, 351, 352, 356, 357, 358, 359, 360, 361, 365, 367, 371, 373, 375, 376, 377, 378, 379, 381, 395, 399, 400, 401, 402, 573, 577, 812, 2173, 2350, 2360, 2426, 2428, 2434, 2436, 2442, 2444, 2449, 2462, 2463, 2467, 2468, 2469, 2474, 2476, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2489, 2490, 2491, 2492, 2498, 2499, 2500
Betaglucosidasea	350, 351, 361, 360, 2500, 2499

4.3.3. Mid-IR and XRF spectromicroscopy

Based on the RFE results (Table 2 and 3) and the description of the Mid IR spectra, specific bands were identified that allowed the speciation of the activity of the microorganisms. Figures 6, 7 and 8 show the mapping of the spectral ranges with the most characteristic responses of the activity of the microorganisms for the total, POM and mOM fractions, respectively, of sample P16, which is the sandiest. Figure 9 corresponds to the total fraction of the most clayey sample, P35. The range between 600 to 1000 cm^{-1} , in which

several interactions of microorganism activity were identified, could not be mapped due to the resolution of the equipment. However, the range from 1000 to 1500 cm^{-1} , which also presented high importance, was analyzed and the 1150-1160 cm^{-1} region (Figure 6 A, 7 A, 8 A), related to Urease and phosphatase activity (Table 3, Figures 2,3,4), was mapped, where green and blue colors are related to low reflectance and orange to reddish colors to higher reflectances.

In relation to microbiological activity, the analysis of these peaks and valleys becomes a bit confusing, because some valleys relate a lower reflectance to a higher presence of the analyzed variable, on the contrary, the peaks may indicate a higher presence. For example, in the reflectance range 1800 to 1700 cm^{-1} , which is also related to phosphatase, urease and additionally POM activity, the reddish tones of the mapping of the total fraction in Figure 6B, are characterized by a higher reflectance in these zones, possibly indicating higher activity of these enzymes.

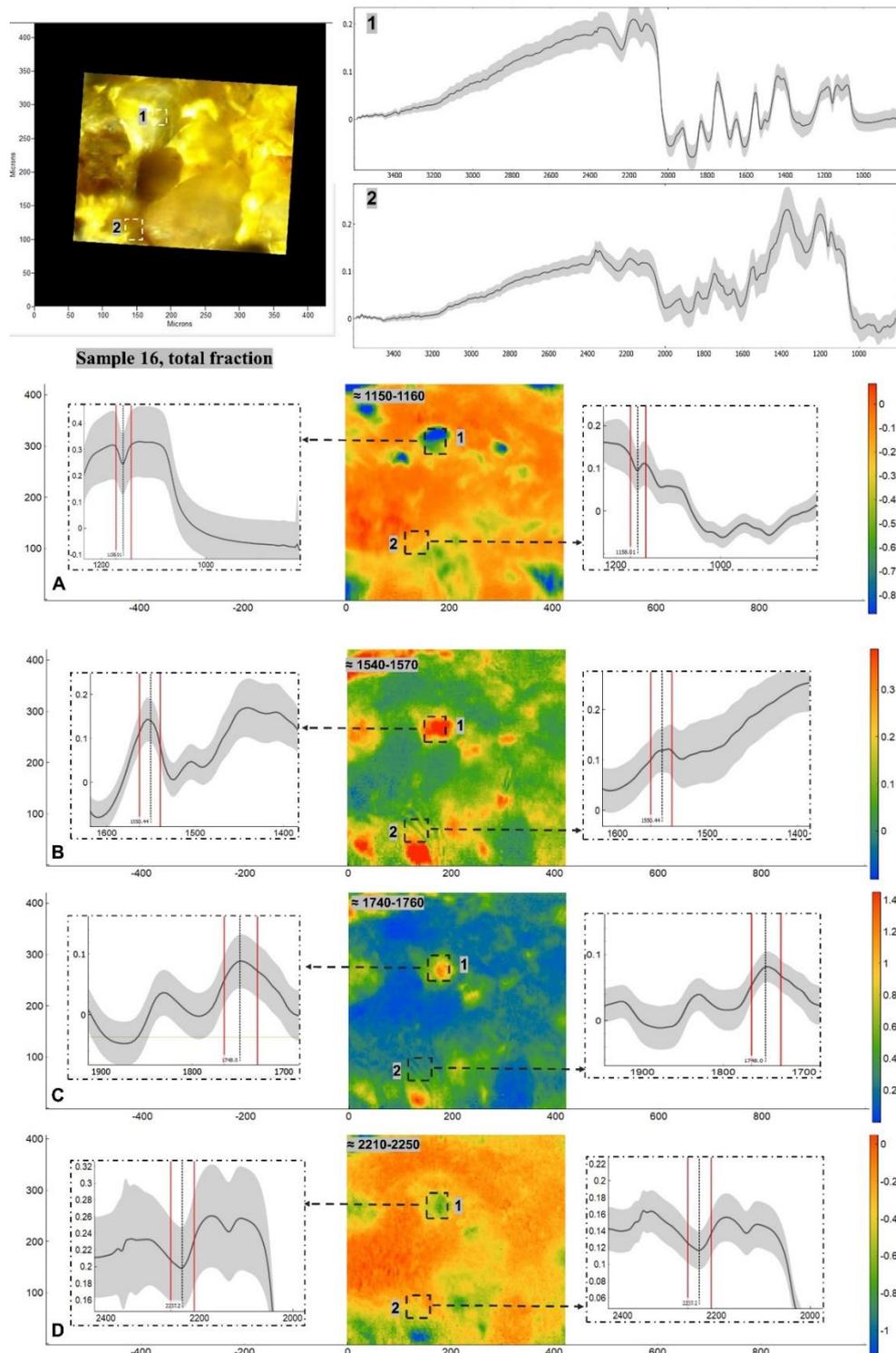


Figure 6. Relative microbiological activity distribution maps of sample 16, for the total fraction and spectra obtained from the different aggregates for regions A ($1150\text{-}1160\text{ cm}^{-1}$), C ($1740\text{-}1760\text{ cm}^{-1}$) and D ($2200\text{-}2300\text{ cm}^{-1}$), associated to the enzymatic activity of ureasea and phostatasea, and region B ($1540\text{-}1570\text{ cm}^{-1}$) related to the enzymatic activity of beta-glucosidasea, Fe, MBC, Ureasea and phosphatasea. Contrasting zones (1, 2) of the mapped particles were chosen to observe the difference in spectral response.

In general, it is observed that the highest reflectance values are located in coarser particles, associated with quartz but presenting some coating (Figures 6 A1, 7 A1, 8 A1, 9 A1). Additionally, it was identified that the spectrum of zones where the reflectance was lower, possibly associated with coatings of other compounds (Figures 6 A2, 7 A2, 8 A2, 9 A2). Figure 10, corresponding to the mapping with the μ -XRF spectrometer, allows to confirm that the coatings of the particles present in the total sample 16 (Figure 10 A), present high aluminum contents and some particles coatings with high Fe contents.

The 1540-1570 cm^{-1} region was also mapped (Figures 6 B, 7 B, 8 B, 9 B), in which the response of the enzymatic activity of betaglucosidasea, Fe, MBC, Ureasea and phosphatasea is identified (Table 3, Figures 2,3,4). The reddish color shades are related to high reflectance and this is maintained in the two mapped regions of the total fraction of sample P16 (Figure 6 B). The reflectance is lost in the POM fraction (Figure 7B), especially in the darker observed particles, because only the phosphatase response is maintained according to Table 3. The response of this enzyme is related to the more stable compounds in the bacterial walls (Omoike & Chorover 2006). In the mOM fraction of sample P16 the reflectance is lower in all regions, due to the fact that in this fraction the finest particles are present, where possibly the bonds with the organic compounds are stronger (Figure 8). In addition, as shown in Figure 10 B, this fraction has a higher content of Al and Fe, which favors organo-mineral interaction. In sample 35 (Figure 8 B) red colors are observed on the reddest particles, and in the darkest particles with some type of coating a decrease in reflectance is observed, but a more pronounced peak. It is observed that the reddest particles have high Fe and Al contents (Figure 10 D), however, it is confirmed that the darkest particles have the highest Fe content.

In general, it is observed that the POM fraction (Figure 7), presents higher reflectance due to the fact that there are only coarse particles larger than 53 μm and the valleys appear to be more pronounced, however, the reflectance decreases in those particles that appear to have some coating, as occurs in regions B and C. On the other hand, the identification of activity in region A (1150 to 1160 cm^{-1}) is not as clear for ureasea, but the response of phosphatasea is maintained according to Table 3. The peaks in region C (1740-1760 cm^{-1}) are also maintained for phosphatasea, but disappear for Ureasea. When analyzing sample 35 (Figure 9), strong C-region reflectances are maintained, which could confirm the quality of organo-mineral bonds, remembering that phosphatasea activity is related to the phosphate groups of bacterial walls (Parikh et al., 2014).

The valley at 2230 cm^{-1} related to ureasease and phosphatase enzyme activity is also observed for the total and POM fractions (Figures 6 D and 7 D, respectively). For the total

fraction, a valley is present indicating that lower reflectance is associated with higher activity. However, for the POM fraction a valley is displaced, and a peak is observed (Figure 7B). Overall, it is possible to confirm the presence of this enzyme activity as a coating on the mineral particles. The urease response is lost in POM possibly because the microorganisms obtain food and energy from this fraction (Parikh et al., 2014).

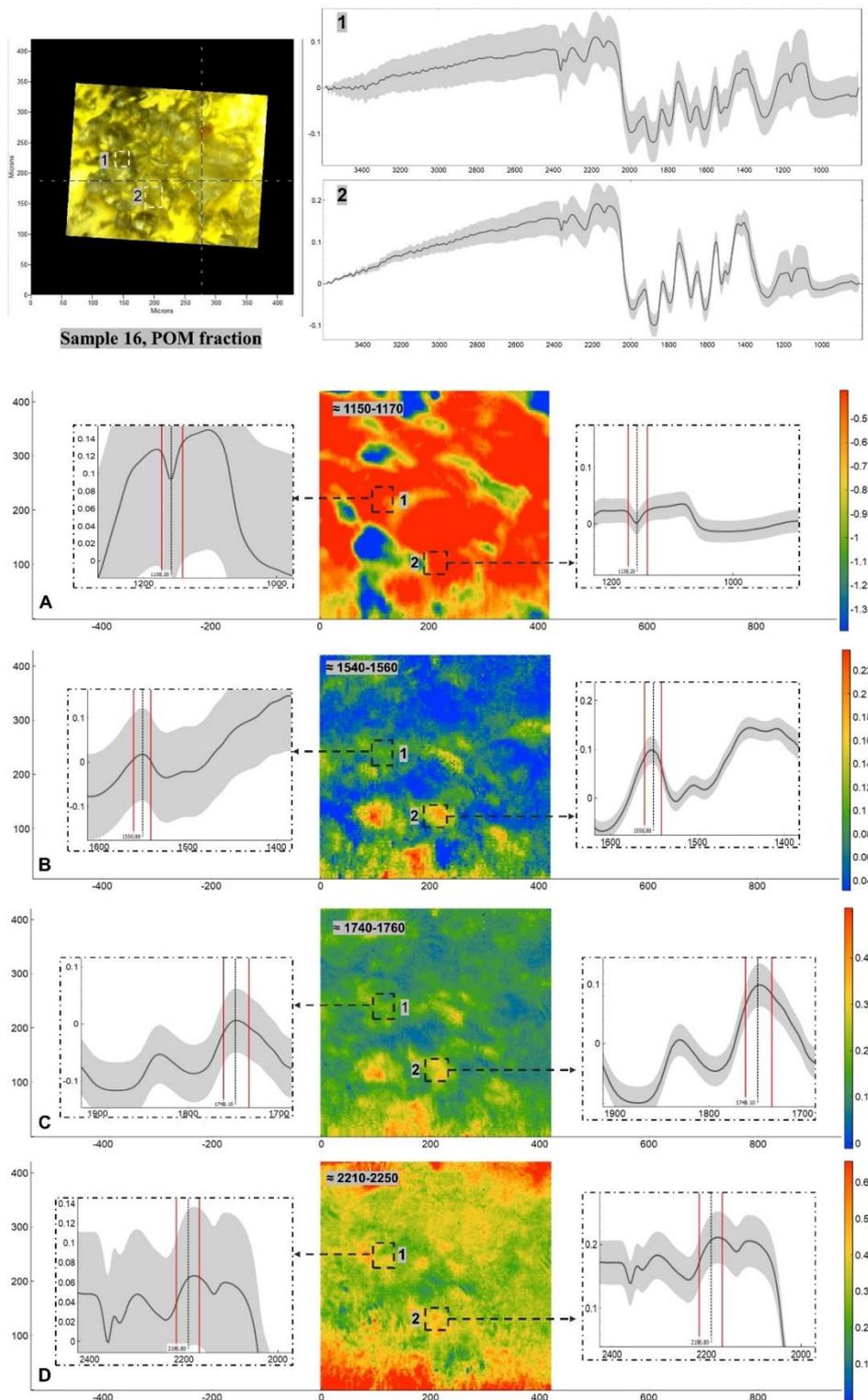


Figure 7. Relative microbiological activity distribution maps of sample 16, for the POM fraction and spectra obtained from the different aggregates for regions A ($1150-1160\text{ cm}^{-1}$), C ($1740-1760\text{ cm}^{-1}$) and D ($2200-2300\text{ cm}^{-1}$), associated to the enzymatic activity of ureasea and phostatasea, and region B ($1540-1570\text{ cm}^{-1}$) related to the enzymatic activity of beta-glucosidasea, Fe, MBC, Ureasea and phosphatasea. Contrasting zones (1, 2) of the mapped particles were chosen to observe the difference in spectral response.

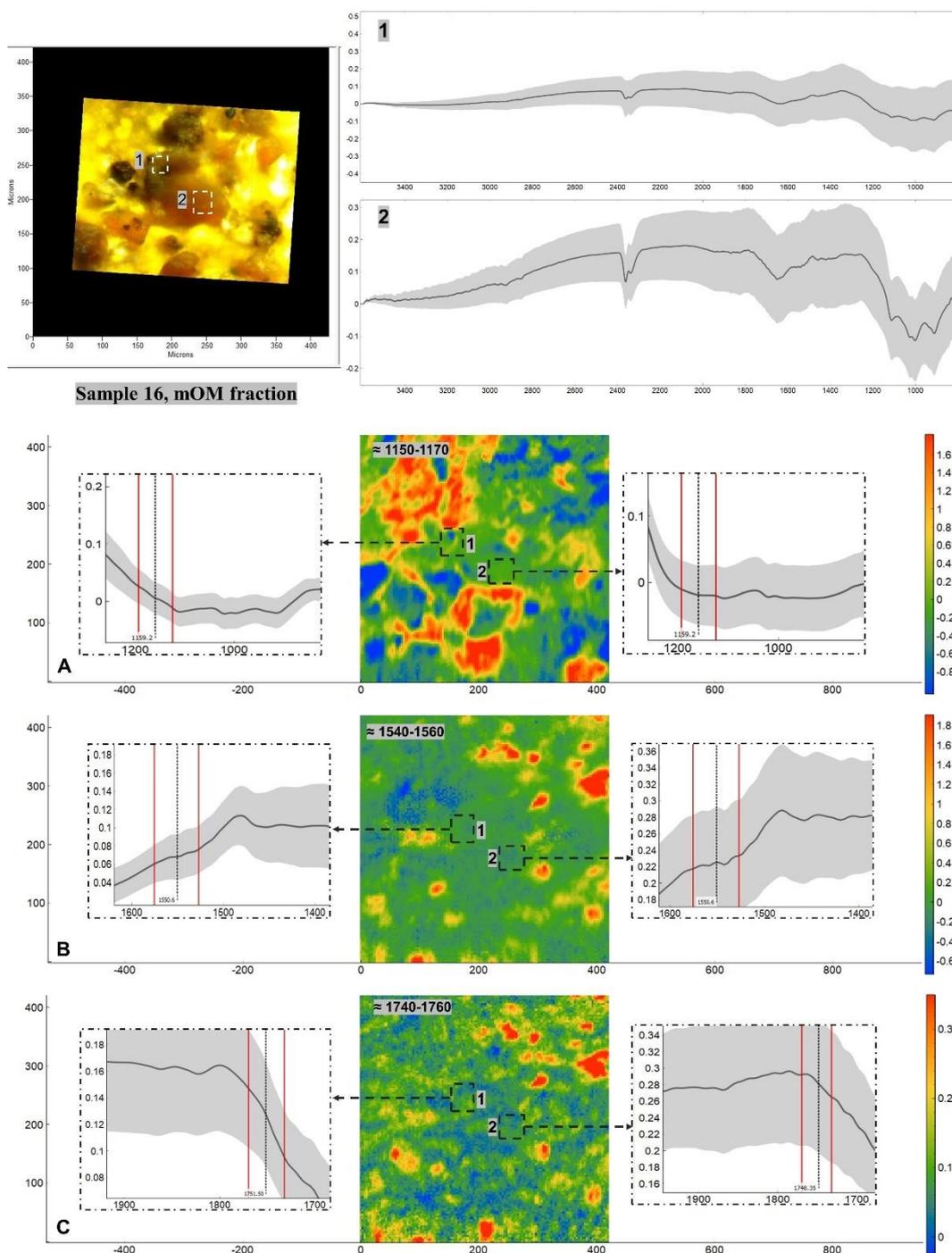


Figure 8. Relative microbiological activity distribution maps of sample 16, for the mOM fraction and spectra obtained from the different aggregates for regions A ($1150-1160\text{ cm}^{-1}$), and C ($1740-1760\text{ cm}^{-1}$), associated to the enzymatic activity of urease and phosphatase, and region B ($1540-1570\text{ cm}^{-1}$) related to the enzymatic activity of beta-glucosidase, Fe, MBC, Urease and phosphatase. Contrasting zones (1, 2) of the mapped particles were chosen to observe the difference in spectral response.

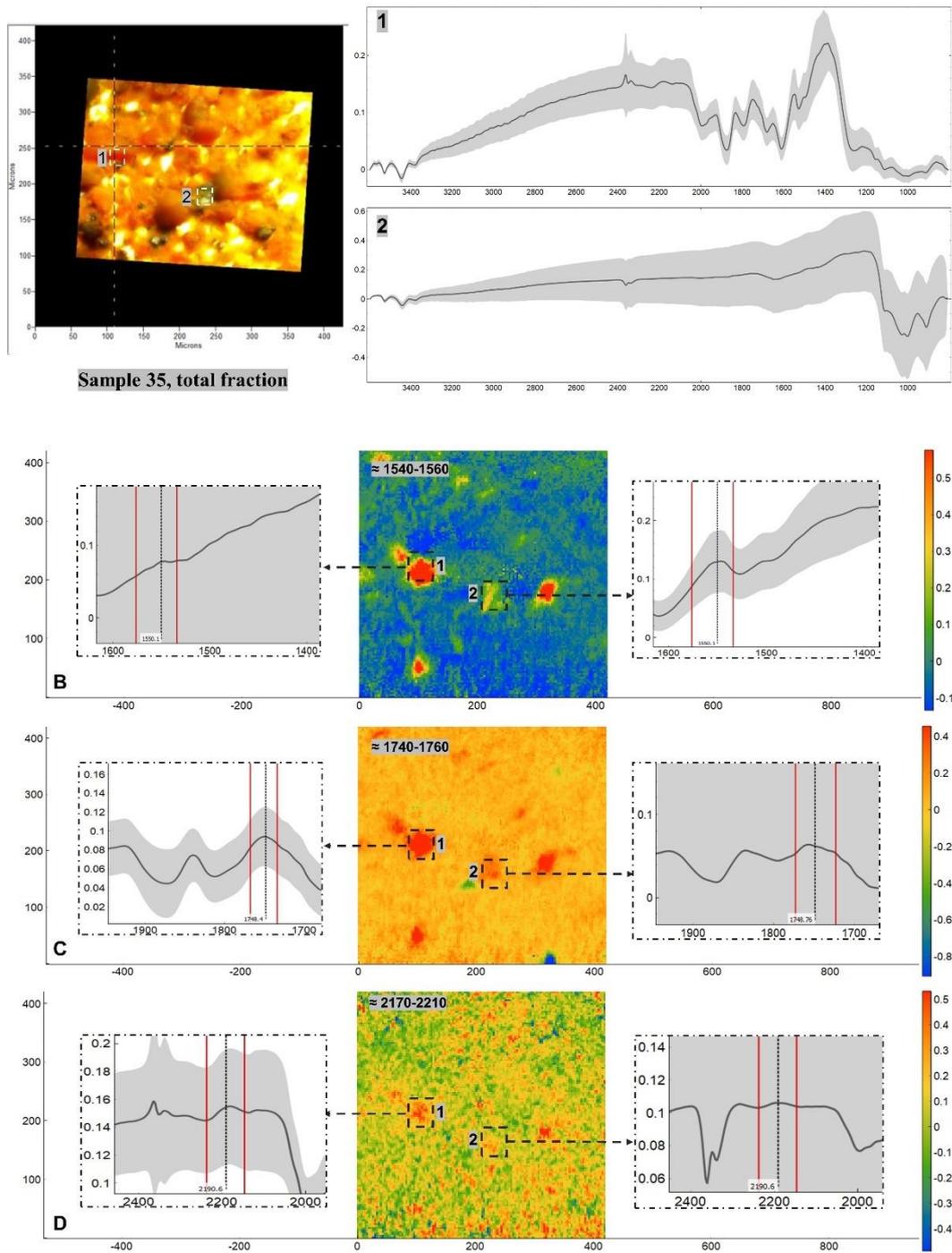


Figure 9. Relative microbiological activity distribution maps of sample 35, for the total fraction and spectra obtained from the different aggregates for regions C ($1740\text{-}1760\text{ cm}^{-1}$) and D ($2200\text{-}2300\text{ cm}^{-1}$), associated to the enzymatic activity of ureasea and phostatasea, and region B ($1540\text{-}1570\text{cm}^{-1}$) related to the enzymatic activity of beta-glucosidasea, Fe, MBC, Ureasea and phosphatasea. Contrasting zones (1, 2) of the mapped particles were chosen to observe the difference in spectral response.

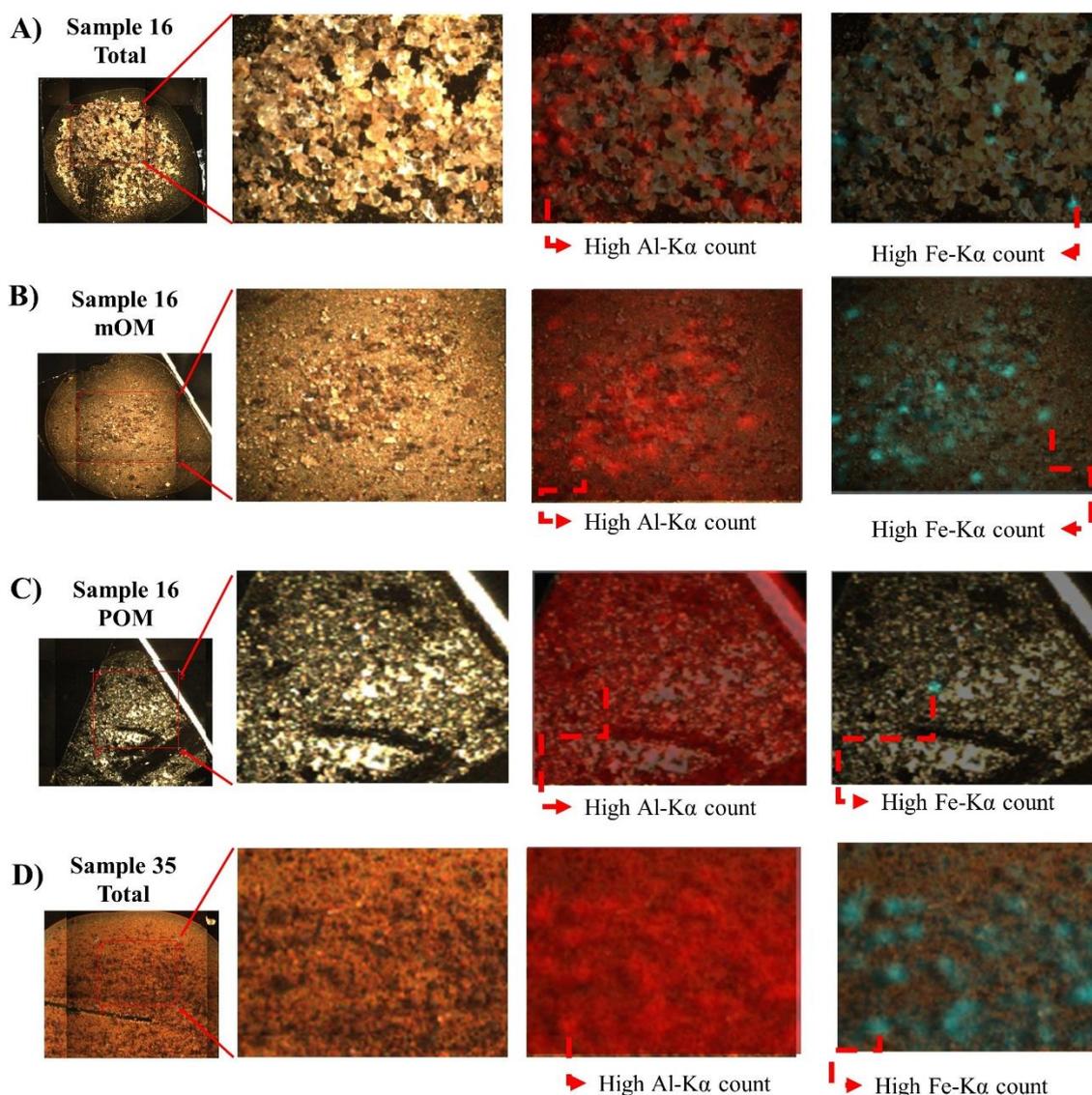


Figure 10. Images of the μ -XRF spectrometer with microprobe for soil samples 16 A) total fraction, B) mOM fraction, C) POM fraction. D) Sample 35 total fraction. Counts of aluminum (Al) (reddish images) and iron (Fe) (bluish images).

4.4. Discussion

4.4.1. Mid IR spectral interpretation

Absorption features associated with organic compounds are identified at 3400, 2930, 2850, 1600 and 1400 cm^{-1} according to Nguyen et al. (1991) and Viscarra Rossel et al. (2006), such zones can be observed in Figures 2, 3 and 4, related to the activity of microorganisms, however, high activity is also observed in the region below 1000 cm^{-1} . According to Ammann & Brandl (2011) the range between 650 - 1800 cm^{-1} is known as the typical fingerprint region for the prediction of microorganism activity, because it includes the response of cellular

carbohydrate and protein compounds. As seen in Figures 2, 3, and 4, this region identifies the spectral response of enzyme activity in general and its interaction with MBC and labile and resistant forms of carbon.

The region between 1800 to 2500 cm^{-1} is also important for the spectral differentiation of microorganism activity (Figures 2,3,4), and according to Parikh et al. (2014) this spectral range is characterized by the response of more stable organic compounds. For Davinic et al. (2012) the 2000-1750 cm^{-1} bands present a strong influence of silicates. Calderón et al. (2011) also highlights that in the region 1790 to 2000 cm^{-1} three characteristic peaks related to silicates are present, thus reporting negative correlations with soil total C and N. In Figures 2, 3 and 4 we see the peaks associated with silicates, in which important relationships of enzymatic activity with Fe were identified, in addition to specific regions associated with MBC and CmOM, considered as stable forms of C, which is in agreement with that reported by Davinic et al. (2012), Calderón et al. (2011) and Parikh et al. (2014).

Betaglucosidase is strongly related to the C cycle, being responsible for catalyzing the hydrolysis and biodegradation of glycosides (Martinez and Tabatabai, 1997), which allows obtaining glucose as a final product, which is an important source of C energy for the activity of microorganisms (Merino et al., 2016). Therefore, the activity of this enzyme is related to the presence of simple sugars, frequently used by microorganisms (Adetunji et al., 2017). The bands of highest significance associated with betaglucosidase activity (Table 3) are within the range reported by Rasche et al. (2013) (2301-1279 cm^{-1}) for this enzyme activity, however, the author does not report activity of this enzyme in the range 600 to 700 cm^{-1} , however, this region is related to carbohydrate and protein response (Ammann & Brandl, 2011). According to Viscarra Rossel et al. (2022) the functional groups CH, NH of amines and CO of carbohydrates may indicate the presence of labile forms of C. As observed in Figures 2, 3 and 4 between 1200 and 1300 cm^{-1} the interaction of the 3 enzymes is present, in addition to the response of carbohydrates and phosphates. According to Gao & Chorover (2009) in 1337-1313 cm^{-1} the influence of Amide III (NH, CN, CH, NH) is identified, and according to Nicolas et al. (2017), the general activity of aliphatic groups is observed in the range 1350-1450 cm^{-1} .

Parikh et al. (2014) highlights the influence of the C=O ester group in the range 1330 to 1315 cm^{-1} , in 1300- 1340 cm^{-1} diamond C=C and in 1300 cm^{-1} the CH overtone. The latter has also been reported by Liu et al. (2015) in the range 1379-1384 cm^{-1} .

As shown in Table 3 the range between 600 to 700 cm^{-1} involves some of the bands with greater importance in the prediction of the different variables analyzed in the spectra of

Figures 2, 3 and 4. For Parikh et al. (2014), the range 650-450 cm^{-1} responds to the vibration of CH_2 stretches of bacterial cellular compounds. According to Demyan et al. (2012) in the range 690-670 cm^{-1} the presence of clay and quartz minerals interacting with carbonyl groups is recognized. Therefore, it is possible that in this spectral range this interaction of microbial activity with soil minerals is being observed, such interaction is maintained in the three fractions. Deo et al. (2001) confirm the interaction of amide IV of bacterial cells with the AIO groups of clay minerals, in the peak 696 cm^{-1} . The phosphatase response in this range may also be due to the presence of phosphate groups in the cell walls that favor binding to the oxide surfaces (Parikh et al., 2014), which explains the Fe response in this spectral region.

It is possible to observe that near the 1000 cm^{-1} range the response of the phosphate groups is intensified (Figures 2,3, 4). According to Ojeda et al. (2008) in the 1084 cm^{-1} band the stretching of P=O groups of polyphosphate products, nucleic acid phosphodiester and phosphorylated proteins is recognized. Gao & Chorover (2009) report in 1080 cm^{-1} , COC, C-C and PO_3^{2-} vibrations. Liu et al. (2015) mentions in the range 1000-1100 cm^{-1} the influence of P-O=R features and in the range 1006-1122 cm^{-1} the C-O stretching. Lad et al. (1993) indicates in the band 1050 cm^{-1} the influence of carbohydrates. For Parikh et al. (2014) the range 1000-1080 cm^{-1} is an indicator of POM cellulose, the range 1160-1020 cm^{-1} reflects the influence of polysaccharides such as cellulose lignin or pectin, and on 1085 cm^{-1} CO and CN stretching are identified, additionally, the authors point out the influence of POFe bonds in the bands 1037, 1045 and 1027 cm^{-1} , product of the interaction of the phosphate groups of the bacterial walls with the surfaces of the metal oxides, thus they specifically report PO phosphate groups in 1075-1028 cm^{-1} , and the phosphorus present in humic fractions in 1100-1000 cm^{-1} . The above, responds to the influence of Fe in the last peak (Figures 2, 3 and 4) and the interaction of Urease due to the presence of nitrogenous compounds. Cao et al. (2011) report in the 1031 cm^{-1} band the symmetric stretching PO, and in 1051 cm^{-1} the PO associated with nucleic acids and the C-OH stretch of phosphorylated proteins, which are shifted to 1049 and 1092 cm^{-1} when they bind to the goethite.

Nicolas et al. (2017) also points out the stretching of carbohydrates in the 1100 cm^{-1} range, which is possible to identify in Figures 2,3 and 4.

Ludwig et al. (2008) indicates that the large range 1057-397 cm^{-1} is related to nitrogenous compounds, which may explain the urease activity observed between 1300 and 600 cm^{-1} (Figures 2, 3 and 4).

Between the range 1500 to 1600 cm^{-1} , interaction of MBC, Fe and betaglucosidase is observed, in addition to the presence of C=C and CPOM bonds (Figures 2,3 and 4). Nicolás et

al. (2017) identified aromatic groups in the 1510 cm^{-1} band. For Ladd et al. (1993), these aromatic groups are specifically related to aromatic C=C bonds at $1630 - 1580\text{ cm}^{-1}$, the authors also point out the influence of amide II at 1550 and lignin at 1500 cm^{-1} . Parikh et al. (2014) reports in the bands $1590-1500\text{ cm}^{-1}$ the influence of Amide II (NH C=N), in 1570 cm^{-1} aromatic CH deformation and in $1550-1500$ and $1600-1580\text{ cm}^{-1}$ aromatic C=C stretching. Cao et al. (2011) have also reported at $1540 - 1549\text{ cm}^{-1}$ the influence of amide II. Omoike & Chorover (2006) report this influence at $1544-1516\text{ cm}^{-1}$. According to Cao et al. (2011) and Mc Whirter et al. (2002) the 1550 cm^{-1} band of amide II may exhibit shifts upon interaction with the surface of iron oxides such as goethite. It is possible that the Fe, MBC and betaglucosidase ratio observed in Figures 2, 3 and 4 is due to this interaction of organic compounds with the surface of Fe oxides.

Between 1800 and 2000 cm^{-1} where the peaks associated with silicates described by Davinic et al. (2012) and Calderón et al. (2011) are observed (Figures 2,3 and 4), interactions between SiO bonds of quartz minerals and carbonyl stretching are also present, specifically in the range $1980-1870\text{ cm}^{-1}$ according to Demyan et al. (2012). A strong influence of CmOM and MBC is observed in these peaks, thus confirming that interaction of carbon forms with the mineral fraction. Rasche et al. (2013) reports a wide range $1980-1870\text{ cm}^{-1}$ associated with MBC, identified in Figures 2, 3 and 4 and within which the most important bands for variable prediction are located (Table 3).

Between 2000 and 2500 cm^{-1} a region with influence of amino groups, interaction of the 3 enzymes, Fe, MBC and carbon labile forms (CPOM) is identified. Liu et al. (2015) points out the influence of NH^+ in the region $2322-2360\text{ cm}^{-1}$, clearly identified in the spectra of Figures 2,3 and 4.

The region 2600 to 2700 cm^{-1} that relates to phosphatase activity (Figures 2, 3 and 4) can be related to the range $2600- 2500\text{ cm}^{-1}$ reported by Parikh et al. (2014), associated with the H bridges of the OH bonds of carboxylic acids. Rasche et al. (2013) reports in the range $2640-2399\text{ cm}^{-1}$ enzymatic activity in general.

Between 3500 to 3400 cm^{-1} phosphatase interaction with amino groups -NH, hydroxyl (-OH) as well as MBC and CmOM is observed (Figures 2, 3 and 4). Davinic et al. (2012) identified in the 3400 cm^{-1} band OH and NH stretching related to the presence of labile forms of C. According to Jiang et al. (2004) hydroxyl, carboxyl, phosphoryl and amide groups are common in bacterial cell walls, and the negative charges that allow interaction with the surface of minerals result from the deprotonation of carboxyls and phosphates. Rasche et al. (2013) report another important MBC range between 3658 and 3317 cm^{-1} . In the present study

we identified specific bands in this wide range that are observed in Figures 2,3 and 4 and in Tables 2 and 3.

Between 3500 - 3900 cm^{-1} an important region is also observed where high interaction of urease, phosphatase, Fe activity and labile and resistant forms of carbon as well as MBC are identified (Figures 2, 3 and 4). According to Liu et al. (2015), a strong influence of mineralogy is observed in this region, with high dominance of structural OH hydroxyl groups of clay minerals, in the bands 3591-3626 cm^{-1} . It is possible that we have identified specific regions of organo-mineral associations because MBC and enzymatic activity in general are strongly associated with the stretching vibrations of the hydroxyl (OH) groups of gibbsite and kaolinite between 3700 to 3500 cm^{-1} (Parikh et al. 2014, Wang and Johnston 2000). Additionally, Silvero et al. (2020) indicates that the range between 3700 and 3400 cm^{-1} is characterized by related absorptions associated with Al-OH bonds. Davinic et al. (2012) reports in the range 3700 to 3200 cm^{-1} , interactions of OH and NH groups, which could explain the influence of urease in these regions (Figures 2, 3 and 4). Such OH and NH groups in this region are part of humic and fulvic acids (Parikh et al. 2014, Davinic et al. 2012), so they respond to more stable forms of C.

4.4.2. Spectral interpretation Vis NIR SWIR

In relation to the Vis NIR-SWIR spectral range in the 350 to 400 nm region, the interaction of the enzymatic activity of the three enzymes, MBC and CmOM, is observed (Figure 5). Sharma et al. (2021) reported in the 355 and 395 nm peaks the interaction of fulvic acids. Yang et al. (2022) reported at 390 nm the interaction of fungi with iron oxides. It is possible that in this region it relates the more stable forms of C interacting with the surfaces of Fe oxides.

Sharma et al. (2021) points out that the range 600-670 nm is related to strongly humified material with high presence of aromatic groups. In Figure 5 and Table 3, it is observed that the range between 600 to 620nm is associated with total C content and phosphatase activity. According to Kirsten et al. (2021), this more stable mOM-related material contributes a higher proportion to the total COS content. Older and processed C are generally related to the clay fraction and contain more than half of the total soil C (Parikh et al., 2014).

According to Table 4 and Figure 5, labile forms are related in the range 790-821nm. Stenberg et al. (2010) has reported NH, CH and CO relationships in the 820nm peaks, and at 825 nm recognizes the response of aromatic compounds (Viscarra-Rossel & Behrens, 2010).

This presence of NH groups may explain the response of Urease in this range of the spectrum.

In the 1000 nm band, the phosphatase response identified by Zornoza et al. (2008) (1014-1195nm) was recognized. Viscarra-Rossel and Behrens (2010) report the presence of amines (1000, 1500 nm) and aromatic bonds (1100 nm) in this range.

Zornoza et al. (2008) report the influence of beta-glucosidase activity in the range 1372-2272nm and MBC in 1374-2092nm. Figure 5 identifies the influence of MBC and beta-glucosidase between 1360-1380nm.

The activity of urease and MBC identified between 1900 and 2000nm (Figure 5, Table 4), responds to that proposed by Zornoza et al. (2008) who report the activity of this enzyme between 1912-2094nm. Viscarra-Rossel and Behrens (2010) indicate at 1930nm the influence of carboxylic acid and at 1961nm polysaccharides. Another urease response is also observed between 2100 to 2150 nm, which coincides with the response of amine NH functional groups, phenolic OH, and aliphatic CH reported by Fidêncio et al. (2002) and Cozzolino & Morón (2003) in the 2200nm band, which could also explain the CmOM response at 2173nm. Viscarra-Rossel and Behrens (2010) indicate the response of polysaccharides at 2137nm. Additionally, fungal and clay mineral interactions have been reported in the range 2140-2150 nm (Yang et al., 2022).

Zornoza et al (2008), reports an additional range for phosphatase between 2270-2630 nm, which may explain the response of this enzyme observed between 2200 to 2300 nm (Figure 5). Additionally, at 2300 nm it is also possible to observe the response of MBC. According to Viscarra-Rossel & Behrens (2010) at 2336 nm the interaction of carbonates is observed, at 2381 nm carbohydrates and in the range 2307-2469nm methyl functional groups.

In the ranges 2340 -2360 nm the influence of urease and CmOM is observed and above 2440nm the response of MBC, Urease and phosphatase (Figure 5). According to Yang et al. (2022) at 2450 and 2440nm the interaction of the methyl CH group is recognized. Viscarra-Rossel & Behrens (2010) recognize the presence of fulvic acids and carbohydrates at 2381nm. In general, Zornoza et al (2008) report the ranges 2270-2632nm for the influence of bacteria, fungi, actinomycetes and vesicular-arbuscular mycorrhizal fungi.

4.4.3. Mid-IR and XRF spectromicroscopy

As named in the section associated with the interpretation of Mid IR spectra, the range between 650-1800 cm^{-1} represents the typical fingerprint region for microorganisms (Ammann & Brandl, 2011). Figures 6, 7, 8 and 9 refer to the mapping of the ranges of that

region, corresponding to 1150-1160 cm^{-1} (region A), 1540-1560 cm^{-1} (region B) and 1740-1760 cm^{-1} (region C). The ranges of region A and C are related to the activity of the enzymes urease and phosphatase, and region B to the activity of β -glucosidase (Table 3). Different authors have identified in the 1000-1200 cm^{-1} region the presence of sugars (De Beeck et al., 2021), carbohydrates (Lad et al. 1993, Nicolas et al., 2017), P=O stretch (Ojeda et al. 2008, Cao et al. 2011, Liu et al., 2015), COC, C-C and PO₃²⁻ vibrations (Gao & Chorover, 2009), C-O stretch (Liu et al., 2015), polysaccharides such as cellulose lignin or pectin, CN stretching, POFe bonds resulting from the interaction of bacterial wall functional groups and Fe oxide surfaces (Parikh et al. 2014, Cao et al. 2011). It is to be expected that the interaction observed in the figures corresponds to compounds of the bacterial walls bound to the surface of the minerals, since in Figure 10, it is confirmed that the darker coatings stand out for high contents of Al and Fe, which explains why in the POM fraction (Figure 7) the phosphatase response is maintained and the urease response disappears, since this phosphatase activity is strongly related to the most stable compounds of the bacterial activity (Omoike & Chorover 2006, Parikh et al. 2014) and urease activity represents the energy contribution to microorganisms that is most easily used in this fraction. Parikh et al. (2014) demonstrated that coarser fractions are poor in the activity of most oxygen-containing functional groups, since it implies greater decomposition of organic compounds due to their easy access (Witzgall et al 2021).

Region B (1540-1560 cm^{-1}) is related to β -glucosidase, urease, phosphatase, MBC and Fe activity (Table 3). Aromatic C=C bonds (Ladd et al., 1993), amide II (Cao et al. 2011, Parikh et al. 2014), aromatic CH deformation (Parikh et al., 2014) have been identified in this region. According to Cao et al. (2011) and Mc Whirter et al. (2002) in this region the interaction of amide II with the surfaces of Fe oxides is identified. This could explain the loss of reflectance in the particles with higher coating observed in Figure 7 B and in the reddish particles of the mOM fraction of sample P16 and the total fraction of P35, due to the high Al and Fe contents in these coatings (Figure 10).

The C region between 1700 to 1800 cm^{-1} noted in Figures 6, 7, 8 and 9, associated with the response of Urease, Phosphatase and POM, according to Parikh et al. (2014) and Nicolas et al. (2017), is related to a high presence of carboxylic acids (1778 cm^{-1}) and carbonyl (1700 cm^{-1}), which are oxygen-containing groups that can be used by microorganisms. According to Calderon et al. (2011) and Davinic et al. (2012) in this region, characteristic peaks associated with silicates are present. Therefore, it is possible to observe this organo-mineral interaction, with higher enzymatic activity in the total fraction (Figure 6), which is

reduced in the POM fraction (Figure 7), since the functional groups that are recognized in this range contain oxygen and in the POM fraction are characterized by being easily available to microorganisms, so they have to be depleted quickly (Parikh et al., 2014).

The 2200-2300 cm^{-1} region is also observed, in which the urease and phosphatase response is also characterized in the total fraction (Figure 6), but the urease response is lost in the POM fraction (Figure 7), due to the depletion of the functional groups associated with this enzymatic activity. According to Silvero et al. (2020) in the 2230 cm^{-1} region Si-O stretching is present. For Liu et al. (2015) at 2300 cm^{-1} characteristic NH^+ peaks are presented. It is then identified in these regions of Figures 6, 7, 8 and 9 the interaction of minerals with organic compounds, product of the activity of microorganisms, also remembering that carboxyl, phosphate and amide groups are constituents of the cell walls of bacteria (Jiang et al., 2004).

Kleber et al. (2021) indicate that there is competition between the different organic compounds for adsorption sites with inorganic ions, prevailing those stronger bonds, such as covalent or ionic interactions where the chemical state of the organic molecule changes, as usually occurs in bindings with metal oxides, silicates and clays 1:1. Clearly the most stable compounds of bacterial activity associated with phosphatase activity are the most prevalent (Figures 6, 7, 8, 9 and 10) due to the quality of PO bonds with minerals (Omoike & Chorover 2006, Parickh et al. 2014).

According to Kirsten et al. (2021) kaolinite, gibbsite, goethite and hematite present functional groups reactive towards OM under acidic pH conditions, which is common in tropical soils. Cao et al. (2011), highlights the interaction between extracellular enzymes, which contain traces of proteins and nucleic acids with clay minerals and goethite, through covalent bonds and van der Waals association. Clearly most studies highlight the interaction goethite and bacterial cell wall compounds (Ojeda et al 2008, Gao and Chorover 2009, Cao et al 2011, Parikh et al 2014), however, it cannot be ignored that both goethite and gibbsite have high density of reactive adsorption sites that favor organo-mineral interactions (Kaiser and Guggenberger, 2000). As shown in Figures 6, 7, 8 and 9, the reddish and grayish coatings on the particles favor these bonds, and it is confirmed in Figure 10 that these compounds have high Fe and Al contents.

4.5. Conclusions

It was observed that the spectral peaks and valleys reported for the functional groups CH, NH, CO, COH, CO and PO are related to the enzymatic activity of betaglucosidase, Urease

and phosphatase. Mid-IR spectral characterization allowed to recognize 4 important regions for the differentiation of microbiological activity, the region 600 - 1800 cm^{-1} in which cellular compounds containing carbohydrates and proteins were detected, and a strong spectral response of enzymatic activity and its interaction with MBC, POM and mOM was observed. Near 1000 cm^{-1} the response of phosphate groups was intensified. The 1800-2500 cm^{-1} region in which important relationships between enzyme activity and Fe, and specific bands of MBC and C of mOM were identified. Between 2000 and 2500 cm^{-1} a strong influence of amino groups was also identified. The 3700 to 2800 cm^{-1} region in which the interaction of phosphatase with amino groups, MBC and C of mOM was observed. In addition, a strong interaction of clay minerals and enzyme activity was observed.

Throughout the Mid IR spectrum, it was recognized that compounds associated with the cell walls of microorganisms such as phosphate and amino groups, present high interaction with Fe and with the bands reported in the literature for different clay minerals and Fe and Al oxides.

The Vis NIR SWIR spectral characterization was more punctual and did not present strong overlapping of the different variables, as occurred in Mid IR, therefore it is possible that it has greater specificity in this spectral range, also a strong differentiation between samples was not observed, the greatest differentiation corresponds to variation of the reflectance value, so that the different bands identified could be recognized in any soil sample.

In general, the Vis NIR SWIR and Mid IR characterization, allowed to corroborate that β -glucosidase is related to labile compounds of the SOM so pronounced peaks were identified in the POM fraction, whose amplitude is considerably reduced when analyzed in the mOM fraction. Urease activity was related to NH functional groups, and its detection was also characteristic in the POM fraction. Phosphatase activity is strongly related to phosphate groups (PO) and contains the most stable compounds of the cell walls, so it is also related to amines and aromatic bonds and prevails in the different fractions, due to the quality of the organo-mineral bonds.

It was corroborated that the soil particles present Al and Fe coatings that favor the interaction with the extracellular enzymes. Additionally, it is concluded that the POM fraction includes the organic compounds that activate the enzymatic degradation, and the spectral analysis of this fraction favors the discrimination of the different enzymes, without discarding that a joint analysis of the fractions helps in the identification of the activity of the microorganisms that together with regression and machine learning techniques will favor the

quantification and characterization of the microbial activity and the understanding of the destiny of the C forms.

References

- ADAMCZYK, B., SIETIÖ, O.M., STRAKOVÁ, P., PROMMER, J., WILD, B., HAGNER, M., PIHLATIE, M., FRITZE, H., RICHTER, A. & J. HEINONSALO. 2019. Plant roots increase both decomposition and stable organic matter formation in boreal forest soil. *NATURE COMMUNICATIONS* (2019) 10:3982. <https://doi.org/10.1038/s41467-019-11993-1>
- ADETUNJI, A., LEWU, F., MULIDZI, R. AND B. NCUBE. 2017. The biological activities of β -glucosidase, phosphatase and urease as soil quality indicators: a review. *Journal of soil science and plant nutrition*, 17(3), 794-807. <https://dx.doi.org/10.4067/S0718-95162017000300018>.
- ALVARES, C.A., STAPE, J.L., SENTELHAS, P.C., DE MORAES GONÇALVES, J.L. & G. SPAROVEK. 2013. Koppen's climate classification map for Brazil. *Meteorol. Z.* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- AMMANN, A. & H. BRANDL. 2011. Detection and differentiation of bacterial spores in a mineral matrix by Fourier transform infrared spectroscopy (FTIR) and chemometrical data treatment. *BMC Biophysics* 2011, 4:14. <http://www.biomedcentral.com/2046-1682/4/14>
- BUCKERIDGE, K., LA ROSA, A., MASON, K., WHITAKER, J., MCNAMARA, N., GRANT, H. & N. OSTLE. 2020. Sticky dead microbes: Rapid abiotic retention of microbial necromass in soil *Soil Biology and Biochemistry* (2020).
- CALDERÓN, F.J., MIKHA, M.M., VIGIL, M.F., NIELSEN, D.C., BENJAMIN, J.G. AND J.B. REEVES III, J.B. 2011. Diffuse-reflectance mid-infrared spectral properties of soils under alternative crop rotations in a semi-arid climate. *Commun. Soil Sci. Plant Anal.* 42, 2143–2159.
- CAO, Y., WEI, X., CAI, P., HUANG, Q., RONG, X. & W. LIANG. 2011. Preferential adsorption of extracellular polymeric substances from bacteria on clay minerals and iron oxide. *Colloids and Surfaces B: Biointerfaces* 83 (2011) 122–127
- CHEN, Q., MENG, Z., LIU, X., JIN, Q. & R. SU. 2018. Decision variants for the automatic determination of optimal feature subset in RF-RFE. *Genes* 9. <https://doi.org/10.3390/genes9060301>
- CHODAK, M. 2011. Near-infrared spectroscopy for rapid estimation of microbial properties in reclaimed mine soils. *Journal of Plant Nutrition and Soil Science* 174, 702-709.
- COHEN, M.J., PRENGER, J.P. & W.F. DEBUSK. 2005. Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. *Journal of Environmental Quality* 34, 1422e1434.
- COTRUFO, M.F., RANALLI, M.G., HADDIX, M.L., SIX, J. & E. LUGATO. 2019. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nature Geoscience*, 12(12), pp.989-994. <https://doi.org/10.1038/s41561-019-0484-6>
- COÛTEAUX, M-M., BERG, B. AND P. ROVIRA. 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. *Soil Biology & Biochemistry*. 2003; 35:1587–1600.

- COZZOLINO, D. & A. MORÓN. 2003. The potencial of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *Journal of Agricultural Science*. 2003; 140:65–71.
- DALMOLIN, R.S.D., GONÇALVES, C.N., KLAMT, E. & D.P. DICK. 2005. Relação entre os constituintes do solo e seu comportamento espectral. *Ciência Rural* 35, 481–489. <https://doi.org/10.1590/s0103-84782005000200042>.
- DAVINIC, M., FULTZ, L.M., ACOSTA-MARTINEZ, V., CALDERON, F.J., COX, S.B., DOWD, S.E., ALLEN, V.G., ZAK, J.C. & J. MOORE-KUCERA. 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. *Soil Biol. Biochem.* 46, 63–72. <https://doi.org/10.1016/j.soilbio.2011.11.012>
- DE BEECK, M., PERSSON, P. & A. TUNLID. 2021. Fungal extracellular polymeric substance matrices – Highly specialized microenvironments that allow fungi to control soil organic matter decomposition reactions. *Soil Biology and Biochemistry* 159 (2021) 108304. <https://doi.org/10.1016/j.soilbio.2021.108304>
- DEMATTÊ, J.A.M., BELLINASSO, H., ROMERO, D.J. AND C.T. FONGARO. 2014. Morphological interpretation of reflectance spectrum (MIRS) using libraries looking towards soil classification. *Sci. Agric.* 71, 509–520. <https://doi.org/10.1590/0103-9016-2013-0365>.
- 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., BELLINASSO, H., FRANCELINO, M.R., SCHAEFER, C.E.G.R., VICENTE, L.E., DOS SANTOS, U.J., DE SÁ BARRETTO SAMPAIO, E. V., MENEZES, R.S.C., DE SOUZA, J.J.L.L., ABRAHÃO, W.A.P., COELHO, R.M., GREGO, C.R., LANI, J.L., FERNANDES, A.R., GONÇALVES, D.A.M., SILVA, S.H.G., DE MENEZES, M.D., CURI, N., COUTO, E.G., DOS ANJOS, L.H.C., CEDDIA, M.B., PINHEIRO, É.F.M., GRUNWALD, S., VASQUES, G.M., MARQUES JÚNIOR, J., DA SILVA, A.J., BARRETO, M.C. D. V., NÓBREGA, G.N., DA SILVA, M.Z., DE SOUZA, S.F., VALLADARES, G.S., VIANA, J.H.M., DA SILVA TERRA, F., HORÁK-TERRA, I., FIORIO, P.R., DA SILVA, R.C., FRADE JÚNIOR, E.F., LIMA, R.H.C., ALBA, J.M.F., DE SOUZA JUNIOR, V.S., BREFIN, M.D.L.M.S., RUIVO, M.D.L.P., FERREIRA, T.O., BRAIT, M.A., CAETANO, N.R., BRINGHENTI, I., DE SOUSA MENDES, W., SAFANELLI, J.L., GUIMARÃES, C.C.B., POPPIEL, R.R., E SOUZA, A.B., QUESADA, C.A. & DO COUTO, H.T.Z. 2019. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma* 354, 113793. <https://doi.org/10.1016/j.geoderma.2019.05.043>
- DEMYAN, M., RASCHE, F., SCHULZ, E., BREULMANN, M., MULLER, T. & G. CADISCH. 2012. Use of specific peaks obtained by diffuse reflectance Fourier transform mid-infrared spectroscopy to study the composition of organic matter in a Haplic Chernozem. *European Journal of Soil Science*, April 2012, 63, 189 – 199. doi: 10.1111/j.1365-2389.2011.01420.x
- DEO, N., NATARAJAN, K. & P. SOMASUNDARAN. 2001. Mechanisms of adhesion of *Paenibacillus polymyxa* onto hematite, corundum and quartz. *Int. J. Miner. Process.* 62 2001 27–39
- DICK, R.P.; BREAKWELL, D.P. & R.F. TURCO. 1996. Soil enzymes activities and biodiversity measurements as integrative microbiological indicators. In: DORAN, J.W.; & JONES, A.J., eds. *Methods for assessing soil quality*. Soil Soc. Am., 1996, p. 247-271. Special publication.

- FIDÊNCIO, PH., POPPI, R.J., ANDRADE, J.C. & H. CANTARELLA. 2002. Determination of organic matter in soil using nearinfrared spectroscopy and partial least squares regression. *Communications in Soil Science and Plant Analysis*. 2002; 33:1607–1615.
- FOLLETT, R., KIMBLE, J. & R. LAL. 2001. The potential of US grazing lands to sequester soil carbon. Lewis Publishers, Boca Raton, Florida (2001), pp. 401-430
- GAO, X. & J. CHOROVER. 2009. In-situ monitoring of *Cryptosporidium parvum* oocyst surface adhesion using ATR-FTIR spectroscopy. *Colloids and Surfaces B: Biointerfaces* 71 (2009) 169–176. doi:10.1016/j.colsurfb.2009.02.003
- GÓMEZ, A., VAN LIER, Q., SILVERO, N., INFORSATO, L., DE MELO, M., RODRÍGUEZ-ALBARRACÍN, H., ROSIN, N., FIM ROSAS, J., RIZZO, R. & J.A.M DEMATTÊ. 2023. Digital mapping of the soil available water capacity: tool for the resilience of agricultural systems to climate change. *Science of the Total Environment* 882 (2023) 163572
- GOMES, M., MACHADO, B., RODRIGUES, E., MONTANHA, G., ROSSI, M., OTTO, R., LINHARES, F. AND H. W.P. CARVALHO. 2019. In Vivo Evaluation of Zn Foliar Uptake and Transport in Soybean Using X-ray Absorption and Fluorescence Spectroscopy. *J. Agric. Food Chem.* 2019, 67, 44, 12172–12181. <https://doi.org/10.1021/acs.jafc.9b04977>
- GRESCHUK, L., ARAÚJO, M., RODRÍGUEZ-ALBARRACÍN, H., BELLINASSO, H., SILVERO, N., PAIVA, A., POPPIEL, R., ROSIN, N., CAMPOS, L., DALMOLIN, R., BALLESTER, M. & J. DEMATTÊ. 2022. Combining spectral ranges for soil discrimination: A case study in the State of Maranhao – Brazil. *Geoderma Regional* 29 (2022) e00507. <https://doi.org/10.1016/j.geodrs.2022.e00507>
- GUYON, I., WESTON, J., BARNHILL, S. & V. VAPNIK. 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. <https://doi.org/10.1023/A:1012487302797>
- HEGLER, F., POSTH, N., JIANG, J. & A. KAPPLER. 2008. Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. *FEMS Microbiol Ecol* 66 (2008) 250–260. DOI:10.1111/j.1574-6941.2008.00592.x
- HOTA, S. 2021. *Synchrotron Based Techniques in Soil Analysis: A Modern Approach*. IntechOpen. doi: 10.5772/intechopen.99176
- IUSS WORKING GROUP WRB. 2015. *World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*. FAO, Rome, Italy (2015)
- JIANG, W., SAXENA, A., SONG, B., WARD, B.B., BEVERIDGE, T.J. AND S.C. MYNENI. 2004. Elucidation of functional groups on gram-positive and gram-negative bacterial surfaces using infrared spectroscopy. *Langmuir* 20, 11433–11442.
- JINDALUANG, W., KHEORUENROMNE, I., SUDDHIPRAKARN, A., SINGH, B. & B. SINGH. 2013. Influence of soil texture and mineralogy on organic matter content and composition in physically separated fractions soils of Thailand. *Geoderma* 195–196 (2013) 207–219. <http://dx.doi.org/10.1016/j.geoderma.2012.12.003>

- JUNIOR, G., NUNES, L., GOMES, M., DE ALMEIDA, E. AND H.W. DE CARVALHO. 2019. Direct determination of mineral nutrients in soybean leaves under vivo conditions by portable X-ray fluorescence spectroscopy. *X-Ray Spectrometry*. 2020; 49: 274–283. <https://doi.org/10.1002/xrs.3111>
- KAISER, K. AND G. GUGGENBERGER. 2000. The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. *Organic Geochemistry* 31:711-725.
- KIRSTEN, M., MIKKUTA, R., VOGEL, C., THOMPSON, A., MUELLER, C., KIMARO, D., BERGSMAN, H., FEGER, K. & K. KALBITZ. 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. *Scientific Reports* | (2021) 11:5076
- KLEBER, M., EUSTERHUES, K., KEILUWEIT, M., MIKUTTA, C., MIKUTTA, R. & P.S. NICO. 2015. Mineral-organic associations: formation, properties, and relevance in soil environments. *Advances in Agronomy* 1–140.
- KLEBER, M., BOURG, I., COWARD, E., HANSEL, C., MYNENIS, S. AND N. NUNAN. 2021. Dynamic interactions at the mineral–organic matter interface. *Nat Rev Earth Environ* 2, 402–421 (2021). <https://doi.org/10.1038/s43017-021-00162-y>
- KUHN, M. 2021. caret: Classification and Regression Training. R package version 6.0-88. <https://CRAN.R-project.org/package=caret>
- LADD, J., FOSTER, R.C. & J.O. SKJEMSTAD. 1993. Soil structure: carbon and nitrogen metabolism. *Geoderma*, 56 (1993) 401-434. Elsevier Science Publishers B.V., Amsterdam. <https://doi.org/10.1016/B978-0-444-81490-6.50034-0>
- LEHMANN, J., KINYANGI, J. AND D. SOLOMON. 2007. Organic matter stabilization in soil microaggregates: implications from spatial heterogeneity of organic carbon contents and carbon forms. *Biogeochemistry* 85,45–57.
- LEHMANN, J. AND D. SALOMON. 2010. Chapter 10 - Organic Carbon Chemistry in Soils Observed by Synchrotron-Based Spectroscopy. Editor(s): Balwant Singh, Markus Gräfe, *Developments in Soil Science*, Elsevier, Volume 34, 2010, Pages 289-312, ISSN 0166-2481, ISBN 9780444532619, [https://doi.org/10.1016/S0166-2481\(10\)34010-4](https://doi.org/10.1016/S0166-2481(10)34010-4).
- LIANG, C., AMELUNG, W., LEHMANN, J. & M. KASTNER. 2019. Quantitative assessment of microbial necromass contribution to soil organic matter. *Glob Change Biol*. 25, 3578–3590.
- LIU, Y., LI, O., LI, Y., BAO, J., HU, Z., HAO, D., SONG, D., WANG, Y. & M. YANG. 2015. Detection of Nanoscale Soil Organic Matter by Middle Infrared Spectrum for Forensic Science. *Hindawi Publishing Corporation Journal of Chemistry* Volume 2015, Article ID 189421. <http://dx.doi.org/10.1155/2015/189421>
- LUDWIG, B., NITSCHKE, R., TERHOEVEN-URSELMANS, T., MICHEL, K. & H. FLESSA. 2008. Use of mid-infrared spectroscopy in the diffuse reflectance mode for the prediction of the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science* 171, 384e391.
- MACEDO, F., MONTANHA, G., DE CARVALHO, H.W. AND W. DE MELO. 2021. Nickel Influences Urease Activity and Calcium Distribution in Tomato Fruits. *ACS Agric. Sci. Technol*. 2021, 1, 1, 29–34. Publication Date: January 27, 2021 <https://doi.org/10.1021/acsagscitech.0c00003>

- MADEJOVÁ, J., GATES, W.P., AND S. PETIT. 2017. IR Spectra of Clay Minerals, in: *Developments in Clay Science*. Elsevier B.V., pp.107–149. <https://doi.org/10.1016/B978-0-08100355-8.00005-9>.
- MARTINEZ, C. AND M. TABATABAI. 1997. Decomposition of biotechnology by-products in soils. *Journal of Environmental Quality*. 26, 625–632.
- MCWHIRTER, M.J., MCQUILLAN, A.J. & P.J. BREMER. 2002. Influence of ionic strength and pH on the first 60 min of *Pseudomonas aeruginosa* attachment to ZnSe and to TiO₂ monitored by ATR-IR spectroscopy. *Colloids Surf. B Biointerfaces* 26, 365–372.
- MERINO, C., GODOY, R. & F. MATUS. 2016. Soil enzymes and biological activity at different levels of organic matter stability. *Journal of Soil Science and Plant Nutrition*. 16, 14–30.
- MÜLLER, B., 2015. Experimental interactions between clay minerals and bacteria: a review. *Pedosphere* 25, 799–810.
- NATH, D., LAIK, R., MEENA, V., PRAMANICK, B. & SINGH. 2021. Can mid-infrared (mid-IR) spectroscopy evaluate soil conditions by predicting soil biological properties?. *Soil Security* 4 (2021) 100008. <https://doi.org/10.1016/j.soisec.2021.100008>
- NGUYEN, T.T., JANIK, L.J. AND M. RAUPACH. 1991. Diffuse reflectance infrared Fourier transform (Drift) spectroscopy in soil studies. *Aust. J. Soil Res.* 29, 49–67. <https://doi.org/10.1071/SR9910049>.
- NICOLÁS, C., ALMEIDA, J., ELLSTRÖM, M., BAHR, A., BONE, S., ROSENSTOCK, N., BARGAR, J., TUNLID, A., PERSSON, P. & H. WALLANDER. 2017. Chemical changes in organic matter after fungal colonization in a nitrogen fertilized and unfertilized Norway spruce forest. *Plant Soil* (2017) 419:113–126. DOI 10.1007/s11104-017-3324-8
- OJEDA, J.J., ROMERO-GONZALEZ, M.E., POURAN, H.M. & S.A. BANWART. 2008. In situ monitoring of the biofilm formation of *Pseudomonas putida* on hematite using flow-cell ATR-FTIR spectroscopy to investigate the formation of inner-sphere bonds between the bacteria and the mineral. *Mineral. Mag.* 72, 101–106.
- OMOIKE, A. & J. CHOROVER. 2006. Adsorption to goethite of extracellular polymeric substances from *Bacillus subtilis*. *Geochim. Cosmochim. Acta* 70, 827–838.
- PARIKH, S., GOYNE, K., MARGENOT, A., MUKOME, F. & F. CALDE-RON. 2014. Soil Chemical Insights Provided through Vibrational Spectroscopy. Publications from USDA-ARS / UNL Faculty. Paper 1471.
- RASCHE, F., MARHAN, S., BERNER, D., KEIL, D., KANDELER, E. & G. CADISCH. 2013. midDRIFTS-based partial least square regression analysis allows predicting microbial biomass, enzyme activities and 16S rRNA gene abundance in soils of temperate grasslands. *Soil Biology & Biochemistry* 57 (2013) 504–512
- RASCHE, F., KNAPP, D., KAISER, C., KORANDA, M., KITZLER, B., ZECHMEISTER-BOLTENSTERN, S., RICHTER, A. & A. SESSITSCH. 2011. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *The ISME Journal* 5, 389–402

- RINNAN, R. & A. RINNAN. 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. *Soil Biol. Biochem.* 39 (7), 1664–1673. <https://doi.org/10.1016/j.soilbio.2007.01.022>.
- RODRÍGUEZ-ALBARRACÍN, H., DEMATTÊ, J., ROSIN, N., DARGHAN, E., SILVERO, N., CERRI, C., MENDES, W. & M. TAYEBI. 2023. Potential of soil minerals to sequester soil organic carbon. *Geoderma* 436 (2023) 116549. <https://doi.org/10.1016/j.geoderma.2023.116549>
- RONG, X., CHEN, W., HUANG, Q., CAI, P. & W. LIANG. 2010. Pseudomonas putida adhesion to goethite: studied by equilibrium adsorption, SEM, FTIR and ITC. *Colloids Surf. B Biointerfaces* 80, 79–85.
- SHARMA, V., CHAUHAN, R. & R. KUMAR. 2021. Spectral characteristics of organic soil matter: A comprehensive review. *Microchemical Journal* 171 (2021) 106836. <https://doi.org/10.1016/j.microc.2021.106836>
- SHEN, Z. AND R.A. VISCARRA ROSSEL. 2021. Automated spectroscopic modelling with optimised convolutional neural networks, *Sci. Rep.*, 11, 208, <https://doi.org/10.1038/s41598-020-80486-9>, 2021.
- SILVERO, N.E.Q., DI RAIMO, L.A.D.L., PEREIRA, G.S., DE MAGALHAES, L.P., DA TERRA, F.S., DASSAN, M.A.A., SALAZAR, D.F.U. AND J.A.M. DEMATTE. 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>.
- SIX, J., BOSSUYT, H., DEGRYZE, S. AND K. DENEFF. 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil Tillage Res.* 79, 7–31.
- SOUZA, A.B., DEMATTE, J.A.M., MELLO, F.A.O., SALAZAR, D.F.U., MENDES, W.S. & J.L. SAFANELLI. 2020. Ratio of clay spectroscopic indices and its approach on soil morphometry. *Geoderma* 357, 113963.
- STENBERG, B., VISCARRA ROSSEL, R.A., MOUAZEN, A.M. & J. WETTERLIND. 2010. Chapter five— visible and near infrared spectroscopy in soil science. In: Donald, L.S. (Ed.), *Advances in Agronomy*, vol. 107. Academic Press, pp. 163–215.
- TABATABAI, M.A. 1994. Soil enzymes. In: WEAVER, R.W.; SCOTT, A. & BOTTOMELEY, P.J., eds. *Methods of soil analysis: microbiological and biochemical properties*. Madison, Soil Sci. Soc. Am., 2:778-835, 1994. (Special Publication, 5)
- TERRA, F.S., DEMATTE, J.A.M. AND R.A. VISCARRA ROSSEL. 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>.
- TORN, M., KLEBER, M., ZAVALETA, E., ZHU, B., FIELD, C. & S.E. TRUMBORE. 2013. A dual isotope approach to isolate soil carbon pools of different turnover times. *Biogeosciences* 10, 8067–8081.
- VANCE, E.D., BROOKES, P.C., JENKINSON, D.S., 1987. An extraction method for measuring soil microbial biomass C. *Soil Biol. Biochem.* 19 (6), 703–707. [https://doi.org/10.1016/0038-0717\(87\)90052-6](https://doi.org/10.1016/0038-0717(87)90052-6).

- VISCARRA-ROSSEL, R. & T. BEHRENS. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158 (2010) 46–54. doi:10.1016/j.geoderma.2009.12.025
- VISCARRA ROSSEL, R.A., WALVOORT, D.J.J., MCBRATNEY, A.B., JANIK, L.J., AND J.O. SKJEMSTAD. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>.
- VISCARRA ROSSEL, R.A. & W.S. HICKS. 2015. Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. *European Journal of Soil Science* 66, 438–450. <https://doi.org/10.1111/ejss.12237>.
- VISCARRA ROSSEL, R. A., BEHRENS, T., BEN-DOR, E., BROWN, D., DEMATTÉ, J., SHEPHERD, K., SHI, Z., STENBERG, B., STEVENS, A., ADAMCHUK, V., AICHI, H., BARTHES, B., BARTHOLOMEUS, H., BAYER, A., BERNOUX, M., BOTTCHEK, K., BRODSKY, L., DU, C., CHAPPELL, A., FOUAD, Y., GENOT, V., GOMEZ, C., GRUNWALD, S., GUBLER, A., GUERRERO, C., HEDLEY, C., KNADEL, M., MORRAS, H., NOCITA, M., RAMÍREZ LÓPEZ, L., ROUDIER, P., CAMPOS, E., SANBORN, P., SELBITTO, V., SUDDUTH, K., RAWLINS, B., WALTER, C., WINOWIECKI, L., HONG, S. AND 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>.
- VISCARRA-ROSSEL, R.A., YANG, Y., BISSETT, A., BEHRENS, T., DIXON, K., NEVIL, P. & S. LI. 2022. Environmental controls of soil fungal abundance and diversity in Australia's diverse ecosystems. *Soil Biology and Biochemistry* 170 (2022) 108694. <https://doi.org/10.1016/j.soilbio.2022.108694>
- TEIXEIRA, P.C., et al. *Manual de métodos de análise de solo EMBRAPA*, Brasília (2017), p. 212
- 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, 10.1097/00010694-193401000-00003
- WAN, J., TYLISZCZAK, T. AND T.K. TOKUNAGA. 2007. Organic carbon distribution, speciation, and elemental correlation within soil microaggregates: application of STXM and NEXAFS spectroscopy. *Geochim. Cosmochim. Acta* 71, 5439–5449.
- WANG, S.L. AND C.T. JOHNSTON. 2000. Assignment of the structural OH stretching bands of gibbsite. *Am. Mineral.* 85, 739–744.
- WANG, Y. & Y. LI. 2023. Mapping the ratoon rice suitability region in China using random forest and recursive feature elimination modeling. *Field Crops Research* 301 (2023) 109016.
- WEBER, K. A., ACHENBACH, L. A. & J. COATES. 2006. Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat. Rev. Microbiol.* 4, 752–764. doi: 10.1038/nrmicro1490
- WITZGALL, K., VIDAL, A., SCHUBERT, D., HÖSCHEN, C., SCHWEIZER, S., BUEGGER, F., POUTEAU, V., CHENU, C. & C. MUELLER. 2021. Particulate organic matter as a functional soil component for persistent soil organic carbon. *NATURE COMMUNICATIONS* | (2021) 12:4115 | <https://doi.org/10.1038/s41467-021-24192-8>

- YANG, L., LI, X., SHIB, J., SHEN, F., QI, F., GAO, B., CHEN, Z., ZHU, A. & C. ZHOU. 2020. Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. *Geoderma* 369 (2020) 114337
- YANG, J., ZHANG, X., BOURG, A. AND H. STONE. 2021. 4D imaging reveals mechanisms of clay-carbon protection and release. *Nat Commun* 12, 622 (2021). <https://doi.org/10.1038/s41467-020-20798-6>
- YANG, Y., SHEN, Z., BISSETT, A. & R. VISCARRA-ROSSEL. 2022. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. *SOIL*, 8, 223–235, 2022. <https://doi.org/10.5194/soil-8-223-2022>
- ZHANG, L., CHEN, X., XU, Y., JIN, M., YE, X., GAO, H., CHU, W., MAO, J. & M. THOMPSON. 2020. Soil labile organic carbon fractions and soil enzyme activities after 10 years of continuous fertilization and wheat residue incorporation. *Scientific Reports* (2020) 10:11318. <https://doi.org/10.1038/s41598-020-68163-3>.
- ZHANG, L. & L. WANG. 2023. Optimization of site investigation program for reliability assessment of undrained slope using Spearman rank correlation coefficient. *Computers and Geotechnics* 155 (2023) 105208
- ZORNOZA, R., GUERRERO, C., MATAIX-SOLERA, J., SCOW, K., ARCENEGUI, V. & J. MATAIX-BENEYTO. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem.* 2008 July; 40(7): 1923–1930. doi:10.1016/j.soilbio.2008.04.003.

5. SPATIALIZATION OF SOIL BETA-GLUCOSIDASE, PHOSPHATASE AND UREASE ENZYME ACTIVITY INDICES FOR BRAZILIAN AGRICULTURAL AREAS: A PARADIGM SHIFT

Abstract

Maintaining or improving soil health is fundamental to achieving the Sustainable Development Goals (SDGs). Therefore, it needs to be evaluated through indicators that inform politicians, farmers and citizens about the impact of human activities on soil and its functions. This evaluation requires a comprehensive consideration of soil chemical, physical and biological (microbial) indicators. Biological indicators include the enzymatic activity of microorganisms, such as beta-glucosidase, phosphatase and urease activity, which are related to the carbon, phosphorus and nitrogen cycles. Such enzymatic activity of microorganisms is closely related to the types and amounts of MOS, and Vis NIR -SWIR and Mid IR wavelengths respond to functional groups of different types of soil organic compounds, iron oxides and clay minerals, which are related to this activity. The use of enzyme activity as an indicator of soil health requires a cost-effective estimation method, such as reflectance spectroscopy analysis that can be used to predict microbiological activity and thus achieve mapping, classification and monitoring methodologies, mainly through machine learning techniques. The objective of this work was to develop a technique to calculate and spatialize the activity indices of beta-glucosidase, phosphatase and urease enzymes for the Brazilian agricultural territory. These indices were constructed from the selection of specific bands of the Vis NIR-SWIR and Mid IR spectral range for each enzyme, obtained with the RFE algorithm, from which multiple linear regression equations were created and applied to 13657 (Vis NIR SWIR) and 2651 (Mid IR) soil samples (0-20 cm depth) included in the Brazilian Soil Spectral Library (BSSL). These indices were mapped under the DSM framework, having as covariates the Synthetic Soil Image (SYSI), variables associated with relief, climate, biomes and mineralogical maps. The first beta-glucosidase, phosphatase and urease activity maps were obtained for the agricultural area of Brazil (3481362.60 km²), with a validation R² ranging from 0.68 to 0.35 and a validation with the oxide maps based on a Spearman correlation of 0.74 to 0.47. These enzyme activity indices at the 30 m scale can be considered an important contribution to the monitoring and mapping of the quality and health of Brazilian soils, as they are sensitive to land use and management.

Keywords: Soil spectroscopy, Enzyme activity, Digital soil mapping, Soil health

5.1. Introduction

Soil is a fundamental natural resource that supports 95 % of food production for humans (Sarkheil et al., 2020), and represents the difference between survival and extinction for most terrestrial life (Doran and Zeiss, 2000), therefore, maintaining or improving the health of our soils is important for achieving the Sustainable Development Goals (SDGs), ensuring the provision of ecosystem services, food security and sustaining life on Earth (Sarkheil et al., 2023).

Soil health is understood as the capacity of this system to function and provide ecosystem services that determine agricultural and forest sustainability, environmental quality and,

consequently, plant, animal and human health (Wei et al. 2022, Costantini and Priori, 2023). Therefore, it is important and mandatory to assess soil health through the use of indicators that can inform policy makers, farmers and citizens about the impact of human activities on soil and its functions (Costantini and Priori, 2023). This assessment requires comprehensive consideration of soil chemical, physical and biological (microbial) indicators (Lehmann et al., 2020). These indicators are chosen according to soil functions (Andrews et al., 2002a), corresponding to (i) food and biomass production, (ii) storage, filtration and transformation of compounds, (iii) habitats for organisms and gene pools, (iv) physical and cultural environment, (v) source of primary materials, (vi) carbon reservoir and (vii) archive of geological and archaeological heritage (European Commission (COM 2006.231). These functions are related to ecosystem services (ES) associated with provisioning (food, fiber and timber production), regulation (climate, flood and water regulation), and cultural and supporting services (nutrient cycling, soil formation) (Silvero et al., 2023).

Biological indicators include enzyme activity, which is closely related to the physical and chemical properties of the soil, and allows predicting global microbial activity (Ma et al., 2021). In addition, such activity influences carbon (C) depletion and sequestration (Zhang et al., 2020), and is considered as one of the indicators with greater sensitivity to soil management practices than other variables (Adetunji et al., 2017).

Enzyme activity plays a vital role in agriculture and nutrient cycling (Balota and Chaves, 2010), especially hydrolases that are associated with carbon (beta-glucosidase), nitrogen (urease) and phosphorus (phosphatase) cycles (Karaca et al., 2010), and are widely used as indicators of soil quality (Bandick and Dick, 1999). Beta-glucosidase catalyzes the hydrolysis and biodegradation of glycosides (Martinez and Tabatabai, 1997), obtaining glucose as the final product, which is an important source of energy C for the activity of microorganisms (Merino et al., 2016). In general, the activity of this enzyme is related to the presence of simple sugars, frequently used by microorganisms and can provide a clear signal of soil organic carbon (SOC) alterations (Adetunji et al., 2017).

Phosphatase catalyzes the hydrolysis of phosphoric acid esters and anhydrides (Condrón et al., 2005). The response of this activity is related to microbial abundance, amount of soil organic matter (SOM) and soil management (Banerjee et al., 2012). Additionally, the demand for phosphorus by plants and microorganisms is related to the enzymatic production of phosphatase in the soil, so it is used as an indicator of the availability of inorganic phosphorus (Piotrowska-Dlugosz and Charzynski, 2015). This enzyme is also related to nitrogen fixation (Makoi and Ndakidemi, 2008) maintaining a strong correlation with soil N availability.

Urease originates from bacteria, yeasts, fungi, algae, animal wastes and plants (Follmer, 2008). This enzyme catalyzes the hydrolysis of urea (Das and Varma, 2010), so its expression is related to the regulation of N supply (Piotrowska-Dlugosz and Charzynski, 2015). This enzyme activity increases with organic fertilization and decreases with soil tillage (Adetunji et al., 2017).

The use of the enzymatic activity of these three enzymes as biological indicators of soil quality requires a simple and cost-effective prediction method (Rasche et al., 2013), such as reflectance spectroscopy analysis in the visible, near-infrared and shortwave infrared (Vis-NIR-SWIR) and mid-infrared (Mid IR) ranges, which can be used to predict the enzymatic activity of microorganisms due to their close association with the types and amounts of SOM, which respond to functional groups of different types of soil organic compounds, iron oxides and clay minerals that are related to such activity and are identified in different bands along the electromagnetic spectrum (Ojeda et al. 2008, Viscarra Rossel and Hicks 2015, Viscarra Rossel et al. 2022, Yang et al. 2021).

In general, there are no specific vis-NIR-SWIR and Mid IR absorptions assigned to soil enzyme activity (Rasche et al., 2011), due to their low contents compared to other edaphic compounds, which makes it difficult to induce activity-specific changes in the spectra. However, their prediction is possible due to the strong relationship with soil organic materials (Cohen et al., 2005). Reflectance spectroscopy analyses on soil samples contribute to the prediction of compounds used by microbes and provide integrated measures of soil mineral-organic composition (Viscarra Rossel et al., 2016). Rasche et al. (2013) points out some bands of the Mid IR spectrum that favor the prediction of betaglucosidase, xylosidase and urease. In relation to Vis-NIR spectra, some predictions associated with enzyme activity in soil have been reported by Cohen et al. (2005), Zornoza et al. (2008) and Chodak (2011), however, these studies indicate wide ranges of these electromagnetic spectra.

NIR and Mid IR reflectance spectra respond to the concentration of compounds related to C-H, N-H, S-H, C=O and O-H chemical bonds (Zornoza et al. 2008, Viscarra Rossel et al. 2022) and such compounds are closely related to enzymatic activity as they are integrated into fungal and bacterial products (Parikh et al., 2014).

Spectral order and specific band identification is important in the prediction of soil properties, because characteristic peaks occur at different positions in the spectrum and may represent differences in information (Zhang et al., 2020). This spectral response information requires mathematical extraction and subsequent correlation with soil properties (Reda et al., 2020), which is achieved by using machine learning techniques and multivariate regression

analysis to extract the wave coverages related to the specific property to be predicted (Zornoza et al., 2008).

Despite the importance of the enzymatic activity of microorganisms as indicators of soil health, few efforts have been made to map these indicators, making it a challenge. However, digital soil mapping (DSM) techniques are available, which promote the construction of maps with high spatial resolution and low uncertainty, even at low sampling densities (Mendes et al., 2022). These DSM techniques combine soil point data with statistically correlated auxiliary data (covariates) (McBratney et al., 2003) and together with reflectance spectroscopy analyses associated with soil microbiological and mineral enzymatic activity, it is possible to generalize methodologies for mapping, classification and monitoring (Di Iorio et al., 2019), of these soil properties, mainly by means of machine learning techniques.

Mathematical models that relate to DSM approaches have the ability to predict soil properties based on environmental covariates, through algorithms associated with these machine learning techniques (Hengl et al., 2015). Among these algorithms, the most commonly used is Random Forest (RF) (Zeraatpisheh et al. 2020, Padarian et al., 2020), because it is a robust model that is composed of multiple decision trees that are not correlated with each other, which gives it a high accuracy in predictions with low possibility of overfitting (Wadoux et al., 2020).

The quality of DSM products is conditional on the environmental covariates used for modeling the attribute of interest. Usually environmental covariates representing physical and chemical processes associated with soil spatial variation (McBratney et al. 2003, Wadoux et al. 2020), and/or representing soil formation factors (Viscarra Rossel, 2011) are used, so digital terrain elevation models, vegetation indices, climatic covariates, geological maps, satellite images of surface reflectance are commonly used (Ma et al., 2019). Recently, bare soil reflectance images have been included as environmental covariates (Rosin et al. 2023, Safanelli et al 2021b). These images are obtained using the GEOS3 (Geospatial Soil Sensing System) technique developed by Dematte et al. (2018), which allows capturing bare soil reflectance from historical series of Landsat images by adding pixels that were exposed at least once throughout the time series to a synthetic soil image (YSI). Silvero et al. (2021), Rizzo et al. (2020), Rosin et al. (2023) have demonstrated the importance of this covariate in soil mapping due to its strong correlation with edaphic attributes.

The objective of this work was to develop a technique based on Vis NIR-SWIR and Mid IR reflectance spectroscopy products, DSM approach and machine learning techniques, to create and spatialize the activity indices of betaglucosidase, phosphatase and urease enzymes

for the Brazilian agricultural territory, having as covariates the Synthetic Soil Image (SYSI), variables associated with relief, climate, biomes and mineralogical maps. It is expected that the maps obtained will cover all agricultural areas of Brazil, with a spatial resolution of 30 m, which will allow researchers, farmers and consultants to understand the dynamics of microbiological activity in the country and contribute to the management of these agricultural areas.

5.2. Methodology

5.2.1. Study area

The study area covers the entire agricultural territory of Brazil. Brazil has an area of approximately 8.5 million km² (Figure 1) and the agricultural area corresponds to 3,481,362.60 km². The country has a high variability of soils, which is due to soil-forming factors such as climate, parent material and topography. The climate of Brazil varies from tropical (81.4%), dry (4.9%) and subtropical (13.7%) according to the Köppen classification (Alvares et al., 2013). The parent materials were formed mainly in the Cenozoic, Mesozoic, Neoproterozoic and Paleoproterozoic eras. These materials gave origin to a geology that is composed of several types of metamorphic, igneous and sedimentary rocks (Gómez et al, 2019). The region comprising the states of Paraná, São Paulo, Rio Grande do Sul, Minas de Gerais, Goiás and Mato Grosso do Sul, presents a large mass of basalt formed by basaltic eruptions, giving rise to extrusive rocks with high levels of iron. Towards the coastal zone of the country, intrusive plutonic rocks formed from continental drift, which present highly crystalline minerals such as quartz.

Brazil also has a variable topography, composed mainly of slightly sloping terrain, with an altitudinal variation of 200 to 400m, where there is low tectonic and volcanic activity (Ross, 2013).

5.2.2. Soil observations

A legacy database was used with spectral data in the Vis NIR-SWIR range (350-2500nm), with spectral resolution of 1 nm from the Brazilian Soil Spectral Library (BSSL) (Dematte et al., 2019), which has 69740 observations, of which 13657 associated to the depth 0 to 20 cm were taken. In addition, 2651 observations of the Mid IR range (4000 - 600 cm⁻¹) with a spectral resolution of 1 cm⁻¹ were used.

5.2.3. Enzyme activity indexes

For the construction of the enzymatic activity indexes of beta-glucosidase, urease and phosphatase, 35 samples located in Piracicaba, Sao Paulo, 34 in the municipality of Tomé-Açu, state of Pará, 15 in the municipality of Bandeirantes and 19 in Dourados, Mato Grosso do Sul, and 15 in Jardim Olinda, Paraná were used. The enzymatic activity of beta-glucosidase, phosphatase and urease were determined following the methodology of Tabatabai (1994) and Dick et al. (1996).

For these samples, spectral readings of the Vis-NIR-SWIR and Mid IR ranges were performed, making use of two sensors, the FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, Col., USA) for obtaining reflectance data from 350 to 2500 nm (Vis-NIR-SWIR) (Demattê et al., 2019) and the Alpha Sample Compartment RT-DLaTGS ZnSe sensor (Bruker Optik GmbH), equipped with a drift attachment (Souza et al., 2020) for obtaining spectral information from the range 4000 to 600 cm^{-1} , corresponding to the mid-infrared, with a resolution of 1.2 cm^{-1} and 64 scans per second according to the description of Terra et al. (2015). For readings with this sensor, the samples ground and sieved at 100 mesh. The spectral resolution of FieldSpec 3, varies from 3 nm for the range between 400 and 700 nm and 10nm for the range from 700 to 2500 nm, so a resampling at 1 nm was performed, obtaining 2151 spectral bands (Greschuk et al., 2022). For the readings with this sensor, the soil samples were ground and sieved at 2 mm.

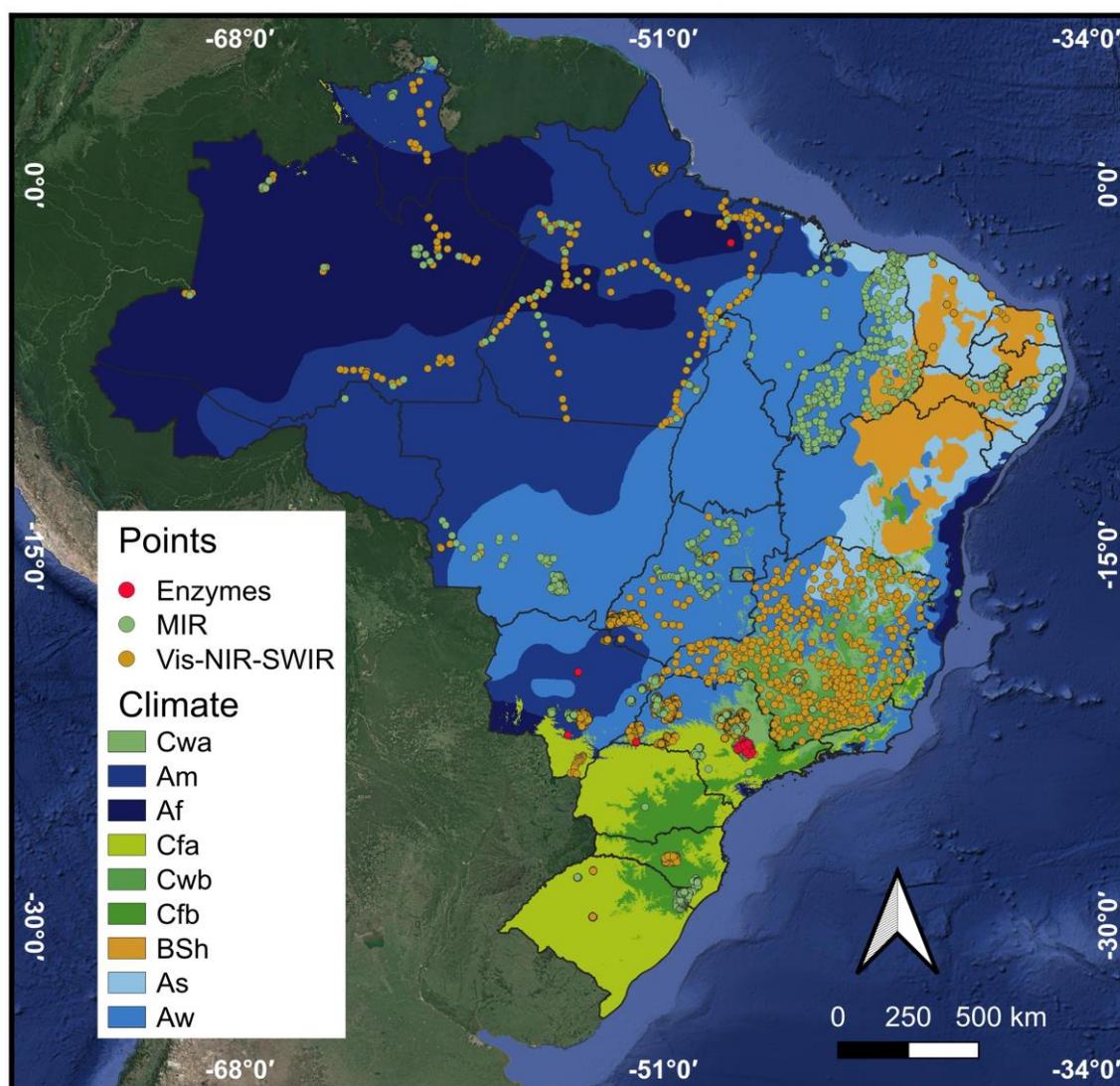


Figure 1. Study area and soil observation points for 0-20 cm depth from the Brazilian Soil Spectral Library (BSSL) dataset, points with data from the Vis NIR SWIR spectrum (orange) and Mid IR spectrum (MIR, green). In red are the samples located in Piracicaba Sao Paulo, Tomé-Açu municipality, Pará state, Bandeirantes municipality and Dourados de Mato Grosso do Sul, which present enzymatic activity analysis of beta-glucosidase, urease and phosphatase. Köppen climate classification map: A = Tropical zone, without dry season (Af), monsoon (Am), dry winter (Aw) or dry summer (As); B = Dry zone with semi-arid climate and low latitude and altitude (BSh); C = Dry zone with semi-arid climate and low latitude and altitude (BSh). Subtropical zone with oceanic climate, without dry season, with hot summer (Cfa) or with temperate summer (Cfb) or subtropical zone with dry winter and hot summer (Cwa) or temperate summer (Cwb), taken from Alvares et al. (2013).

The Recursive Feature Elimination (RFE) algorithm developed by Guyon et al. (2002) and implemented in the "Caret" package of R, using the Random Forest (RF) machine learning method as an internal model (Kuhn, 2021), was used to select the set of bands of

greatest importance for each enzyme activity. This algorithm is a method of wide applicability that allows handling nonlinear and more complex relationships, considering comprehensively the relationships between the analyzed factors (Wang and Li, 2023). RFE allows determining the optimal number of predictors by taking the model with the best performance (RMSE) (Zhang et al., 2023).

For the present work, the RFE algorithm was applied for each enzyme activity and for each range of the electromagnetic spectrum, thus obtaining the most important Vis NIR SWIR and Mid IR bands of each enzyme (Figure 2), from which regression models were constructed to quantify the variables through spectral information.

Multivariate linear regression models (MRA) were tested considering as dependent variables the enzymatic activity of beta-glucosidase, phosphatase and urease, and as explanatory variables the reflectance values of the most important bands for each variable and each Vis NIR-SWIR and Mid IR electromagnetic spectrum. Therefore, regression models were performed for each range of the spectrum and each enzyme.

The choice of the best fit models was based on the correlation value (r) between the observed values of the response of the dependent variables Y and the values estimated by the model (Hoge et al., 2018). In addition, we considered the fulfillment of the assumption of independence of the residuals, using the Moran index test (MIT) with the matrix of weights of all neighbors (Liu and Chen, 2021), where independence is fulfilled if the p-value of the test is greater than 0.05. From these models, enzyme activity was calculated for each observation in the legacy data sets.

5.2.4. Environmental covariates

Soil microbiological enzyme activity is directly related to SOC dynamics (Yang et al., 2021), which is closely related to the soil development process and thus to its forming factors suggested by Jenny (Hobley et al., 2015), among which there are active factors, such as climate and organisms, and passive ones, such as parent materials, relief and time. Based on these forming factors, the covariates used in soil attribute mapping are defined. For the present study, SYSI, soil attributes, climate, mineralogy and biomes were taken into account.

SYSI represents soil directly and parent material and time indirectly (Poppiel et al., 2020). This covariate is obtained using the GEOS3 method proposed by Dematte et al. (2018) and Dematte et al. (2020), applied in Google Earth Engine (GEE) to a collection of Landsat images (5, 6, 7 and 8) from 1984 to 2020. SySI has 6 bands (blue (450-520 nm), green (520-600 nm), red (630-690 nm), NIR (760-900 nm), SWIR1 (1550-1750 nm) and SWIR2 (2080-

2350 nm), processed by the GEOS3 method corresponding to a data mining algorithm that extracts ground features from historical data, related to soil spectral trend, Normalized Difference Vegetation Index (NDVI) and Normalized Burning Index 2 (NBR2), which allows creating a soil mask for the Landsat image collection, by selecting bare soil pixels. Therefore, the SySI pixels correspond to the median of all the bare soil pixels detected in the time series images (Dematte et al., 2020).

In relation to terrain attributes, elevation, slope, aspect, curvatures, valleys, hills, orientation and topographic moisture index were taken into account as covariates, due to their influence on the action of climatic factors on soils (Weil and Brady, 2016, Lamichhane et al. 2019). In addition, these covariates control soil water status, litter mineralization dynamics, erosion, deposition processes (Hengl et al., 2015) and the rate of decomposition of organic materials, due to the direct relationship of elevation with temperature (Schindlbacher et al., 2010). These terrain attributes were determined from the 30-m resolution digital elevation model (DEM) obtained from the Advanced Land Observing Satellite (ALOS) (Japan Aerospace Exploration Agency, 2021) available from GEE (Golerick et al., 2017), and by making use of the Terrain Analysis in GEE (TAGEE) package (Safanelli et al., 2020).

As climatic covariates, temperature and mean annual precipitation were considered, because they are the covariates that most affect C storage (Shi et al., 2020), control its dynamics and by in enzymatic activity (Hengl et al. 2015, Minasny et al. 2013). Because they determine the rate of C decomposition by microorganisms (Lal, 2004). These covariates were taken from the 1970 to 2000 WorldClim2 dataset (Fick and Hijmans, 2014).

Relative abundance maps of hematite, goethite, kaolinite and gibbsite from Rosin et al. (2023) were included as mineralogical covariates. Because functional groups that are related to enzymatic activity present a strong interaction with the surfaces of Fe and Al oxides and clay minerals such as 1:1 clays (Cao et al. 2011, Parikh et al. 2014, Kleber et al. 2021, Yang et al. 2021, Viscarra Rossel et al 2022).

Considering that vegetation also interferes with SOC stocks because it favors the addition of plant biomass, its ease of decomposition (Bui et al., 2009), associated with the type of species (Mueller et al. 2015, Vesterdal et al. 2013). These biomass inputs are relevant in the most superficial horizons (Rasse et al. 2005, Lorenz et al. 2017), therefore, a stratification by biome was performed considering the map of global ecoregions reported by Dinerstein et al. (2017).

5.2.5. Spatial prediction

Spatial prediction was determined from the Random Forest (RF) algorithm, which corresponds to the algorithm with the best performance reported for the prediction and mapping of various soil attributes such as SOC (Gómez et al. 2023, Zeraatpisheh et al. 2020, Padarian et al. 2020). This algorithm is a non-parametric model, based on the combination of different independent decision trees that are created in the training stage, where each tree is generated from a random sampling of identically distributed covariate values, allowing to obtain an average value of the prediction of all trees (Breiman 2001, Gómez et al. 2023). The strength of this algorithm is based on the initial randomization of the data and the selection of random entries (Sothe et al., 2022) with replacement of the original data and internal validation with data not used in the bootstrap procedure (Khaledian and Miller 2020, Zeraatpisheh et al. 2020). This allows capturing complex relationships between variables leading to robust and accurate results (Cutler et al., 2012).

Due to the great number of data represented by the mapping at the national level, the calculated values for each observation were uploaded to Google Cloud Storage through the Google Earth Engine Python API, which has a great set of publicly available geospatial data and it is possible to perform efficient data processing with fast visualizations (Gorelick et al., 2017). The geospatial mapping followed the Pipeline proposed by Van Den Hoogen et al. (2021), which allows creating nationwide predictions with high resolution, allowing also to evaluate bootstrap confidence intervals and standard deviation.

The pipeline makes use of Google Earth Engine (GEE) runs through the Python API. Predictions are created from the dataset of unprocessed observations (Mid IR (n= 2651), Vis NIR SWIR (n= 13657)), using the RF algorithm to identify the relationships with the set of covariates. This set is the combination of the raster of each covariate. Subsequently, the pipeline extracts the information from each covariate to the different unprocessed points, creating a training data set, from which a set of RF models is trained and evaluated. Subsequently, the RF hyperparameters are adjusted by a grid search procedure, and models with different combinations of hyperparameters are evaluated by k-fold random cross validation. The final model, which will be used to create the prediction map, corresponds to the best model from the grid search procedure based on R^2 (coefficient of determination), RMSE (root mean square error) and the ratio of performance to interquartile distance (RPIQ) of the calibration and validation data sets. The R^2 represents the variance explained, the RMSE the precision, and the RPIQ evaluates the error in the interquartile range of the data (Rosin et al., 2023).

To obtain the uncertainty estimates, the pipeline includes a stratified bootstrapping procedure, based on the Bootstrapping technique that simulates the distribution of samples to assign measures of accuracy to the model predictions. Therefore, the training dataset is sampled with replacement, using the biome as the stratification category, where each biome is proportionally represented in each bootstrap sample. One hundred resamples are performed and each is used in the classification of the final composite image containing the layers associated with the prediction map, bootstrap maps (mean, confidence intervals, standard deviation and coefficient of variation), and two interpolations versus extrapolation maps, and metrics of model accuracy and variable importance (Van Den Hoogen et al., 2021).

5.2.6. Data interpretation

Maps of the enzyme activity indexes determined from the Vis NIR SWIR and Mid IR spectral ranges were prepared for the three enzymes, applying color ramps in order to differentiate areas with higher or lower enzyme activity. Additionally, prediction interval maps were prepared in order to identify the areas with greater uncertainty.

5.2.7. Spatial validation

Given the strong interaction between Fe and Al oxides and functional groups associated with the products of beta-glucosidase, phosphatase and urease enzymatic activity (Cao et al. 2011, Parikh et al. 2014, Kleber et al. 2021, Yang et al. 2021, Viscarra Rossel et al. 2022), a visual comparison was made with the Fe, Al and Si oxide maps of Brazil constructed by Rosas et al. (2024) and the SOC map (Gomes et al., 2019).

Additionally, a correlation analysis was performed between the enzyme activity maps and the Fe, Al and Si oxides maps of Rosas et al. (2024). For this purpose, Spearman's correlation coefficient was used, being one of the most suitable and accurate metrics for this type of analysis, being robust and insensitive to outliers, applicable to data with normal and non-normal distribution, and effective in the characterization of linear and non-linear correlations (Zhang and Wang, 2023).

Adetunji et al. (2017) reports that the enzymatic activity of phosphatase and urease increases in leguminous crops, therefore, it is proposed to elaborate a visual validation in the region of Campo Novo do Parecis, Mato Grosso, where there is an area with soybean crops since 1995 next to pasture areas.

5.3. Results and Discussion

5.3.1. Index of enzyme activity

Table 1 shows the most important bands determined for each enzyme activity in each range of the electromagnetic spectrum. These bands were finalized from RFE and the comparison of the bands obtained by Rodriguez-Albarracin et al. (2024) for the enzyme activity mapping of the Piracicaba Sao Paulo area.

For the Vis NIR SWIR range, an R^2 for RFE of 0.67 was obtained for beta-glucosidase, 0.4 for Urease and 0.26 for phosphatase. For the Mid IR range, an R^2 of 0.74 was obtained for beta-glucosidase, 0.35 for urease and 0.26 for phosphatase. The models obtained for each spectral range and each enzyme activity are shown in Table 2.

Table 1. Bands of the Vis NIR SWIR and Mid IR spectral ranges of major importance for the prediction of urease, beta-glucosidase and phosphatase enzyme activity obtained from the Recursive Feature Elimination (RFE) algorithm and bands selected for the multivariate linear regression (MRA) models.

Enzyme	VisNIR SWIR		Mid IR	
	RFE	Selected	RFE	Selected
Urease	350-441, 445, 447, 452, 455, 459, 460, 463, 477, 479, 489, 553, 644, 659, 663, 784, 791-821, 842, 843, 849, 868, 895, 906, 969, 981, 982, 994, 999, 1010, 1050, 1051, 1056, 1057, 1058, 1060, 1061, 1076-1092, 1098, 1106, 1109, 1110, 1116, 2182, 2265, 2360, 2444, 2460-2499	350, 351, 352, 359, 368, 379, 398, 410, 430, 463, 489, 663, 798, 821, 868, 1106, 2360, 2470, 2499	614, 619, 620, 621, 627, 629, 630, 648, 650, 651, 652, 656, 657, 658, 662, 663, 676, 697, 836, 905, 1010, 1304, 1308, 1309, 1310, 1311, 1312, 1316, 1322, 1323, 1627	1627, 1310, 1010, 905, 836, 651, 629, 619, 606
Phosphatase	350-417, 422, 423, 431-439, 445, 711-715, 725, 726, 728, 729, 730, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 750, 751, 752, 755, 757, 758, 934, 940, 1054, 1231, 1240, 1249, 1250, 1259, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1356, 1375, 1376, 1377, 1410, 1411, 1414, 1415, 1416, 1417, 1418, 1523, 1703-1707, 1820-1840, 1869, 1870, 1871, 1873, 1874, 1980, 1983, 1984, 2071, 2072, 2103, 2138, 2139, 2142, 2143, 2180, 2423, 2424, 2428, 2440-2445, 2453, 2459, 2460, 2461, 2462, 2466, 2467, 2469, 2471-2500,	350, 355, 356, 357, 360, 361, 576, 667, 736, 740, 1911, 2441, 2476, 2482, 2498	600-623, 629, 635-646, 652, 656, 657, 661-688, 695-711, 735, 750, 805, 818, 821-824, 838, 875, 882, 887, 888, 889, 890, 913, 914, 920, 951, 1000-1020, 1038, 1040-1068, 1291, 1294, 1301-1324, 1331, 1332, 1334, 1340, 1341, 1342, 1361-1366, 1386, 1396, 1397, 1402-1412, 1458, 1474, 1602, 1604, 1611, 1612, 1641, 1649, 1653, 1869, 1912, 2010, 2011, 2372, 2533, 2534, 2830, 2849, 2893, 2897, 2917, 3449, 3496, 3498, 3613, 3614, 3620, 3622, 3626, 3672, 3694-3729, 3964, 3989, 3990, 3994-4000	3722, 3714, 3703, 3449, 2534, 2372, 1869, 1653, 1474, 1397, 1294, 1068, 875, 735, 670, 624, 616, 612, 606
Beta-glucosidase	360, 361, 362, 366, 369, 376, 388, 394, 398, 423, 432, 446, 535, 536, 537, 538, 541, 542, 543, 544, 545,	360, 361, 376, 388, 394, 423, 535, 543, 2499, 2500	617-640, 664, 683, 696, 705-708, 738, 740, 743, 765, 768, 776, 804, 819, 820, 886, 910-914, 992-1013, 1023, 1027, 1030, 1037-1057, 1060, 1175, 1177, 1186, 1187, 1190, 1199, 1200-1238, 1242, 1300-1341, 1348, 1350, 1351, 1352, 1771, 1774, 1844, 1850-1870, 1879, 2038, 2042, 2048, 3314, 3383, 3384, 3385, 3447, 3479, 3481, 3482, 3484, 3485, 3499, 3501-3504, 3520, 3539, 3540, 3551, 3552, 3568, 3569, 3570-3580, 3600, 3681-3684, 3697, 3698, 3700, 3714-3724, 3911, 3934, 3935, 3952-3959, 3968, 3971-4000	4000, 3956, 3383, 2042, 1857, 1826, 1581, 1311, 1310, 1243, 1054, 912, 886, 776, 667, 629, 617, 608

Table 2. Multivariate linear regression (MRA) equations for the prediction of urease, beta-glucosidase and phosphatase enzyme activity from the selected bands of the Vis NIR SWIR and Mid IR spectral ranges.

Enzyme	VisNIR SWIR	Mid IR
Urease	$12.30 + 1411.36 (410) - 1500.72 (463) + 949.22 (663) - 4750.01 (798) + 4277.31 (821) - 143.03 (1106) - 50.16 (2470)$	$6.13 + 562.28 (1310) - 2948.53 (651) + 3238.62 (629) - 945.31 (619)$
Phosphatase	$539.16 + 3340.61 (350) + 10585.21 (356) - 13872.10 (360) - 9709.36 (576) + 43224.81 (667) - 536484.72 (736) + 502451.57 (740) - 1037.61 (2441)$	$815.84 - 23409.77 (3714) - 11781.39 (2534) + 95234.45 (1653) - 29869.57 (1474) - 10903.26 (1068) - 66866.03 (875) + 95526.29 (735) + 67049.22 (670) - 24877.30 (606)$
Beta-glucosidase	$57.17 + 3140.16 (360) - 5047.09 (361) + 4507.84 (376) - 2815.09 (423) + 412.16 (535) - 129.69 (2500)$	$68.76 - 1551.95 (3956) + 1835.39 (3383) - 1582.82 (2042) - 1871.06 (1857) + 4878.35 (1826) + 57376.11 (1311) - 55386.80 (1310) + 4906.20 (912)$

In the Mid IR range, correlations of 0.72 were obtained for urease, 0.89 beta-glucosidase and 0.76 phosphatase. For the VIS NIR SWIR range 0.84 was obtained for urease, 0.76 betaglucosidase, 0.70 phosphatase. The models obtained satisfied the assumption of spatial independence of the residuals and normality.

The highest coefficients are related to the most important bands, for the case of urease in the Vis NIR SWIR range corresponds to the bands 798 and 821 nm, related to the NH, CH and CO functional groups of the labile forms of OM (Stenberg et al., 2010). The bands of greatest significance for phosphatase correspond to 667, 736 and 740nm. In the range 600-670 nm strongly humified material with strong presence of aromatic groups has been reported Sharma et al. (2021).

It is important to note that the bands between 350 – 376 nm reported in the phosphatase and beta-glucosidase models are related to the interaction of cell wall compounds of microorganisms and Fe oxides (Yang et al., 2022).

In general, it is observed that in this spectral range the bands with less importance in the model are related to the ranges above 2400 nm, in which methyl functional groups have been reported (Viscarra-Rossel and Behrens, 2010).

In relation to the Mid IR range the most important bands for urease correspond to 651 and 629 cm^{-1} , which correspond to CH_2 stretching vibrations of bacterial cellular compounds (Parikh et al., 2014). In the range 600-700 cm^{-1} different authors have reported the interaction of compounds associated with the cell walls of microorganisms with AIO groups of clay minerals (Deo et al. 2001, Demyan et al. 2012). For phosphatase the most important bands

correspond to 1653, 875, 735 and 670 cm^{-1} . The latter band is related to the interaction of bacterial wall phosphate groups with Fe oxides (Parikh et al., 2014). It is possible that the 1653 cm^{-1} band is related to aromatic C=C, amide II, CH bonds interacting with the surfaces of Fe oxides (Cao et al. 2011, McWhirter et al. 2002, Ladd et al. 1993). For beta-glucosidase the most important bands correspond to 1826, 1311, 1310 and 912 cm^{-1} . Between the range 650-1800 cm^{-1} the carbohydrate and protein response of cellular compounds is recognized (Ammann and Brandl, 2011), in the 1300 cm^{-1} ranges the influence of amide II, aliphatic and ester groups have been recognized (Parikh et al. 2014, Nicolas et al. 2017, Gao and Chorover 2009). Near the 1000 cm^{-1} range, the response of phosphate groups, carbohydrates, cellulose and polysaccharides has been recognized (Ojeda et al. 2008, Gao and Chorover 2009, Liu et al. 2015, Parikh et al. 2014).

5.3.2. Environmental covariates

For the mapping of enzyme activity indices related to the Vis NIR SWIR spectral range (Figure 2) the most important covariates correspond to annual precipitation (28.7%), elevation (13.1%), annual temperature (11.8%), SWIR 1 (8.5%), blue (6.8%), goethite (5.4%), SWIR 2 (4.3%), Northness (4%) and hematite (3.9%) for beta-glucosidase. For phosphatase, blue (15.2%), annual temperature (12.9%), annual precipitation (9.9%), SWIR 2 (9.4%), elevation (9.1%), hematite (8.5%), goethite (5.9%), kaolinite (4.6%), SWIR1 (4.4%), gibbsite (4.1%). For urease, precipitation (33.7%), elevation (21.6%), temperature (9%), blue (5%), SWIR 1 (4.4%), Eastness (3.7%) and goethite (3.6%).

For the Mid IR spectral range (Figure 3) the most important covariates for beta-glucosidase enzyme activity index mapping were temperature (13.6%), precipitation (12.2%), blue (11.7%), SWIR 1 (11.1%), goethite (10.3%), elevation (8.2%), hematite (6.1%), SWIR 2 (5.3%) and green (4.6%). For phosphatase were SWIR 2 (14.1%), temperature (13.6%), SWIR 1 (8.7%), kaolinite (7.9%), precipitation (6.8%), vertical curvature (6.6%), hematite (5.8%), blue (5.8%), NIR (5.1%), red (5.0%), horizontal curvature (4.9%). For urease blue (13.4%), precipitation (12.3%), SWIR 1 (11.1%), temperature (10.7%), hematite (10.1%), SWIR 2 (7.6%), kaolinite (6.3%), elevation (5.7%), goethite (4.5%).

In general, the importance of climatic covariates precipitation and temperature in mapping enzyme activity is observed (Figure 2 and 3). Beta-glucosidase has a close relationship with the C cycle (Adetunji et al., 2017) and the formation of simple sugars that are utilized by microorganisms (Xiao-Chang and Qin, 2006). It is possible to assume that, due to this relationship, beta-glucosidase activity responds to the decomposition of C by microorganisms

and therefore temperature and precipitation are important because they interfere with these decomposition rates (Lal, 2004). According to Shi et al., (2020), these variables have the greatest effect on C storage. In addition, precipitation plays a key role in biomass productivity that determines litter input to the soil (Chaplot et al., 2010), due to its influence on the volume, quality and quantity of mineralization (Zeraatpishe and Khormali 2012).

Adetunji et al., (2017) indicates that soil temperature and moisture affect the activity of urease, presenting an increase in activity with increasing temperature and a reduction of this with loss of moisture.

In relation to relief, topographic indices such as elevation, slope and landscape position influence the action of climatic factors on soils (Weil and Brady 2016, Lamichhane et al. 2019). Terrain parameters such as elevation and orientation have been used as predictive covariates of SOC (Sothe et al. 2022, Hengl et al. 2015). Because these variables control soil water status, litter mineralization dynamics, erosion and deposition processes (Hengl et al., 2015). In addition, ground elevation is related to temperature and thus responds to the rate of decomposition (Schindlbacher et al. 2010).

According to Amatulli et al. (2018) topography measured by elevation, orientation and slope, allows characterizing the spatial heterogeneity and abiotic environment for a given area, which is subsequently driven by hydrological, geomorphological and biological processes.

Slope curvatures influence the direction of water flow, soil erosion and soil moisture, thus influencing water availability (Bogaart and Troch, 2006). East and north are derived continuous topographic variables, calculated from orientation and slope (Fassnacht et al 2003), and therefore describe orientation in combination with slope. A value close to 1 for north indicates a slope exposed to low amount of solar radiation and close to -1 indicates steep slopes with high exposure to solar radiation (Amatulli et al., 2018).

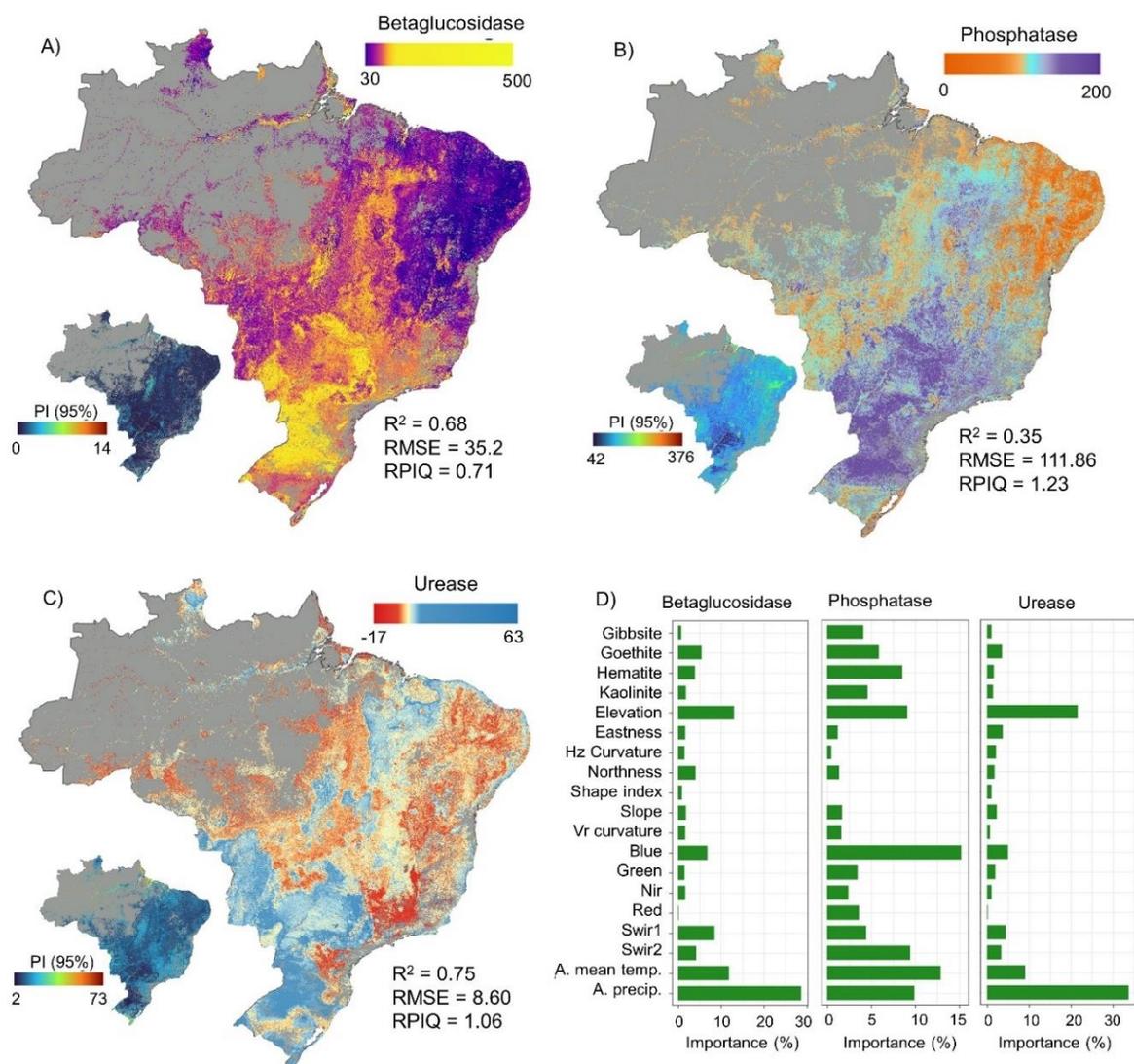


Figure 2. Prediction maps of the enzymatic activity of A) Beta-glucosidase, B) Phosphatase, C) Urease, considering the prediction models created from the Vis NIR SWIR spectral range, together with the prediction interval (PI) maps and D) importance of the covariates used in the mapping by applying the random forest (RF) algorithm. Where A. mean temp (annual mean temperature), A. precip. (annual precipitation), Vr curvature (vertical curvature), Hz curvature (horizontal curvature).

Phosphatase activity is affected by soil moisture, being reduced under drought effects (Adetunji et al., 2017), which explains the importance of temperature, precipitation and elevation in the prediction of this enzyme activity (Figure 2). It is possible to observe that the most important variable was the SYSI blue band, this band includes the range between 450 and 520 nm, and it is observed that, in the model associated with the phosphatase activity index, the 576 nm band contributes. Yang et al. (2022) reported the interaction of Fe oxides with fungal diversity in the 410 and 460 nm bands, that is, the blue band may be related to the

interaction of cell wall compounds that are related to phosphatase activity (Parikh et al., 2014) and Fe oxide surfaces. Dematte et al. (2014) indicates the presence of hematite in the 550 nm band and goethite around 440-470nm, and clearly the importance of these two oxides as covariates in mapping enzyme activity in general is observed. This blue band was also of importance in urease mapping, because the 410 and 463 nm bands are important in modeling the activity index of this enzyme (Table 2).

SWIR 2 (2080-2350 nm) is also important in phosphatase mapping (Figure 2), possibly due to its proximity to the 2441 nm band, which participates in the model of the activity index of this enzyme. Zornoza et al. (2008) report the range 2270-2630 nm, associated with phosphatase. Viscarra-Rossel and Behrens (2010) report the presence of fulvic acids and carbohydrates at 2336 and 2381 nm.

Hematite and goethite were important covariates in the mapping of the 3 enzymes (Figure 2 and 3), different authors have reported the interaction of functional groups present in the cell walls with these Fe oxides (Yang et al. 2022, Parikh et al. 2014, Sharma et al. 2021, Ojeda et al. 2008, Gao and Chorover 2009, Cao et al. 2011). Kaolinite also appears as an important covariate for phosphatase, because the interaction of phosphate and amide II groups with AlO groups of clay minerals has been reported (Deo et al., 2001).

5.3.3. Spatial prediction

Prediction maps of enzyme activity indices based on the Vis NIR SWIR spectral range had an R^2 of 0.68 and RMSE of 35.2 for beta-glucosidase, 0.35 and 111.86 for phosphatase, 0.75 and 8.6 for urease (Figure 2). For the Mid IR spectral range, R^2 values of 0.38 and RMSE of 27.2 were obtained for beta-glucosidase, 0.11 and 760 for phosphatase and 0.45 and 3.64 for urease (Figure 3).

Safanelli et al. (2021b) reported an R^2 of 0.5 for the SOC map and 0.74 for clay in the agricultural areas of Brazil obtained from the RF algorithm. Gomes et al. (2019) reported an R^2 of 0.3 for the SOC map of Brazil. Rosas et al. (2024) reports R^2 of 0.65 for Fe oxide maps of agricultural areas of Brazil. Rosin et al. (2023) elaborated hematite, goethite, gibbsite and kaolinite maps for Brazil and obtained R^2 of 0.64, 0.4, 0.29 and 0.2, respectively. The maps obtained in the present study (Figure 2 and 3) present a high coefficient of determination especially with the Vis NIR SWIR spectral range, with the exception of phosphatase.

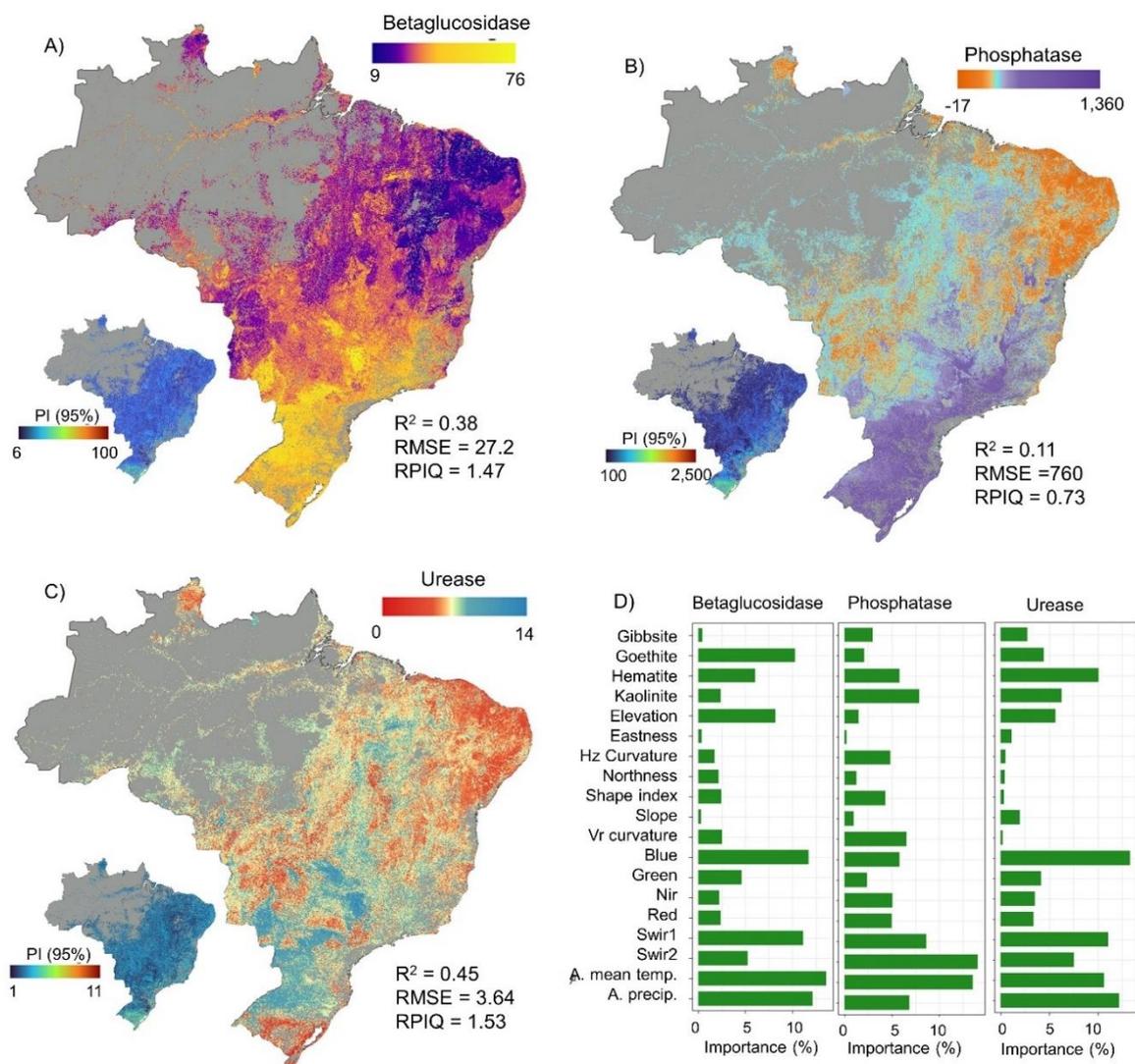


Figure 3. Prediction maps of the enzymatic activity of A) Beta-glucosidase, B) Phosphatase, C) Urease, considering the prediction models created from the Mid IR spectral range, together with the prediction interval (PI) maps and D) importance of the covariates used in the mapping by applying the random forest (RF) algorithm. Where A. mean temp (annual mean temperature), A. precip. (annual precipitation), Vr curvature (vertical curvature), Hz curvature (horizontal curvature).

There are not many studies that have mapped the enzymatic activity of the soil, Rodriguez-Albarracín et al. (2024), elaborated the maps of the enzymatic activity indexes for the region of Piracicaba, Sao Paulo, and obtained R^2 values of 0.43 for beta-glucosidase, 0.52 for urease and 0.49 for phosphatase, using the Vis NIR SWIR spectral range, and R^2 of 0.3 for beta-glucosidase, 0.42 urease and 0.49 phosphatase with the Mid IR spectral range.

The prediction interval (PI) maps are observed (Figures 2 and 3), which indicate that higher values are related to greater uncertainty. It is observed that beta-glucosidase and urease

both in the predictions with Vis NIR SWIR and Mid IR are the ones that present lower uncertainty, however, it is important to highlight that the urease model with Mid IR presents improvements contrary to beta-glucosidase. The most difficult variable to map was phosphatase because it presented the worst validation metrics.

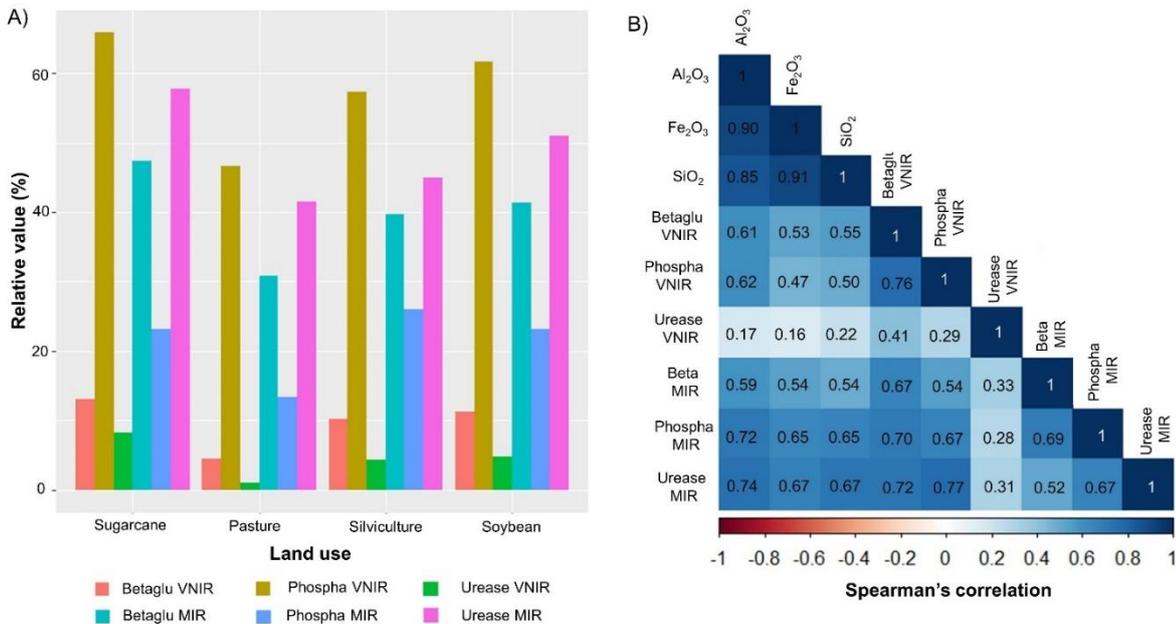


Figure 4. Distribution of relative enzyme activity of beta-glucosidase (Betaglu), Phosphatase (Phospha) and Urease, determined from the Vis NIR SWIR (VNIR) and Mid IR (MIR) prediction models, according to the agricultural uses reported by MAPBIOMAS (2022).

Spatially, urease and phosphatase show similar behavior, observing that areas with lower phosphatase activity also coincide with low urease activity and vice versa (Figure 2 and 3). The areas with the greatest differences are observed in the states of Minas Gerais, Paraná, Santa Catarina and Rio Grande do Sul (Figure 2), where areas with lower urease activity present higher phosphatase activity, which seems to follow the behavior of the Fe oxide map elaborated by Rosas et al. (2024). On the other hand, areas with lower urease activity seem to follow the pattern of the gibbsite map elaborated by Rosin et al. (2023) and the aluminum oxide map, however, a higher urease activity would be expected due to the relationship presented by the NH functional groups and Al-OH bonds (Davinic et al., 2012), however, Kleber et al. (2021) indicates that there is competition between different organic compounds for adsorption sites with inorganic ions, where stronger bonds prevail, which in these areas could be associated with phosphate group ligands on mineral surfaces (Omoike and Chorover 2006, Parickh et al. 2014). However, the behavior of urease in these areas may be due to the

higher elevation of the terrain found in this area (Rosin et al., 2023), resulting in a possible reduction in temperature and, therefore, a decrease in the activity of this enzyme (Machuca et al., 2015).

It is observed that in the southern part of the state of Rio Grande do Sul, there is a decrease in the activity of urease and phosphatase (Figures 2 and 3). In this area the slopes are lower (Rosin et al., 2023) and according to IBGE (2021) there are "planosolo" soils that are flooded, additionally, according to MapBiomass (2020) it is an area where rice is cultivated, which explains the reduction in the activity of this enzyme, due to the fact that commonly in this land use fertilization is conducted where nitrification inhibitors are applied (Marchesan et al., 2013).

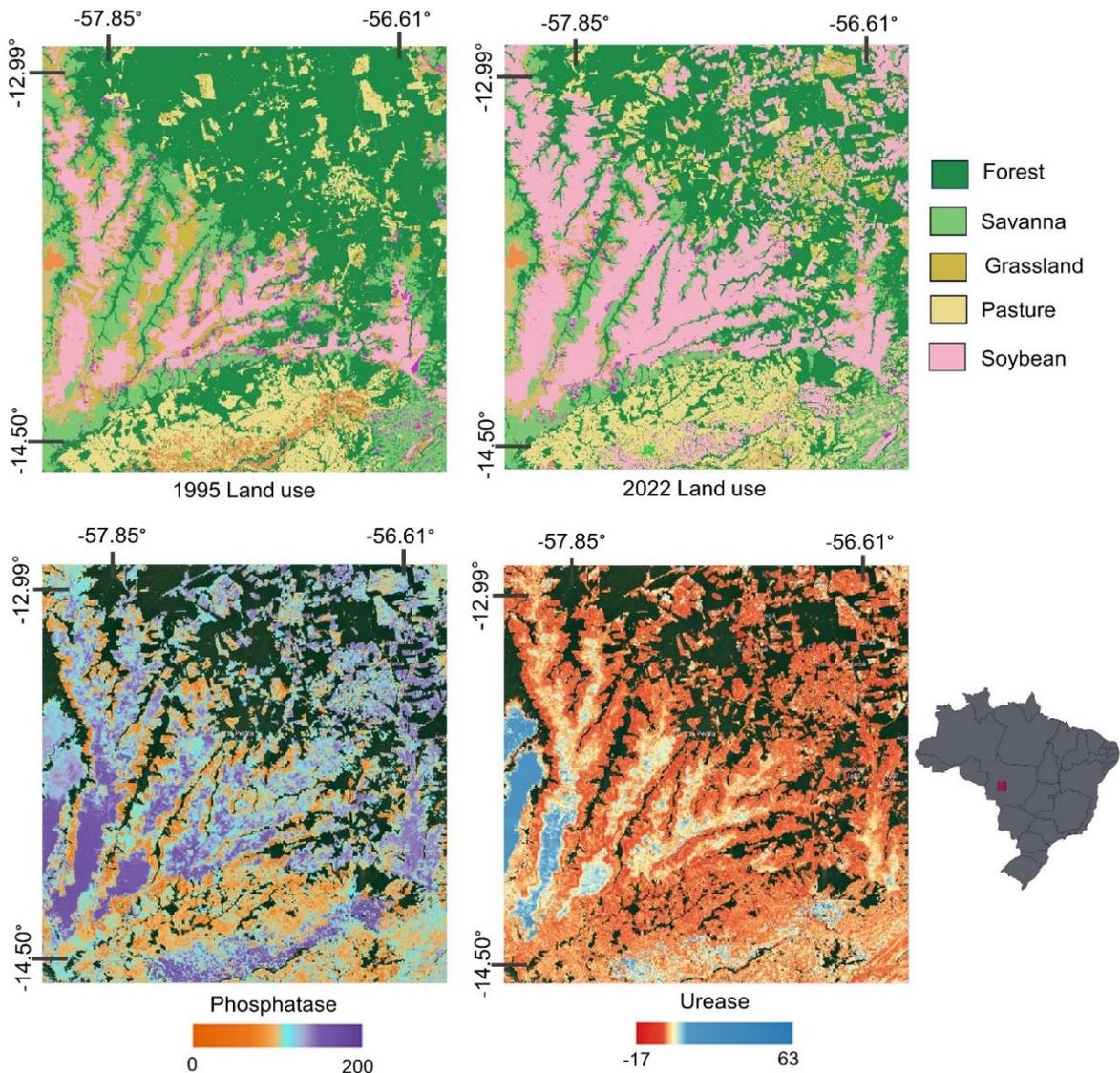


Figure 5. Analysis of phosphatase and urease enzyme activity determined from Vis NIR SWIR prediction models in an area traditionally cultivated with soybean (1995-2022) in the region of Campo Novo do Parecis, Mato Grosso.

Beta-glucosidase activity follows the distribution of the SOC map (Gomes et al., 2019), where higher enzymatic activity is related to higher C content (Figures 2 and 3), which is in agreement with that reported by Zhang et al. (2020) and Henríquez et al. (2014), who point out a strong correlation between beta-glucosidase activity and SOC. Some areas in the state of São Paulo and Mato Grosso do Sul with high beta-glucosidase enzymatic activity respond to hematite and goethite dominance (Rosin et al., 2023). However, when comparing the response of this enzymatic activity with the Fe and Al oxide maps of Rosas et al. (2024), a reduction of enzymatic activity is observed with increasing oxide content, this is because a large part of the

functional groups associated with betaglucosidase are related to labile forms of C and are the main source of energy for microorganisms (Adetunji et al., 2017).

5.3.4. Spatial validation

In general, high correlations were obtained between Fe, Al and Si oxide maps and enzyme activity (Figure 4). High correlations of beta-glucosidase with Fe_2O_3 (0.53), Al_2O_3 (0.61) and SiO_2 (0.55) were observed, because the activity of this enzyme is also related to aromatic, ester and carbonyl functional groups interacting with the oxides surfaces (Gao and Chorover 2009, Parikh et al 2014). It is important to clarify that SiO_2 does not correspond to quartz, this is found in the structure of clay fraction minerals such as hematite, goethite, gibbsite, kaolinite and montmorillonite (Rosas et al., 2024), therefore, the correlation reflects that interaction of the products of the enzymatic activity with the surfaces of the clay minerals, especially the carbonyl functional groups that interact with the Si oxides (Calderón et al. 2011, Davinic et al. 2012).

The correlations between phosphatase activity and oxides are also high (Figure 4) and improve when compared to the Mid IR spectrum maps, possibly due to the fact that in this spectral range there are more specific bands related to the products of this enzyme (Table 1). The lowest correlations are presented for the enzymatic activity of urease with the Vis NIR SWIR spectral range, possibly because in this spectral range there are fewer specific bands associated with this enzymatic activity and, like phosphatase, there is a large number of bands resulting from the RFE algorithm (Table 1), contrary to the Mid IR range which has a more specific number of bands for urease activity and this is reflected in the lower uncertainty of the map (Figure 3) and greater correlation with the different oxides.

Adetunji et al. (2017) and Meyer et al. (2015) highlight the influence of land use on enzyme activity, and this is identified in Figure 4, where lower activity is presented in pasture areas, contrary to forest, soybean and sugarcane areas. In these last two land uses, higher enzyme activity is observed in general. Beta-glucosidase activity is improved because in these uses commonly crop material is left covering the soil, which can be converted into labile forms of C for energy production by microorganisms (Kleber et al. 2015, Parikh et al., 2014) and furthermore this enzyme activity is related to the rate of OM oxidation (Henriquez et al., 2014) and increases with OM input to the soil (Meyer et al., 2015) and is higher in soils where litter is maintained than in those without (Yang et al., 2003). Phosphatase activity is higher in soybean crops because legumes require more phosphorus in the N fixation process (Makoi and Ndakidemi, 2008). Higher urease activity is observed in sugarcane crops because nitrogen

fertilization is commonly used, unlike soybean crops, and this type of fertilization decreases the activity of this enzyme (Mohammadi, 2011).

A case study in Mato Grosso, in the region of Campo Novo do Parecis (Figure 5) in an area traditionally used in Soybean crops (1995 - 2022), close to a pasture area, confirms the strong relationship of phosphatase enzyme activity with legume crops reported by Makoi and Ndakidemi, (2008) and Adetunji et al. (2017), because compared to pasture areas, a high activity of this enzyme is observed, as well as of urease, due to the fact that N availability favors the increase of urease activity (Adetunji et al. 2017, Henríquez et al. 2014).

5.4. Conclusions

Equations were obtained for the determination of the enzymatic activity of beta-glucosidase, phosphatase and urease for each spectral range (Vis NIR SWIR and Mid IR) based on multiple linear regression models, which presented correlations that oscillated from 0.89 to 0.7, presenting the highest correlations with the models based on Mid IR.

The bands involved in the prediction models and therefore the most important, for the Mid IR spectral range for beta-glucosidase corresponded to 3956, 3383, 2042, 1857, 1826, 1311, 1310 and 912 cm^{-1} , for urease 1310, 651, 6249 and 619 cm^{-1} , for phosphatase 3714, 2534, 1653, 1474, 1068, 875, 735, 670 and 606 cm^{-1} . For Vis NIR-SWIR spectral range for beta-glucosidase corresponded to 360, 361, 376, 423, 535 and 2500 nm, urease 410, 463, 663, 798, 821, 1106 and 2470 nm, phosphatase 350, 356, 360, 576, 667, 736 and 2441 nm.

Prediction maps of enzyme activity indices were obtained based on the DSM framework and Vis NIR-SWIR spectral range information with R^2 of 0.68 for beta-glucosidase, 0.35 phosphatase and 0.75 urease. For the Mid IR spectral range much lower R^2 values were obtained, however, the prediction interval (PI) decreased for urease.

Climatic covariates were the most important for spatial prediction, due to their influence on enzyme activity. Additionally, the blue, SWIR 1 and SWIR 2 bands of the SYSI also presented high importance because the range of each band presents strong relationships with the bands of higher importance that participated in the prediction models. The mineralogical variables also presented high importance, especially hematite and goethite due to the interaction between the surfaces of these minerals and the functional groups related to the activity of these three enzymes. Additionally, a strong correlation was observed between the enzymatic activity maps and the Fe, Al and Si oxide maps, which oscillated from 0.74 to 0.47, being higher with the maps based on the Mid IR spectral range.

Variation of enzyme activity was identified in the different land uses and it was observed that phosphatase and urease activity is higher in crops with soybean tradition, such activity decreases in pasture areas. Crops that commonly leave material in the soil after harvest, such as sugar cane and soybeans, show an increase in beta-glucosidase activity. The latter enzyme had a strong visual correlation with the SOC map due to its close relationship with the C cycle.

References

- ADETUNJI, A., LEWU, F., MULIDZI, R. AND B. NCUBE. 2017. The biological activities of β -glucosidase, phosphatase and urease as soil quality indicators: a review. *Journal of soil science and plant nutrition*, 17(3), 794-807. <https://dx.doi.org/10.4067/S0718-95162017000300018>.
- ALVARES, C.A., STAPE, J.L., SENTELHAS, P.C., DE MORAES GONÇALVES, J.L. & G. SPAROVEK. 2013. Koppen's climate classification map for Brazil. *Meteorol. Z.* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- AMATULLI, G., DOMISCH, S., TUANMU, M., PARMENTIER, B., RANIPETA, A., MALCZYK, J. AND W. JETZ. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *SCIENTIFIC DATA* | 5:180040 | DOI: 10.1038/sdata.2018.40
- AMMANN, A. & H. BRANDL. 2011. Detection and differentiation of bacterial spores in a mineral matrix by Fourier transform infrared spectroscopy (FTIR) and chemometrical data treatment. *BMC Biophysics* 2011, 4:14. <http://www.biomedcentral.com/2046-1682/4/14> (14 July 2011)
- BALOTA, E.L. AND J.C. CHAVES. 2010. Enzymatic activity and mineralization of carbon and nitrogen in soil cultivated with coffee and green manures. *Revista Brasileira de Ciência do Solo*. 34, 1573-1583.
- BANDICK, A.K. AND R.P. DICK. 1999. Field management effects on soil enzyme activities. *Soil Biology and Biochemistry*. 31, 1471-1479.
- BANERJEE, A., SANYAL, S. AND S. SEN. 2012. Soil phosphatase activity of agricultural land: A possible index of soil fertility. *Agricultural Science Research Journals*. 2, 412-419.
- BOGAART, P. W. AND P.A. TROCH. 2006. Curvature distribution within hillslopes and catchments and its effect on the hydrological response. *Hydrology and Earth System Sciences Discussions* 3, 1071–1104
- BREIMAN, L. 2001. Random forests. *Machine Learning*, 45 (2001), pp. 5-32

- BUI, E., HENDERSON, B. & K. VIERGEVER. 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Glob. Biogeochem. Cycles*, 23 (4) (2009)
- CALDERÓN, F.J., MIKHA, M.M., VIGIL, M.F., NIELSEN, D.C., BENJAMIN, J.G. AND J.B. REEVES III, J.B. 2011. Diffuse-reflectance mid-infrared spectral properties of soils under alternative crop rotations in a semi-arid climate. *Commun. Soil Sci. Plant Anal.* 42, 2143–2159.
- CAO, Y., WEI, X., CAI, P., HUANG, Q., RONG, X. & W. LIANG. 2011. Preferential adsorption of extracellular polymeric substances from bacteria on clay minerals and iron oxide. *Colloids and Surfaces B: Biointerfaces* 83 (2011) 122–127
- CHAPLOT, V., BOUAHOM, B. & C. VALENTIN. 2010. Soil organic carbon stocks in Laos: spatial variations and controlling factors *Glob. Chang. Biol.*, 16 (4) (2010), pp. 1380-1393
- CHODAK, M. 2011. Near-infrared spectroscopy for rapid estimation of microbial properties in reclaimed mine soils. *Journal of Plant Nutrition and Soil Science* 174, 702-709.
- COHEN, M.J., PRENGER, J.P. & W.F. DEBUSK. 2005. Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. *Journal of Environmental Quality* 34, 1422e1434.
- CONDRON, L.M., TURNER, B.L., CADE-MENUN, B.J., SIMS, J. AND A. SHARPLEY. 2005. Chemistry and dynamics of soil organic phosphorus. *Phosphorus: Agriculture and the environment*. 87-121.
- COSTANTINI, E. & S. PRIORI. 2023. Soil quality and health key indicators. *Encyclopedia of Soils in the Environment*, Second Edition. <https://doi.org/10.1016/B978-0-12-822974-3.00024-0>
- CUTLER, A., CUTLER, D.R. AND J.R. STEVENS. 2012. Random Forests, in: *Ensemble Machine Learning*. Springer, New York, NY, New York, NY, pp. 157 175. https://doi.org/10.1007/978-1-4419-9326-7_5
- DAVINIC, M., FULTZ, L.M., ACOSTA-MARTINEZ, V., CALDERON, F.J., COX, S.B., DOWD, S.E., ALLEN, V.G., ZAK, J.C. & J. MOORE-KUCERA. 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. *Soil Biol. Biochem.* 46, 63–72. <https://doi.org/10.1016/j.soilbio.2011.11.012>
- DEMATTÊ, J.A.M., BELLINASSO, H., ROMERO, D.J. AND C.T. FONGARO. 2014. Morphological interpretation of reflectance spectrum (MIRS) using libraries looking towards soil classification. *Sci. Agric.* 71, 509–520. <https://doi.org/10.1590/0103-9016-2013-0365>.
- DEMATTÊ, J., FONGARO, C., RIZZO, C. & J.L. SAFANELLI. 2018. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.*, 212 (2018), pp. 161-175.

- DEMATTÊ, J.A.M., DOTTO, A.C., PAIVA, A.F.S., SATO, M. V., DALMOLIN, R.S.D., DE ARAÚJO, M. DO S.B., DA SILVA, E.B., NANNI, M.R., TEN CATEN, A., NORONHA, N.C., LACERDA, M.P.C., DE ARAÚJO FILHO, J.C., RIZZO, R., BELLINASO, H., FRANCELINO, M.R., SCHAEFER, C.E.G.R., VICENTE, L.E., DOS SANTOS, U.J., DE SÁ BARRETTO SAMPAIO, E. V., MENEZES, R.S.C., DE SOUZA, J.J.L.L., ABRAHÃO, W.A.P., COELHO, R.M., GREGO, C.R., LANI, J.L., FERNANDES, A.R., GONÇALVES, D.A.M., SILVA, S.H.G., DE MENEZES, M.D., CURI, N., COUTO, E.G., DOS ANJOS, L.H.C., CEDDIA, M.B., PINHEIRO, É.F.M., GRUNWALD, S., VASQUES, G.M., MARQUES JÚNIOR, J., DA SILVA, A.J., BARRETO, M.C. D. V., NÓBREGA, G.N., DA SILVA, M.Z., DE SOUZA, S.F., VALLADARES, G.S., VIANA, J.H.M., DA SILVA TERRA, F., HORÁK-TERRA, I., FIORIO, P.R., DA SILVA, R.C., FRADE JÚNIOR, E.F., LIMA, R.H.C., ALBA, J.M.F., DE SOUZA JUNIOR, V.S., BREFIN, M.D.L.M.S., RUIVO, M.D.L.P., FERREIRA, T.O., BRAIT, M.A., CAETANO, N.R., BRINGHENTI, I., DE SOUSA MENDES, W., SAFANELLI, J.L., GUIMARÃES, C.C.B., POPPIEL, R.R., E SOUZA, A.B., QUESADA, C.A. & DO COUTO, H.T.Z. 2019. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma* 354, 113793. <https://doi.org/10.1016/j.geoderma.2019.05.043>
- DEMATTÊ, J.A.M., SAFANELLI, J.L., POPPIEL, R.R., RIZZO, R., SILVERO, N.E.Q., MENDES, W. DE S., BONFATTI, B.R., DOTTO, A.C., SALAZAR, D.F.U., MELLO, F.A. DE O., PAIVA, A.F. DA S., SOUZA, A.B., SANTOS, N.V. DOS, MARIA NASCIMENTO, C., MELLO, D.C. DE, BELLINASO, H., GONZAGA NETO, L., AMORIM, M.T.A., RESENDE, M.E.B. DE, VIEIRA, J. DA S., QUEIROZ, L.G. DE, GALLO, B.C., SAYÃO, V.M. AND C.J. LISBOA. 2020. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. *Sci Rep* 10, 1 11. <https://doi.org/10.1038/s41598-020-61408-1>
- DEMYAN, M., RASCHE, F., SCHULZ, E., BREULMANN, M., MULLER, T. & G. CADISCH. 2012. Use of specific peaks obtained by diffuse reflectance Fourier transform mid-infrared spectroscopy to study the composition of organic matter in a Haplic Chernozem. *European Journal of Soil Science*, April 2012, 63, 189 – 199. doi: 10.1111/j.1365-2389.2011.01420.x
- DEO, N., NATARAJAN, K. & P. SOMASUNDARAN. 2001. Mechanisms of adhesion of *Paenibacillus polymyxa* onto hematite, corundum and quartz. *Int. J. Miner. Process.* 62 2001 27–39
- DICK, R.P.; BREAKWELL, D.P. & R.F. TURCO. 1996. Soil enzymes activities and biodiversity measurements as integrative microbiological indicators. In: DORAN, J.W.; & JONES, A.J., eds. *Methods for assessing soil quality*. Soil Soc. Am., 1996, p. 247-271. Special publication
- DI IORIO, E., CIRCELLI, L., LORENZETTI, R., COSTANTINI, E., PERL, S. & C. COLOMBO. 2019. Estimation of andic properties from Vis-NIR diffuse reflectance spectroscopy for volcanic soil classification. *Catena* 182 (2019) 104109

- DINERSTEIN, E., OLSON, D., JOSHI, A., VYNNE, C., BURGESS, N., WIKRAMANAYAKE, E., HAHN, N., PALMINTERI, S., HEDAO, P., NOSS, R., HANSEN, M., LOCKE, H., ELLIS, E., JONES, B., BARBER, C., HAYES, R., KORMOS, C., MARTIN, V., CRIST, E., SECHREST, W., RICE, L., BAILLIE, J., WEEDEN, D., SUCKLING, K., DAVIS, C., SIZER, N., MOORE, R., THAU, D., BIRCH, T., POTAPOV, P., TURUBANOVA, S., TYUKAVINA, A., DE SOUZA, N., PINTEA, L., BRITO, J., LLEWELLYN, O., MILLER, A., PATZELT, A., GHAZANFAR, S., TIMBERLAKE, J., KLÖSER, H., SHENNAN-FARPÓN, Y., KINDT, R., LILLESØ, J., van BREUGEL, P., GRAUDAL, L., VOGEL, M., AL-SHAMMARI, K. AND M. SALEEM. 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience* June 2017 / Vol. 67 No. 6. <https://doi.org/10.1093/biosci/bix014>.
- DORAN, J.W., ZEISS, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. *Applied Soil Ecology*. Vol. 15
- FICK, S.E. & R.J. HIJMANS. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>.
- FOLLMER, C. 2008. Insights into the role and structure of plant ureases. *Phytochemistry*. 69, 18-28.
- FASSNACHT, S. R., DRESSLER, K. A. AND R.C. BALES. 2003. Snow water equivalent interpolation for the Colorado River Basin from snow telemetry (SNOTEL) data. *Water Resources Research* 39, 8
- GAO, X. AND J. CHOROVER. 2009. In-situ monitoring of *Cryptosporidium parvum* oocyst surface adhesion using ATR-FTIR spectroscopy. *Colloids and Surfaces B: Biointerfaces* 71 (2009) 169–176. doi:10.1016/j.colsurfb.2009.02.003
- GORELICK, N., HANCHER, M., DIXON, M., ILYUSHCHENKO, S., THAU, D. & R. MOORE. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202 (2017) 18–27. <http://dx.doi.org/10.1016/j.rse.2017.06.031>
- GÓMEZ, J., SCHOBENHAUS, C. AND N.E. MONTES. 2019. Geological Map of South America 2019. Scale 1:5 000 000. 10.32685/10.143.2019.929.
- GOMES, L., FARIA, R., DE SOUZA, E., VELOSO, G., SCHAEFER, C. & E.I.F. FILHO. 2019. Modelling and mapping soil organic carbon stocks in Brazil *Geoderma*, 340 (2019), pp. 337-350
- GÓMEZ, A., VAN LIER, Q., SILVERO, N., INFORSATO, L., DE MELO, M., RODRÍGUEZ-ALBARRACÍN, H., ROSIN, N., FIM ROSAS, J., RIZZO, R. & J.A.M DEMATTÊ. 2023. Digital mapping of the soil available water capacity: tool for the resilience of agricultural systems to climate change. *Science of the Total Environment* 882 (2023) 163572

- GRESCHUK, L., ARAÚJO, M., RODRÍGUEZ-ALBARRACÍN, H., BELLINASO, H., SILVERO, N., PAIVA, A., POPPIEL, R., ROSIN, N., CAMPOS, L., DALMOLIN, R., BALLESTER, M. & J. DEMATTÊ. 2022. Combining spectral ranges for soil discrimination: A case study in the State of Maranhao – Brazil. *Geoderma Regional* 29 (2022) e00507. <https://doi.org/10.1016/j.geodrs.2022.e00507>
- GUYON, I., WESTON, J., BARNHILL, S. & V. VAPNIK. 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. <https://doi.org/10.1023/A:1012487302797>
- HENGL, T., HEUVELINK, G., KEMPEN, B., LEENAARS, J., WALSH, M., SHEPHERD, K., SILA, A., MACMILLAN, R., MENDES DE JESUS, J., TAMENE, L. & J.E. TONDOH. 2015. Mapping soil properties of Africa at 250 m resolution: Random Forests significantly improve current predictions. *PLoS One*, 10 (6) (2015), Article e0125814, [10.1371/journal.pone.0125814](https://doi.org/10.1371/journal.pone.0125814)
- HENRÍQUEZ, C., URIBE, L., VALENCIANO, A. AND R. NOGALES. 2014. Actividad enzimática del suelo -Deshidrogenasa, β -Glucosidasa, Fosfatasa y Ureasa- bajo diferentes cultivos. *Agronomía Costarricense*, 38(1), 43-54. Retrieved January 31, 2024, from http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0377-94242014000100003&lng=en&tlng=es.
- HOBLEY, E., WILSON, B., WILKIE, A., GRAY, J AND T. KOEN. 2015. Drivers of soil organic carbon storage and vertical distribution in Eastern Australia. *Plant Soil*, 390 (2015), pp. 111-127
- HOGUE, M., WOHLING, T. & W. NOWAK. 2018. A primer for Model Selection: The Decisive Role of Model Complexity. *Water Resources Research*, 54, 1688-1715. <https://doi.org/10.1002/2017WR021902>
- KARACA, A., CETIN, S.C., TURGAY, O.C. AND R. KIZILKAYA. 2010. Soil enzymes as indication of soil quality. *Soil enzymology*. Springer.
- KHALEDIAN, Y. AND B.A. MILLER. 2020. Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model.* 81, 401–418. <https://doi.org/10.1016/j.apm.2019.12.016>
- KLEBER, M., EUSTERHUES, K., KEILUWEIT, M., MIKUTTA, C., MIKUTTA, R. & P.S. NICO. 2015. Mineral-organic associations: formation, properties, and relevance in soil environments. *Advances in Agronomy* 1–140.
- KLEBER, M., BOURG, I., COWARD, E., HANSEL, C., MYNENIS, S. AND N. NUNAN. 2021. Dynamic interactions at the mineral–organic matter interface. *Nat Rev Earth Environ* 2, 402–421 (2021). <https://doi.org/10.1038/s43017-021-00162-y>
- KUHN, M. 2021. caret: Classification and Regression Training. R package version 6.0-88. <https://CRAN.R-project.org/package=car>
- LAL R. 2004. Soil carbon sequestration impacts on global climate change and food security *Science*, 304 (2004), pp. 1623-1627

- LADD, J., FOSTER, R.C. & J.O. SKJEMSTAD. 1993. Soil structure: carbon and nitrogen metabolism. *Geoderma*, 56 (1993) 401-434. Elsevier Science Publishers B.V., Amsterdam. <https://doi.org/10.1016/B978-0-444-81490-6.50034-0>
- LAMICHHANE, S., KUMAR, L. & B. WILSON. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* Volume 352, 15 October 2019, Pages 395-413
- LEHMANN, J., BOSSIO, D.A., KÖGEL-KNABNER, I. AND M.C. RILLIG. 2020. The concept and future prospects of soil health *Nature Reviews Earth & Environment*, 1 (10) (2020), pp. 544-553
- LIU, Y., LI, O., LI, Y., BAO, J., HU, Z., HAO, D., SONG, D., WANG, Y. & M. YANG. 2015. Detection of Nanoscale Soil Organic Matter by Middle Infrared Spectrum for Forensic Science. *Hindawi Publishing Corporation Journal of Chemistry* Volume 2015, Article ID 189421. <http://dx.doi.org/10.1155/2015/189421>
- LIU, X. AND J. CHEN. 2021. Variable Selection for the Spatial Autoregressive Model with Autoregressive Disturbances. *Mathematics* 2021, 9, 1448. <https://doi.org/10.3390/math9121448>
- LORENZ, K., LAL, R., PRESTON, C. & K.G.J. NIEROP. 2017. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules. *Geoderma*, 142 (2007), pp. 1-10, [10.1016/j.geoderma.2007.07.013](https://doi.org/10.1016/j.geoderma.2007.07.013)
- MA, Y., MINASNY, B., MALONE, B.P. AND A.B. MCBRATNEY. 2019. Pedology and digital soil mapping (DSM). *Eur J Soil Sci* 70, 216–235. <https://doi.org/10.1111/EJSS.12790>
- MA, C.Y., CI, K.D., ZHU, J., SUN, Z.L., LIU, Z.X., LI, X.Y., TANG, C., WANG, P. & Z.M. LIU. 2021. Impacts of exogenous mineral silicon on cadmium migration and transformation in the soil-rice system and on soil health. *Sci Total Environ.* 759, 143501.
- MARCHESAN, E., GROHS, M., WALTER, M., DA SILVA, L. AND T. FORMENTINI. 2013. Agronomic performance of rice to the use of urease inhibitor in two cropping systems. *Rev. Ciênc. Agron.*, v. 44, n. 3, p. 594-603, jul-set , 2013
- MACHUCA, A., CUBA-DÍAZ, M. AND C. CÓRDOVA. 2015. Enzymes in the rhizosphere of plants growing in the vicinity of the polish arctowski antarctic station. *Journal of soil science and plant nutrition.* 15, 833-838.
- MAKOI, J.H. AND P.A. NDAKIDEMI. 2008. Selected soil enzymes: Examples of their potential roles in the ecosystem. *African Journal of Biotechnology.* 7.
- MARTINEZ, C. AND M. TABATABAI. 1997. Decomposition of biotechnology by-products in soils. *Journal of Environmental Quality.* 26, 625-632.
- MCBRATNEY, A., SANTOS, M. & B. MINASNY. 2003. On digital soil mapping. *Geoderma*, 117 (2003), pp. 3-52

- MCWHIRTER, M.J., MCQUILLAN, A.J. & P.J. BREMER. 2002. Influence of ionic strength and pH on the first 60 min of *Pseudomonas aeruginosa* attachment to ZnSe and to TiO₂ monitored by ATR-IR spectroscopy. *Colloids Surf. B Biointerfaces* 26, 365–372.
- MENDES, W., DEMATTE, J.A.M., ROSIN, N., TERRA, F., POPPIEL, R., URBINA-SALAZAR, D., BOECHAT, C., SILVA, E., CURI, N., SILVA, S., SANTOS, U. & G. VALLADARES. 2022. The Brazilian soil Mid-infrared Spectral Library: The Power of the Fundamental Range. *Geoderma* 415 (2022) 115776. <https://doi.org/10.1016/j.geoderma.2022.115776>
- MERINO, C., GODOY, R. & F. MATUS. 2016. Soil enzymes and biological activity at different levels of organic matter stability. *Journal of Soil Science and Plant Nutrition*. 16, 14-30.
- MEYER, A.H., WOOLDRIDGE, J. & J.F. DAMES. 2015. Variation in urease and β -glucosidase activities with soil depth and root density in a cripp's pink /m7 apple orchard under conventional and organic management. *South African Journal of Plant and Soil*. 32, 227-234.
- MINASNY, B., MCBRATNEY, A. & B.P. MALONE. 2013. Wheeler Digital mapping of soil carbon. *Adv. Agron.*, 118 (2013), pp. 1-47
- MOHAMMADI, K. 2011. Soil microbial activity and biomass as influenced by tillage and fertilization in wheat production. *American-Eurasian Journal Agricultural & Environmental Sciences Deira*. 10, 330-337.
- MUELLER, K., HOBBIE, S., CHOROVER, J., REICH, P., EISENHAEUER, N., CASTELLANO, M., CHADWICK, O., DOBIES, T., HALE, C., JAGODZIŃSKI, A., KAŁUCKA, I., KIELISZEWSKA-ROKICKA, B., MODRZYŃSKI, J., ROŽEN, A., SKORUPSKI, M., SOBCZYK, L., STASIŃSKA, M., TROCHA, L., WEINER, J., WIERZBICKA, A. & J. OLEKSYN. 2015. Effects of litter traits, soil biota, and soil chemistry on soil carbon stocks at a common garden with 14 tree species. *Biogeochemistry*, 123 (2015), pp. 313-327, [10.1007/s10533-015-0083-6](https://doi.org/10.1007/s10533-015-0083-6)
- NICOLÁS, C., ALMEIDA, J., ELLSTRÖM, M., BAHR, A., BONE, S., ROSENSTOCK, N., BARGAR, J., TUNLID, A., PERSSON, P. & H. WALLANDER. 2017. Chemical changes in organic matter after fungal colonization in a nitrogen fertilized and unfertilized Norway spruce forest. *Plant Soil* (2017) 419:113–126. DOI [10.1007/s11104-017-3324-8](https://doi.org/10.1007/s11104-017-3324-8)
- OJEDA, J.J., ROMERO-GONZALEZ, M.E., POURAN, H.M. & S.A. BANWART. 2008. In situ monitoring of the biofilm formation of *Pseudomonas putida* on hematite using flow-cell ATR-FTIR spectroscopy to investigate the formation of inner-sphere bonds between the bacteria and the mineral. *Mineral. Mag.* 72, 101–106.
- PADARIAN, J., MINASNY, B. & A.B. MCBRATNEY. 2020. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* 6, 35–52. <https://doi.org/10.5194/soil-6-35-2020>

- PARIKH, S., GOYNE, K., MARGENOT, A., MUKOME, F. & F. CALDE-RON. 2014. Soil Chemical Insights Provided through Vibrational Spectroscopy. Publications from USDA-ARS / UNL Faculty. Paper 1471.
- PIOTROWSKA-DLUGOSZ, A. AND P. CHARZYNSKI. 2015. The impact of the soil sealing degree on microbial biomass, enzymatic activity, and physicochemical properties in the ekranic technosols of toruń (poland). *Journal of Soils and Sediments*. 15, 47-59.
- POPPIEL, R.R., DEMATTE, J.A.M., ROSIN, N.A., CAMPOS, L.R., TAYEBI, M., BONFATTI, B.R., AYOUBI, S., TAJIK, S., AFSHAR, F.A., JAFARI, A., HAMZEHPUR, N., TAGHIZADEH- MEHRJARDI, R., OSTOVARI, Y., ASGARI, N., NAIMI, S., NABIOLLAHI, K., FATHIZAD, H., ZERAATPISHEH, M., JAVAHERI, F., DOUSTAKY, M., NADERI, M., DEHGHANI, S., ATASH, S., FARSHADIRAD, A., MIRZAEI, S., SHAHRIARI, A., GHORBANI, M., RAHMATI, M., 2021. High resolution middle eastern soil attributes mapping via open data and cloud computing. *Geoderma* 385, 114890. <https://doi.org/10.1016/j.geoderma.2020.114890>.
- RASCHE, F., MARHAN, S., BERNER, D., KEIL, D., KANDELER, E. & G. CADISCH. 2013. midDRIFTS-based partial least square regression analysis allows predicting microbial biomass, enzyme activities and 16S rRNA gene abundance in soils of temperate grasslands. *Soil Biology & Biochemistry* 57 (2013) 504-512
- RASCHE, F., KNAPP, D., KAISER, C., KORANDA, M., KITZLER, B., ZECHMEISTER-BOLTENSTERN, S., RICHTER, A. AND A. SESSITSCH. 2011. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *The ISME Journal* 5, 389-402.
- RASSE, D., RUMPEL, C. AND M.F. DIGNAC. 2005. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant and Soil*, 269 (2005), pp. 341-356, [10.1007/s11104-004-0907-y](https://doi.org/10.1007/s11104-004-0907-y)
- REDA, R., SAFFAJ, T., EDDINE, S., BOUZIDA, I., SAIDI, O., YAAKOUBI, B., LAKSSIR, B., MERNISSI, N., MESTAFI, E. AND E. HADRAMI. 2020. Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 242 (2020) 118736
- RIZZO, R., MEDEIROS, L.G., MELLO, D.C. DE, MARQUES, K.P.P., MENDES, W. DE S., QUIÑONEZ SILVERO, N.E., DOTTO, A.C., BONFATTI, B.R., DEMATTÊ, J.A.M. 2020. Multi temporal bare surface image associated with transfer functions to support soil classification and mapping in southeastern Brazil. *Geoderma* 361, 114018. <https://doi.org/10.1016/J.GEODERMA.2019.114018>
- RODRIGUEZ-ALBARRACIN, H.S., DEMATTE, J.A.M., ROSIN, N.A., AMORIM, M.T.A., DARGHAN, A.E., ANDREOTE, F.D. AND J. T. ROSAS. 2024. Soil organic carbon sequestration potential explained by mineralogy and microbiological activity by spectral transfer functions.

- ROSAS, J.T.F., DEMATTE, J.A.M., ROSIN, N.A., POPPIEL, R.R., SILVERO, N.E.Q., AMORIM, M.T.A., RODRIGUEZ-ALBARRACIN, H.S. 2024. The Brazilian soils clay fraction oxides and weathering index spatialized by Earth Observation Strategy
- ROSS, J.L.S., 2013. Brazilian Relief: Structures and Forms. *Rev. do Dep. Geogr.* –USP 25, 20–36. <https://doi.org/10.7154/RDG.2013.0025.0002>.
- ROSIN, N.A., DEMATTÊ, J.A.M., POPPIEL, R.R., SILVERO, N.E.Q., RODRIGUEZ ALBARRACIN, H.S., ROSAS, J.T.F., GRESCHUK, L.T., BELLINASSO, H., MINASNY, B., GOMEZ, C., MARQUES JÚNIOR, J., FERNANDES, K. 2023. Mapping Brazilian soil mineralogy using proximal and remote sensing data. *Geoderma* 432, 116413. <https://doi.org/10.1016/J.GEODERMA.2023.116413>
- SAFANELLI, J.L., POPPIEL, R.R., RUIZ, L.F.C., BONFATTI, B.R., MELLO, F.A.O., RIZZO, R. & J.A.M. DEMATTÊ. 2020. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. *ISPRS Int. J. Geo-Inf.* 2020, 9, 400. DOI: <https://doi.org/10.3390/ijgi9060400>
- SAFANELLI, J.L., DEMATTÊ, J.A.M., DOS SANTOS, N.V., ROSAS, J.T.F., SILVERO, N.E.Q., BONFATTI, B.R. AND W. DE S. MENDES. 2021b. Fine scale soil mapping with Earth Observation data: a multiple geographic level comparison. *Rev Bras Cienc Solo* 45, e0210080. <https://doi.org/10.36783/18069657RBCS20210080>
- SARKHEIL, H., TAHERY, B., RAYEGANI, B., RAMEZANI, J., GOSHTASB, H. & A. JAHANI. 2020. Evaluating the current status of the national health, safety, and environment management system for integration, harmonization, and standardization of environmental protection. *Health Risk Analysis* 18–24.
- SARKHEIL, H., NOUGHABI, K., AZIMI, Y. & S. RAHBARI. 2023. Fuzzy soil quality index using resistivity and induced polarization for contamination assessment in a lead and zinc drainage irrigation field study. *Ecological Indicators* 152 (2023) 110362
- SHARMA, V., CHAUHAN, R. & R. KUMAR. 2021. Spectral characteristics of organic soil matter: A comprehensive review. *Microchemical Journal* 171 (2021) 106836. <https://doi.org/10.1016/j.microc.2021.106836>
- SHI, Z., ALLISON, S. D., HE, Y., LEVINE, P. A., HOYT, A. M., BEEM-MILLER, J., ZHU, Q., WIEDER, W. R., TRUMBORE, S., AND RANDERSON, J. T. 2020. The age distribution of global soil carbon inferred from radiocarbon measurements, *Nature Geoscience*, <https://doi.org/10.1038/s41561-020-0596-z>, 2020.
- SCHINDLBACHER, a., DE GONZALO, C., DÍAZ-PINÉS, E., GORRÍA, P., MATTHEWS, B., INCLÁN, R., ZECHMEISTER-BOLTENSTERN, S., RUBIO, AND R. JANDL. 2010. Temperature sensitivity of forest soil organic matter decomposition along two elevation gradients *J Geophys Res* (2010), 10.1029/2009JG001191
- SILVERO, N.E.Q., DEMATTE, J.A.M., AMORIM, M.T.A., SANTOS, N.V. DOS, RIZZO, R., SAFANELLI, J.L., POPPIEL, R.R., MENDES, W. DE S., BONFATTI, B.R., 2021. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. *Remote Sens. Environ.* 252, 112117. [10.1016/j.rse.2020.112117](https://doi.org/10.1016/j.rse.2020.112117).

- SILVERO, N.E.Q., DEMATTE, J.A.M., MINASNY, B., ROSIN, N.A., NASCIMENTO, J.G., RODRIGUEZ-ALBARRACIN, H.S., BELLINASSO, H. AND A.M. GÓMEZ. 2023. Chapter Three - Sensing technologies for characterizing and monitoring soil functions: A review. Editor(s): Donald L. Sparks. *Advances in Agronomy*, Academic Press, Volume 177, 2023, Pages 125-168, ISSN 0065-2113, ISBN 9780443192586, <https://doi.org/10.1016/bs.agron.2022.08.002>.
- SOTHE, C., GONSAMO, A., ARABIAN, J. & J. SNIDER. 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. *Geoderma* Volume 405, 1 January 2022, 115402
- SOUZA, A.B., DEMATTE, J.A.M., MELLO, F.A.O., SALAZAR, D.F.U., MENDES, W.S. & J.L. SAFANELLI. 2020. Ratio of clay spectroscopic indices and its approach on soil morphometry. *Geoderma* 357, 113963.
- STENBERG, B., VISCARRA ROSSEL, R.A., MOUAZEN, A.M. & J. WETTERLIND. 2010. Chapter five— visible and near infrared spectroscopy in soil science. In: Donald, L.S. (Ed.), *Advances in Agronomy*, vol. 107. Academic Press, pp. 163–215.
- TABATABAI, M.A. 1994. Soil enzymes. In: WEAVER, R.W.; SCOTT, A. & BOTTOMELEY, P.J., eds. *Methods of soil analysis: microbiological and biochemical properties*. Madison, Soil Sci. Soc. Am., 2:778-835, 1994. (Special Publication, 5)
- TERRA, F.S., DEMATTE, J.A.M. AND R.A. VISCARRA ROSSEL. 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>.
- VAN DEN HOOGEN, J., ROBMANN, N., ROUTH, D., LAUBER, T., VAN TIEL, N., DANYLO, O. AND T. CROWTHER. 2021. A geospatial mapping pipeline for ecologists. bioRxiv, p. 2021.07.07.451145, 1 jan. 2021. <https://doi.org/10.1101/2021.07.07.451145>
- VISCARRA ROSSEL, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *J. Geophys. Res.* 116, F04023. <https://doi.org/10.1029/2011JF001977>.
- VISCARRA-ROSSEL, R. AND T. BEHRENS. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158 (2010) 46–54. doi:10.1016/j.geoderma.2009.12.025
- VISCARRA ROSSEL, R.A. AND W.S. HICKS. 2015. Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. *European Journal of Soil Science* 66, 438–450. <https://doi.org/10.1111/ejss.12237>.
- VISCARRA-ROSSEL, R.A., YANG, Y., BISSETT, A., BEHRENS, T., DIXON, K., NEVIL, P. & S. LI. 2022. Environmental controls of soil fungal abundance and diversity in Australia's diverse ecosystems. *Soil Biology and Biochemistry* 170 (2022) 108694. <https://doi.org/10.1016/j.soilbio.2022.108694>

- VISCARRA ROSSEL, R. A., BEHRENS, T., BEN-DOR, E., BROWN, D., DEMATTÊ, J., SHEPHERD, K., SHI, Z., STENBERG, B., STEVENS, A., ADAMCHUK, V., AICHI, H., BARTHES, B., BARTHOLOMEUS, H., BAYER, A., BERNOUX, M., BOTTCHEK, K., BRODSKY, L., DU, C., CHAPPELL, A., FOUAD, Y., GENOT, V., GOMEZ, C., GRUNWALD, S., GUBLER, A., GUERRERO, C., HEDLEY, C., KNADEL, M., MORRAS, H., NOCITA, M., RAMÍREZ LÓPEZ, L., ROUDIER, P., CAMPOS, E., SANBORN, P., SELITTO, V., SUDDUTH, K., RAWLINS, B., WALTER, C., WINOWIECKI, L., HONG, S. AND 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>.
- WADOUX, A.-M.-J.-C., MINASNY, B. AND A.B. MCBRATNEY. 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Rev.* 210, 103359 <https://doi.org/10.1016/j.earscirev.2020.103359>.
- WANG, Y. AND Y. LI. 2023. Mapping the ratoon rice suitability region in China using random forest and recursive feature elimination modeling. *Field Crops Research* 301 (2023) 109016.
- WEI, L., LI, Y., ZHU, Z., WANG, F., LIU, X., ZHANG, W., XIAO, M., LI, G., DING, J., CHEN, J., KUZYAKOV, Y. & T. GE. 2022. Soil health evaluation approaches along a reclamation consequence in Hangzhou Bay, China. *Agriculture, Ecosystems and Environment* 337 (2022) 108045
- WEIL, R. AND N.C. BRADY. 2016. *The Nature and Properties of Soils* (15th edition), Pearson Education (2016)
- XIAO-CHANG, W. AND L. QIN. 2006. Beta-glucosidase activity in paddy soils of the taihu lake region, china. *Pedosphere*. 16, 118-124.
- YANG, J., ZHANG, X., BOURG, A. AND H. STONE. 2021. 4D imaging reveals mechanisms of clay-carbon protection and release. *Nat Commun* 12, 622 (2021). <https://doi.org/10.1038/s41467-020-20798-6>
- YANG, Y., SHEN, Z., BISSETT, A. & R. VISCARRA-ROSSEL. 2022. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. *SOIL*, 8, 223–235, 2022. <https://doi.org/10.5194/soil-8-223-2022>
- YANG, Y.J., DUNGAN, R.S., IBEKWE, A.M., VALENZUELA-SOLANO, C., CROHN, D.M. AND D.E. CROWLEY. 2003. Effect of organic mulches on soil bacterial communities one year after application. *Biology and Fertility of Soils* 38: 273–281.
- ZERAATPISHEH, M., JAFARI, A., BODAGHABADI, M.B., AYOUBI, S., TAGHIZADEH-MEHRJARDI, R., TOOMANIAN, N., KERRY, R. & M. XU. 2020. Conventional and digital soil mapping in Iran: past, present, and future *Catena*, 188 (2020), p. 104424

- ZERAATPISHE, M. & F. KHORMALI. 2012. Carbon stock and mineral factors controlling soil organic carbon in a climatic gradient, Golestan province. *Journal of soil science and plant nutrition*, 12(4), 637-654.
- ZHANG, L., CHEN, X., XU, Y., JIN, M., YE, X., GAO, H., CHU, W., MAO, J. & M. THOMPSON. 2020. Soil labile organic carbon fractions and soil enzyme activities after 10 years of continuous fertilization and wheat residue incorporation. *Scientific Reports* (2020) 10:11318. <https://doi.org/10.1038/s41598-020-68163-3>.
- ZHANG, X., CHEN, S., XUE, J., WANG, N., XIAO, Y., CHEN, Q., HONG, Y., ZHOU, Y., TENG, H., HU, B., ZHUO, Z., JI, W., HUANG, Y., GOU, Y., RICHER-DE-FORGES, A., ARROUAYS, D. & Z. SHI. 2023. Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping. *Geoderma* 432 (2023) 116383
- ZHANG, L. AND L. WANG. 2023. Optimization of site investigation program for reliability assessment of undrained slope using Spearman rank correlation coefficient. *Computers and Geotechnics* 155 (2023) 105208.
- ZORNOZA, R., GUERRERO, C., MATAIX-SOLERA, J., SCOW, K., ARCENEGUI, V. & J. MATAIX-BENEYTO. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem.* 2008 July; 40(7): 1923–1930. doi:10.1016/j.soilbio.2008.04.003.