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

Potential productive of Brazilian agricultural soils

Lucas Tadeu Greschuk

Dissertation presented to obtain the degree of Master in Science. Area: Soil and Plant Nutrition

Piracicaba
2022

Lucas Tadeu Greschuk
Agronomist

Potential productive of Brazilian agricultural soils

versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

Dissertation presented to obtain the degree of Master in Science. Area: Soil and Plant Nutrition

Piracicaba
2022

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

Greschuk, Lucas Tadeu

Potential productive of Brazilian agricultural soils / Lucas Tadeu Greschuk - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2022.

53 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Atributos do solo 2. Mapeamento digital do solo 3. Aprendizado de máquina 4. Monitoramento do solo I. Título

DEDICATÓRIA

Aos meus pais,

ESTANISLAU E LURDES, que

ME APOIARAM INCONDICIONALMENTE,

DEDICO!

AGRADECIMENTOS

Primeiramente agradeço à Deus, pela minha vida e pelas bênçãos recebidas ao decorrer desta caminhada.

Aos meus pais, Estanislau Greschuk e Maria de Lurdes Ruthes Greschuk, pelos ensinamentos, pela educação, pelas orações e pelos momentos de apoio para ajudar-me a alcançar meus objetivos.

Ao meu orientador e amigo Prof. Dr. José Alexandre M. Demattê, pelo incentivo, cobranças e ensinamentos durante a jornada no curso de mestrado.

A todos os professores das disciplinas que cursei no mestrado na Escola Superior de Agricultura "Luiz de Queiroz" – ESALQ/USP pelos ensinamentos e auxílios.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelos recursos investidos em minha formação.

Aos meus colegas e amigos do Grupo de Pesquisa de Geotecnologias em Ciência do Solo (GEOCIS), pela ajuda do desenvolvimento deste trabalho.

A todos que diretamente contribuíram de alguma forma para a realização deste trabalho.

OBRIGADO!

“Pra quem tem pensamento forte,
o impossível é só questão de opinião”.

Charlie Brown Jr.

SUMÁRIO

RESUMO

Potencial produtivo dos solos agrícolas brasileiros

A principal questão para o bem mundial é a segurança alimentar. O Brasil desempenha um papel importante na agricultura global, podendo auxiliar o suprimento da demanda alimentar global no futuro. As propriedades químicas, físicas e biológicas do solo influenciam diretamente a produção de plantas cultivadas. Compreender o potencial produtivo das terras agrícolas brasileiras é de grande importância. Portanto, o objetivo desse estudo é desenvolver uma estratégia para mapear e quantificar o potencial produtivo dos solos agrícolas brasileiros através de técnicas de mapeamento digital de solos. Foram utilizadas aproximadamente 70.000 amostras de solo (0 - 1,0m) georreferenciadas com informações de propriedades químicas, físicas e biológicas de solos agrícolas brasileiros. Cada atributo do solo foi avaliado por meio de informações literárias. Em seguida foi desenvolvida uma equação de artifício ponderado através da análise de componentes principais das amostras de solo. Portanto, cada amostra de recebeu uma nota variado de 0 a 100 referente ao seu potencial do solo em produzir biomassa vegetal, ou seja, quanto maior é a nota, consequentemente maior é o seu potencial. As pontuações das amostras foram preditas via aprendizado de máquina. Foram mapeados 80% (205 milhões de hectares) das áreas agrícolas brasileiras, considerando áreas de plantas cultivadas e pastagens. Através da quantificação territorial do potencial produtivo dos solos, a possibilidade de enxergar as melhores e piores categorias de solos agrícolas em cada bioma brasileiro. Os melhores solos agrícolas do Brasil que pertencem às categorias de SoilPP "muito alto e alto", foram encontrados nos biomas Mata Atlântica (11,8 Mha), Amazônia (7,6 Mha) e Cerrado (4,4 Mha). Por outro lado, os solos que variaram de potencial médio/alto a muito baixo foram encontrados nos biomas Cerrado (77,9 Mha), Caatinga (28,9 Mha) e Mata Atlântica (34,5 Mha). Através da avaliação das terras agrícolas e de valores médios de produtividade, observamos que de 2304 municípios produtores de soja, 896 possuem capacidade para aumentar a produtividade média da soja até um determinado nível. Para a cana de açúcar, de 2468 municípios avaliados, 1056 podem elevar sua produtividade média. Portanto, essa técnica pode auxiliar no desenvolvimento de políticas globais de segurança alimentar e do solo e ser replicada em outros países, regiões, municípios ou fazendas. O SoilPP*map* para as áreas agrícolas brasileiras teve por principais limitações a representatividade espacial para todos os tipos de solos agrícolas do Brasil e também a baixa acurácia do modelo de predição, possivelmente relacionada a avaliação de atributos químicos do solo. Entretanto, a estratégia utilizada foi efetiva para mapear e quantificar o potencial produtivo dos solos agrícolas brasileiros.

Palavras-chave: Atributos do solo, Mapeamento digital de solos, Aprendizado de máquinas, Monitoramento do solo

ABSTRACT

Potential productive of Brazilian agricultural soils

The main issue for the world's good is food security. Brazil plays an important role in global agriculture and can help meet global food demand in the future. The chemical, physical and biological properties of the soil directly influence the production of cultivated plants. Understanding the productive potential of Brazilian agricultural land is of great importance. Therefore, the objective of this study is to develop a strategy to map and quantify the productive potential of Brazilian agricultural soils through digital soil mapping techniques. Approximately 70,000 soil samples (0 - 1.0m) georeferenced with information on chemical, physical and biological properties of Brazilian agricultural soils were used. Each soil attribute was evaluated using literary information. Then, a weighted artifice equation was developed through the analysis of principal components of the soil samples. Therefore, each sample received a score ranging from 0 to 100 referring to its soil potential to produce plant biomass, that is, the higher the score, the higher its potential. Sample scores were spatialized via machine learning. 80% (205 million hectares) of Brazilian agricultural areas were mapped, considering areas of cultivated plants and pastures. Through the territorial quantification of the productive potential of soils, the possibility of seeing the best and worst categories of agricultural soils in each Brazilian biome. The best agricultural soils in Brazil that belong to the "very high and high" SoilPP categories were found in the Atlantic Forest (11.8 Mha), Amazon (7.6 Mha) and Cerrado (4.4 Mha) biomes. On the other hand, soils that varied from medium/high to very low potential were found in the Cerrado (77.9 Mha), Caatinga (28.9 Mha) and Atlantic Forest (34.5 Mha) biomes. Through the evaluation of agricultural land and average productivity values, we observed that of 2304 soybean producing municipalities, 896 have the capacity to increase average soybean productivity to a certain level. For sugarcane, out of 2468 municipalities evaluated, 1056 can increase their average productivity. Therefore, this technique can assist in the development of global food and soil security policies and be replicated in other countries, regions, municipalities or farms. The SoilPPmap for Brazilian agricultural areas had as main limitations the spatial representativeness for all types of agricultural soils in Brazil and also the low accuracy of the prediction model, possibly related to the evaluation of soil chemical attributes. However, the strategy used was effective in mapping and quantifying the productive potential of Brazilian agricultural soils.

Keywords: Soil attributes, Digital soil mapping, Machine learning, Soil monitoring

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

SoilPP	Soil Productive Potential
SoilPPI	Soil Productive Potential Index
SoilPP*map*	Map of soil productive potential
SoilPP*score*	Soil productive potential score evaluated through equation (5)
PES	Production Environment System
Mha	Million hectares
SySI	Synthetic Soil Image
K	Magnetic susceptibility of the soil
PCA	Principal Component Analysis

1. INTRODUCTION

The continuous increase in demand for food and fuel has put considerable pressure on agricultural areas around the world. Strategies to face these demands have led to improvements in food production by optimizing natural resources that are still available for agricultural production (Ferreira and Férez, 2020; Fróna et al., 2019). Brazil is one of the largest producers and exporters of agricultural products in the world (FAOSTAT, 2022), exploring about 30% of its territory (263 million hectares) with agricultural activities (MapBiomas, 2020). The expansion of soil tilling activities to new Brazilian areas is a unsustainable option to meet future food demands (Bordonal et al., 2018). Thus, new strategies should be developed to improve yields where crops are already cultivated, which will allow optimizing the use of current resources (Gerland et al., 2014; Runyan and Stehm, 2019), especially the soil, which is under constant anthropogenic pressure in the last decades (Montanarella et al., 2016). However, there are numerous challenges to proper evaluate and map agricultural areas when considering the large territorial extents within diverse biomes (Gomes et al., 2019; Guerra et al., 2020), climate types (Alvares et al., 2013) and soils (Santos et., 2018), such as the case of Brazil.

Soil provides a wide range of ecosystem services that meet human needs (Adhikari and Hartemink, 2016; Dominati et al., 2014; Kopittke et al., 2019; Robinson et al., 2012; Tóth et al., 2013), including the provision of food, fiber, and fuel (FAO and ITPS, 2015; Johnston and Poulton, 2018). The chemical, physical and biological properties of a soil directly influence crop yields (Beerling et al., 2018; Bishopp and Lynch, 2015; Huang et al., 2021; Vogel et al., 2019), and its productive potential is a result of biomass production capacity (Vogel et al., 2019; Yang et al., 2003). However, estimating the productive potential that for a given soil is a major question that remains in soil science (Greiner et al., 2017; Yang et al., 2003). Therefore, measuring the productive potential of an agricultural environment is an important step to find ways to increase production in a sustainable and efficient way.

We can observe some ways to measure the yield potential of an agricultural soil. Vogel et al. (2019) presented methods used in Germany to assess the potential of soils for wheat biomass production through water availability and soil texture. Huang et al. (2021) evaluated the effect of texture and soil organic carbon (SOC) amounts on yield responses of seven major crops in the United States of America from 1958 to 2019 and found that crop yields were higher in soils with fine texture and higher SOC concentration, due to increased available water, improved soil structure and nutrient retention (Huang et al., 2021). The Production Environment System (PES) was developed after some empirical studies initiated by Coopersucar (1994) and ratified by (Landell et al., 2003), which observed the strong relationship between soil types and sugarcane productivity. The PES can be defined as the junction of one or more soil mapping units with similar production capacity, expressed in Tons of Cane per Hectare (TCH) and classified in seven different classes of sugar cane yield (Demattê and Demattê, 2009). However, this involves high complexity and requires time and resources to map in detail large territories such as the case of Brazil.

Digital Soil Mapping (DSM) is an indispensable tool for environmental studies, making it possible to estimate soil information at multiple scales (Demattê et al., 2020; McBratney et al., 2003). Assessing and quantifying the intrinsic potential of Brazilian agricultural soils for biomass production (Vogel et al., 2019), through scoring methods (Cherubin et al., 2016; Kumar et al., 2019; Lehmann et al, 2020) based on soil chemical, physical and biological indices, associated with digital soil mapping techniques (Demattê et al., 2020, 2018; McBratney et al., 2003; Safanelli et al., 2021, 2020), can be a good strategy to produce information for large geographical extents as Brazil.

These measures can assist in increasing food security for the rapidly growing global population under a changing climate (Huang et al., 2021; McBratney et al., 2014). Obtaining soil and terrain information is necessary for policy formulation ("Computing the Iowa Corn Suitability Rating for Your Farm | Ag Decision Maker," 2013), agricultural resource management, and soil security (Lehmann et al., 2020b; McBratney et al., 2014; Mulder et al., 2011). It can also assist the prediction of provisioning services (direct or indirect food for humans, timber, fiber, and fuel) and supporting services (nutrient cycling and production) (Adhikari and Hartemink, 2016; Dominati et al., 2014; Kopittke et al., 2019; Robinson et al., 2012; Tóth et al., 2013). On another hand, land use requires the formulation of data-driven policy and management strategies, which is hampered by the high cost and time-consuming characteristics of soil data acquisition (McBratney et al., 2014).

Thus, there is a need for the development of methods that can monitor soil over large areas with a good level of detail, assisting agricultural decision making to sustain food security for the population, and helping in the process of public policy making. It is expected that the creation of a weighted index by principal component analysis (PCA) to evaluate the chemical, physical and biological attributes of the soil associated with digital soil mapping techniques can represent the productive potential of agricultural lands in the Brazil, i.e, the best and worst Brazilian agricultural lands. In this sense, the objective of this study is to develop a new method for mapping the productive potential of Brazilian soils at 30m spatial resolution from agricultural areas using remote sensed data coupled with scoring functions and machine learning algorithm.

2. METHODOLOGY

This study has five main steps (Fig. 1). First, a database with approximately 70,000 soil samples was organized with information on the geographic coordinates, chemical, physical and biological properties of Brazilian agricultural soils for layers A (0-20cm), B (40-60cm) and C (80-100cm) (Step 1). Subsequently, soil properties were integrated using scoring functions with values ranging from 0 to 100, that is, the higher the score, the greater the productive potential of the soil (Step 2). Principal component analysis (PCA) made it possible to create a weighted index (SoilPPI) to score all soil samples in the range from 0 to 100 (Step 3) for each layer.

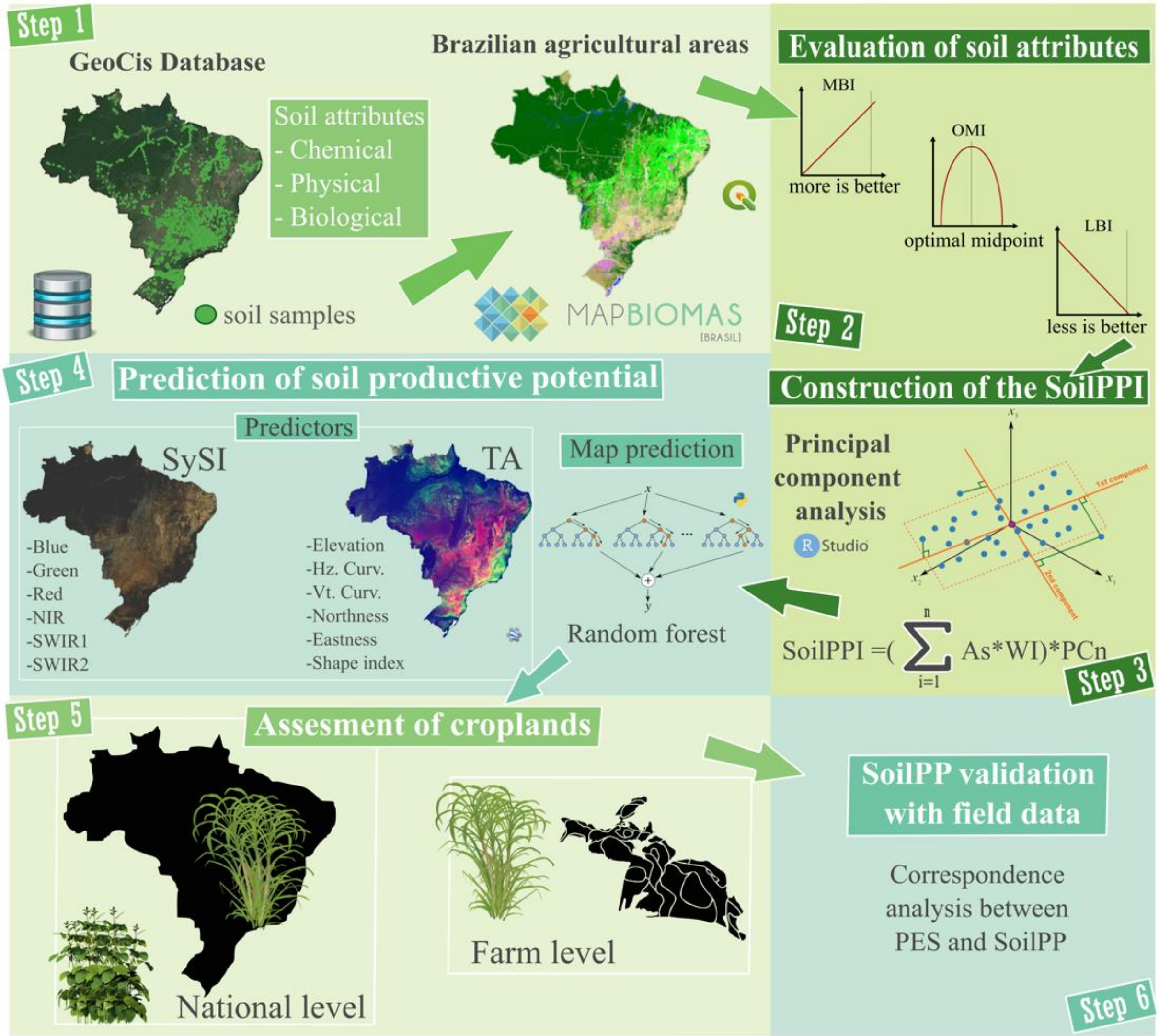

Fig. 1: Methodological flowchart. SySI: Synthetic Soil Image. TA: Terrain Attributes.

Then, we used the scores obtained for each soil sample and covariates to obtain the soil productive potential map (SoilPPmap) using the random forest machine learning algorithm

(Step 4). Finally, an evaluation was made of the crop productivity historical statistics of soybean (*Glycine max*) and sugarcane (*Saccharum officinarum*) in Brazilian municipalities, followed by a case study at the farm level (Step 5). We used the correspondence analysis to understand the categorical similarity between SoilPP and a traditional system adapted by Demattê and Demattê (2009) to assess sugarcane productivity through pedological classification, called production environments system (PES) (Step 6).

2.1. Study area

The study area comprises the entire agricultural areas from Brazil, distributed in about 66.5 million hectares for crops, 68 million hectares for native pasture and 112 million hectare for planted pastures, totaling 246.5 million hectares (Embrapa Territorial, 2018). These data were collected through the Rural Environmental Registry (CAR), and the survey was carried out at the property level throughout the country (EMBRAPA Territorial, 2018). The 6.0 collection provided by MapBiomas was used as a spatial mask for agricultural areas, totaling about 263,045,118 million hectares (MapBiomas, 2020), an increase of 6.28% in agricultural areas compared to 2016.

The Brazilian territorial extension contains an enormous biodiversity, with six biomes and different climatic conditions, mainly the tropical type (Gomes et al., 2019; Guerra et al., 2020). The soils found in Brazil are mostly derived from sedimentary and igneous rocks, such as Latosols (Ferralsols) and Argisols (Acrisols) (Santos et al., 2018). The Amazon biome covers 49.29% of Brazil, with a humid tropical climate (Af, Am and Aw). The Cerrado (Brazilian savannas) is the second largest biome, covering 22% of the territory, with a predominantly semi-humid climate (Aw). The Atlantic Forest biome, which extends to the east of Brazil, has the greatest diversity of environments and there are several types of climate (Cfb, Cfa, Cwb, Aw and As). The Caatinga biome is the driest, under a semi-arid climate (Bsh), with an average annual rainfall of 500 mm and an average annual temperature of 20 to 29 °C. The Pantanal biome is characterized by long periods of flooding, seasonal climate Aw. The pampa biome is located in southern Brazil, covered by temperate grasslands with a Cfa climate. Throughout the territory, Brazil has a large extension of protected natural areas, especially in the Amazon biome (Gomes et al., 2019).

.

2.2. Soil sampling and laboratory analysis

The georeferenced soil data were acquired from soil surveys carried out in agricultural areas by the Geotechnologies in Soil Science group and collaborators (Demattê et al., 2019). These field data were obtained in transects (e.g., along toposequences) or grid schemes using global positioning systems. As a compilation of national soil data, soils were sampled over different periods of time during the last two decades by independent field campaigns. Soil samples (number of soil samples: layer A "22122"; layer B "21160"; layer C "25992") had their physical and chemical soil attributes analyzed for the surface and subsurface layer of soil (A: 0-20; B: 40-60; C: 80-100 cm).

The physical-chemical analysis followed the Brazilian standards of soil analysis (Teixeira et al., 2017), gathering information on clay, silt and sand (g kg $^{-1}$) contents; soil organic carbon (SOC, g kg $^{-1}$); soil organic matter (SOM, g kg $^{-1}$); pH determined in aqueous solution (pH in water); pH determined in KCl; cation exchange capacity (CEC pH7); Ca^{+2} (mmolc kg $^{-1}$); Mg^{+2} (mmolc kg $^{-1}$); K^{+} (mmolc kg $^{-1}$); Al^{+3} (mmolc kg $^{-1}$); H+Al (mmolc kg $^{-1}$) sum of bases (mmolc kg $^{-1}$); base saturation (V%); and Aluminum saturation (m%).

We derived the following soil attributes: slope (º); Soil density "SD" (g cm^{-3}) was estimated using a pedotransfer function obtained by (Benites et al., 2007) using clay and SOC contents for Brazilian soils with R^2 of 0.63 and a standard error of 0.11 g. cm^{-3}. The ΔpH "pH_{KCl} - pH_{H20}", weathering index (Ki) "SiO_2/Al_2O_3 x 1,7"; and clay activity were calculated according to (Prado et al., 2011).

2.3. Soil potential productivity index (SoilPPI)

2.3.1 Scoring soil attributes

For this step, we used methods usually employed for assessing the soil quality (Cherubin et al., 2016; Lehmann et al., 2020). Each soil attribute was scored on a scale ranging from 0 to 100 using three types of functions: 'more is better' index (MBI), which means that higher values of the soil attribute indicate higher SoilPP; 'less is better' index (LBI), with lower values considers the highest SoilPP; and "optimal midpoint" (OMI), where an intermediate value indicates superior soil condition (Cherubin et al., 2016). Clay, Silt, SOC, SOM, Ca^{2+}, Mg^{2+}, K^{+}, CECpH7, SB, V%, Ki and Clay activity were scored using MBI (Equation 1). Sand, BD, Al^{3+}, H^{+}+Al^{3+}, m%, ΔpH and Slope (º) were scored using LBI (Equation 2), while pH in water was scored using OMI (Equation 3).

$$MBI = \frac{a}{\left[1 + \left(\frac{B - UL}{x - UL}\right)^{S}\right]} \tag{1}$$

$$LBI = \frac{a}{\left[1 + \left(\frac{B - LL}{x - LL}\right)^{S}\right]} \tag{2}$$

$$OMI = \begin{cases} \dfrac{a}{\left[100 - \left(\frac{B_L - O}{x - O}\right)^{S}\right]} & for \quad x < O \\ 100 & for \quad x = O \\ \dfrac{a}{\left[100 - \left(\frac{B_U - O}{x - O}\right)^{S}\right]} & for \quad x > O \end{cases} \tag{3}$$

where $MBI_{(1)}$, $LBI_{(2)}$ and $OMI_{(3)}$ are the 'more is better', 'less is better', 'optimal midpoint' scoring functions, respectively; 'a' is the maximum score value (100), 'S' is the slope of the equation, defined as -2.5; 'B' is the baseline value that has a score of 50% (median); 'U_L' is the upper (maximum) limit of soil attribute values; 'L_L' is the lower (minimum) limit of soil attribute values; 'B_L' is the lower baseline of the 'ideal midpoint' curve, with a score of 50%; 'B_U' is the upper baseline of the 'ideal midpoint' curve, with a score of 50%; 'O' is the optimal score value, equal to 100%; and 'x' is the real value of the soil attribute that is being scored (Table 1).

The lower, middle and upper baselines values were determined as the 0.01, 0.5 and 0.99 percentiles of soil data (Table 1). This approach was used to restrict the function parameters to the soil attributes variation obtained through soil analysis. The slope was used as an indirect parameter to assess the depth of the soil, i.e, the greater the slope, the shallower the soil, and the greater the risk of erosion (Santos et al., 2018).

Table 1. Parameters of the scoring functions of soil attributes.

Attribute	Scoring function[1]	LL[2]	B[3]	UL[4]	B_L[5]	O[6]	B_U[7]
Sand (g kg^{-1})	LBI	9.00	747.00	974.00	-	-	-
Silt (g kg^{-1})	MBI	0.00	63.00	846.00	-	-	-
Clay (g kg^{-1})	MBI	5.00	180.00	960.00	-	-	-
Bulk Density (g cm^{-3})	LBI	0.85	1.42	1.55	-	-	-
SOC (g kg^{-1})	MBI	0.01	5.50	59.16	-	-	-
SOM (g kg^{-1})	MBI	1.00	9.40	102.00	-	-	-
pH in water	OMI	3.70	5.50	7.80	5.4	6.5	7.2
Ca^{2+} (mmolckg^{-1})	MBI	0.00	8.01	99.60	-	-	-
Mg^{2+} (mmolckg^{-1})	MBI	0.00	3.70	55.00	-	-	-
K^{+} (mmolckg^{-1})	MBI	0.00	0.70	20.00	-	-	-
Al^{3+} (mmolckg^{-1})	LBI	0.00	1.50	156.50	-	-	-
$H^{+} + Al^{3+}$	LBI	0.10	19.20	336.50	-	-	-
CEC (mmolckg^{-1})	MBI	3.29	35.40	349.50	-	-	-
SB (mmolckg^{-1})	MBI	0.00	13.10	159.80	-	-	-
V %	MBI	0.00	39.60	99.84	-	-	-
m %	LBI	0.00	10.58	100.00	-	-	-
ΔpH	LBI	-3.30	-0.8	2.48	-	-	-
Ki	MBI	0.69	1.62	3.73	-	-	-
Clay activity	MBI	0.72	18.39	1887.65	-	-	-
Slope º	LBI	0.00	3.09	38.36	-	-	-

[1]LBI: "Less is better" index; MBI: "More is better" index; OMI: "optimum mid-point". [2]Lower limit. [3]Baseline. [4]Upper limit. [5]Lower baseline. [6]Optimum [7]Upper baseline. Obs.: The parameters were determined by statistical reductions of the whole Brazilian territory. The minimum and maximum values were determined by the 0.01 and 0.99 percentiles. The baseline was determined by the median value (0.5 percentile). The lower baseline was determined by the middle between the minimum (in parenthesis) and optimum values, while the upper baseline was determined by the optimum and maximum values (in parenthesis).

2.3.2 Indexing the soil attributes into the SoilPPI

The final index (SoilPPI) was obtained by a weighted sum of the soil attributes scored in the 2.3.1 section. Principal component analysis was adopted to determine the weights as described by Cherubin et al. (2016). Due to the different ranges of values and measurement units referring to the soil samples (BDGeoCis), their normalization was performed to rescale the data to have an average of 0 and a standard deviation of 1. This standardization is called z-score. A z score can be calculated with the following formula:

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

where z is represented by the z score, $\boldsymbol{x}$ is the soil attribute value to be normalized, $\boldsymbol{\mu}$ is the mean of the original data, σ is the standard deviation of the original data. Data points above the mean will result in a positive z-score, while data below the mean will get a negative z-score. The absolute value of the score indicates the number of standard deviations between the data points and the mean. Then, principal component analysis (PCA) was performed using the "factoextra", "lifecycle" and "psych" packages in the R environment (R Core Team, 2020).

Through PCA, each soil attribute was weighted according to its absolute loadings and the proportion of variance explained by each component (e.g., % of variation explained by each component divided by the total accumulated variation of all components selected for the study. The weighted values of attributes used in the calculation of the SoilPP are available in the supplementary material (Table A1 and Table A2). The amount of principal components selected to perform the SoilPPI calculation was based on the variance proportion above 80% and standard deviation (SD) value above 1 (Table A2). To calculate the SoilPP the following equation was used:

$$SoilPPI = [(\sum_{i=1}^{n} AS * WI) * PCn] \tag{5}$$

where *'n'* is the number of soil attributes selected, *'AS'* is the attribute scored obtained for each attribute in each soil sample, *'WI'* weighted indicator value within each main component provided through PCA, *'PCn'* is the proportion of variance of the principal component. Finally, all soil samples from each layer were scored on a scale from 0 to 100 using the SoilPPI equation (5). The closer the sample score is to 100, the greater the productive potential of the soil.

2.4 Spatial Prediction of the SoilPPI (Random Forest)

The Synthetic Soil Image (SySI) reflectance bands (Demattê et al., 2018) and terrain attributes in Google Earth Engine (TAGEE) (Safanelli et al., 2020) were used as environmental covariates to calibrate SoilPPmap prediction models for Brazilian agricultural areas. The covariates from SySI are composed of spectral bands: "Blue", "Red", "Green", "Nir", "SWIR 1", and "SWIR 2". The TAGEE attributes used were: Eastness, Northness, Elevation, Shape Index, Vertical Curvature, and Slope. At each soil sampling site, the values of predictor covariates were intersected using the sampling method that considers the value pixels below the point sample with coordinates. A set of bootstrapped and random regression trees from Python's scikit-learn library was used as a machine learning algorithm (Pedregosa et al., 2011).

Instead of fitting simple regression trees, we emulated the Random Forest algorithm (Breiman, 2001) generating bootstrapping trees to aggregate them by the mean, using the methodology of Safanelli et al., (2021). The optimal number of bootstrap trees (forest size), number of covariates to be sampled in tree splits and terminal leaf size was defined with a hyperparameter grid search seeking to control overfitting during calibration. The range of values tested for forest size was 30, 60, 100, 200, 300 and 500 trees. The amount of 1, 2, 3, 5, 8, 11 and 13 predictors were investigated to be used randomly in tree divisions. For the minimum number of observations on the leaves, which defined the size or individual depth of the tree, values of 10, 20, 30, 40, 50, 100, 200 and 500 observations were tested.

The models were adjusted for each depth (H:0-20; B:40-60 and C:80-100cm). The optimal model for each depth of soil was determined by the Root Mean Square Error (RMSE) of the calibration set after testing all combinations of hyperparameters. The calibration set consisted of bootstrap observations, while the remaining observations that were not sampled were used for testing, resulting in an out-of-bootstrapped validation. The accuracy of the predictive model was assessed using the RMSE. The Coefficiene of Determination (R^2) was calculated to assess the explained variance of the prediction models, and the Performance Ratio to Interquartile Range (RPIQ) was calculated to assess the consistency between the predicted values with the variability of the soil dataset. Final assessment metrics were calculated by averaging the statistics from the improvised samples (test observations), reporting the mean and standard deviation.

After the SoilPPmap prediction for the soil layers (A:0-20; B:40-60 and C:80-100cm), the average of the three SoilPPmap layers was calculated as an aggregation alternative to represent the SoilPP up to one-meter depth. Areas with climate types Aw (Cerrado) and Bsh (Caatinga) were penalized due to the higher vapor pressure gradient, which causes an increase in the potential evapotranspiration of the plant (Alvares et al., 2013; Nyle Brady and Ray R. Weil, 2012).

2.5 Assessment of croplands

2.5.1 National extent

The predicted SoilPPmap were intersected with mean crop yield data of soybean and sugarcane from the Brazilian municipalities between 2016 and 2020. SoilPPmap and the municipal average yield data for the respective crops were used to perform this assessment on a national scale. The municipality yield data were obtained as vector files from the Brazilian Institute of Geography and Statistics database (IBGE, https://sidra.ibge.gov.br/tabela/6957), considering the average productivity for a five-year period, corresponding to the 2016 to 2020 harvests. The municipalities that had average production above 35 tons ha^{-1} of sugarcane and 1500 kg ha^{-1} of soy were considered in the analysis. The average productivity values of each municipality for soybean and sugarcane were categorized according to table 2. After the elaboration of the SoilPPmap, the average SoilPP of the agricultural areas that make up all Brazilian municipalities was performed, ie, the result is an average value of SoilPP for each municipality, which was categorized using the Class score 'Cs' (Table 2). The values of Brazilian municipalities with average yield for soybean and sugarcane were divided into seven different classes, called Class yield 'Cy'. The variables Cs and Cy are available in table 2, and are fundamental to perform the yield level calculation, i.e., each Brazilian municipality producing soybeans and sugarcane falls into some category of Cy and Cs.

Table 2. Categorization of municipal productivity values for sugarcane, soybean (IBGE 2022) and SoilPPmap in Class yield and Class score.

Soybean	Sugar Cane		SoilPPI	
Yield (kg ha^{-1})	Yield (ton ha^{-1})	Class yield (Cy)	SoilPP score	Class score (Cs)
3517.0 - 5440.0	79.7 - 130.0	1	85.5 - 87.7	1
3339.0 - 3517.0	69.8 - 79.7	2	83.1 - 85.4	2
3194.0 - 3338.0	58.3 - 69.8	3	80.8 - 83.0	3
3034.0 - 3194.0	52.0 - 58.3	4	78.4 - 80.7	4
2823.0 - 3034.0	45.6 - 52.0	5	76.1 - 78.3	5
2496.0 - 2823.0	40.0 - 45.6	6	73.7 - 76.0	6
1500.0 - 2496.0	35.0 - 40.0	7	71.3 - 73.6	7

* The Class score (Cs) and Class yield (Cy) variables are used in equation 6.

The verification of the productive levels of the Brazilian municipalities allows to visualize at which levels the SoilPP is being explored. To perform this analysis, equation 6 was used:

$$Yield\ level = \mathrm{Cs} - \mathrm{Cy} \quad (6)$$

If the resulting value of the "Yield level" is positive for a Brazilian municipality, it means that: the agricultural areas that cultivate soy or sugar cane located within its territory are exploiting the productive potential of the soil at optimal levels, and the greater the value, the greater its exploitation. However, if the resulting value is negative, municipalities can increase the productivity value of the evaluated crops.

2.5.2 Farm extent

A case study was carried out on a sugar cane farm, located in the municipality of Rafard (State of São Paulo). The farm's area is about 182 ha of the sugar cane. We used this example as a basis for validation and comparison of the Production Environment map (Demattê and Demattê, 2009) prepared through the pedological classification of the area, with information on texture, soil fertility, and productivity. The production environment system (PES) has seven different categories based on sugarcane productivity (A: greater than 100 tons per hectare; B: 91 to 100 tons; C: 86 to 90 tons; D: 81-85 tons, E: 76-80 tons; 70-75 tons and G: < 70 tons per hectare). SoilPPmap relationships with the geology and soil classification of the area were also analyzed through legacy field maps, enabling a better understanding of the characteristics of a given environment. The SoilPPmap (.tif file) of the Brazilian agricultural areas was cut according to the perimeter of the Rafard farm (SP), presenting a spatial resolution of 30 m and soil information up to a depth of one meter. The PES was generated from the physical, chemical and biological properties of the soil, plant evapotranspiration and productivity, which were evaluated in the field.

2.6 Correspondence analysis

Multiple Correspondence Analysis (MCA) was performed with about 21000 soil samples evaluated in the field via PES and used in SoilPPmap prediction, to verify the association of categorical groups between both systems. The variables analyzed were the SoilPP*score* compared to Production Environment System (PES), which is categorized through pedological classification, with information on texture, fertility and evapotranspiration of the plant (Demattê and Demattê, 2009). The PES is divided into seven different categories (A, B, C, D, E, F and G) based on sugarcane yield, as described in the previous item. With this, it is possible to analyze whether the SoilPPI categorical system (equation 5) corresponds with the empirical categorical system performed through the methodology of Demattê and Demattê, (2009). The Chi-square test was applied to assess the significance of the association between the categories of both methods.

3. RESULTS

3.1 Prediction of soil productive potential (SoilPP)

The performance of the SoilPPmap prediction model for layers A (0-20 cm), B (40-60 cm), and C (80-100 cm) using the covariates extracted through soil samples with geographic coordinates was observed (Table 3). Subsequently this was employed in an algorithm of machine learning (Random forest). The performance of the spatial prediction of SoilPP for the three layers has values of R^2 smaller than 0.25, RMSE values ranging from 5.53 to 5.80 and RPIQ values higher than 1.41. Layer A presented the lowest R^2 (0.20), demonstrating the largest size of the forest (FS: 100), where the prediction was calculated using the most unstable hyperparameters of the model, such as the high number of observations on the end sheets (MSL: 500). Layers B and C presented R^2 values of 0.23 and 0.25, respectively. The predictions were calculated from 200 samples in the terminal leaves for both layers B and C. The difference between these layers was in the size of the forest, where layer C had an FS equal to 60, and layer B had an FS equal to 30. The number of hyperparameters of random features tested in each split of the tree was the same for all layers (NRF: 13).

Table 3. Prediction performance metrics and optimal hyperparameters of ensembled bootstrapped regression trees used to map the potential productive of the agricultural soils in Brazil.

Layers[1]	FS[2]	MSL[3]	NRF[4]	[5] R^2	RMSE[6]	RPIQ[7]
A	100	500	13	0.20	5.53	1.41
B	30	200	13	0.23	5.80	1.90
C	60	200	13	0.25	5.57	1.93

[1]A: soil layer 0 - 20 cm of deep, B: soil layer 40 - 60 cm of deep, C: soil layer 80 - 100 cm of deep; [2]FS: forest size, i.e., number of bootstrapped trees; [3]NRF: hyper- parameter number of random features tested in each tree split; [4]MSL: hyper-parameter minimum samples at leaves; [5]R^2: coefficient of determination; [6]RMSE: root mean squared error; [7]RPIQ: ratio of performance to interquartile range.

3.2 SoilPP map for Brazilian agricultural areas

The map of the productive potential of the soil (SoilPPmap) for Brazilian agricultural areas (Fig. 2) through the agricultural area filter (MapBiomas, 2020), represents about 205 million hectares (Mha). The map was generated by evaluating approximately 70,000 soil samples through indexing strategies (equations 1, 2, 3, 4 and 5) associated with DSM techniques in conjunction with the "random forest" machine learning algorithm (Breiman, 2001). The SoilPPmap represents information up to 1m deep. The soil productive potential score

(SoilPPscore) reflects its ability to produce biomass and is divided into seven different classes (A, B, C, D, E, F, and G), and each one is represented by a specific color (Fig. 2). The best agricultural areas (pixels) with the highest SoilPP are those that obtained the highest scores (A, B and C) through the soil productive potential index (SoilPPI 'equation 5'), while the worst areas (E, F and G) are those that obtained the lowest scores.

Soils belonging to class 'A' (SoilPP: very high) (Fig. 2), feature greater depth, good drainage, fine texture, rich in nutrients for plants, the relief can be variable and without the presence of rocky outcrops. On the other hand, the main characteristics of soils that belong to class 'G' (SoilPP: very low) are low capacity to provide nutrients to the plants, usually have a coarse texture (sandy), low water retention, and consequently low availability for the plant. This class also can present varied soil depth and, in some cases, have a high presence of stones. Intermediate categories as 'B and C' have a greater similarity of characteristics with class 'A', while 'E and F' categories are closer to class 'G'.

The Brazilian agricultural soils that presented a very high SoilPP (class A) represented by the green color (Fig. 2), comprised about 11.1 million hectares (M ha) or 4.8% of the total agricultural areas. Areas with high SoilPP (class B) covered 14.5 M ha (5.5% of the total), the medium high SoilPP (class C) corresponded to 44.5 M ha (16.9% of the total). These are the soil classes that have the greatest productive potential, corresponding to about 27.2% of the total agricultural area. Agricultural soils with medium SoilPP, represented by the color orange, was the class that had the largest amount of occupied area, approximately 50 M ha (19% of the total area). Then, soils with medium/low, low and very low productive potential covered 44.1 M ha (16.7% of the total agricultural area), 20.9 M ha (7.9%) and 18.6 M ha (7.0%) respectively.

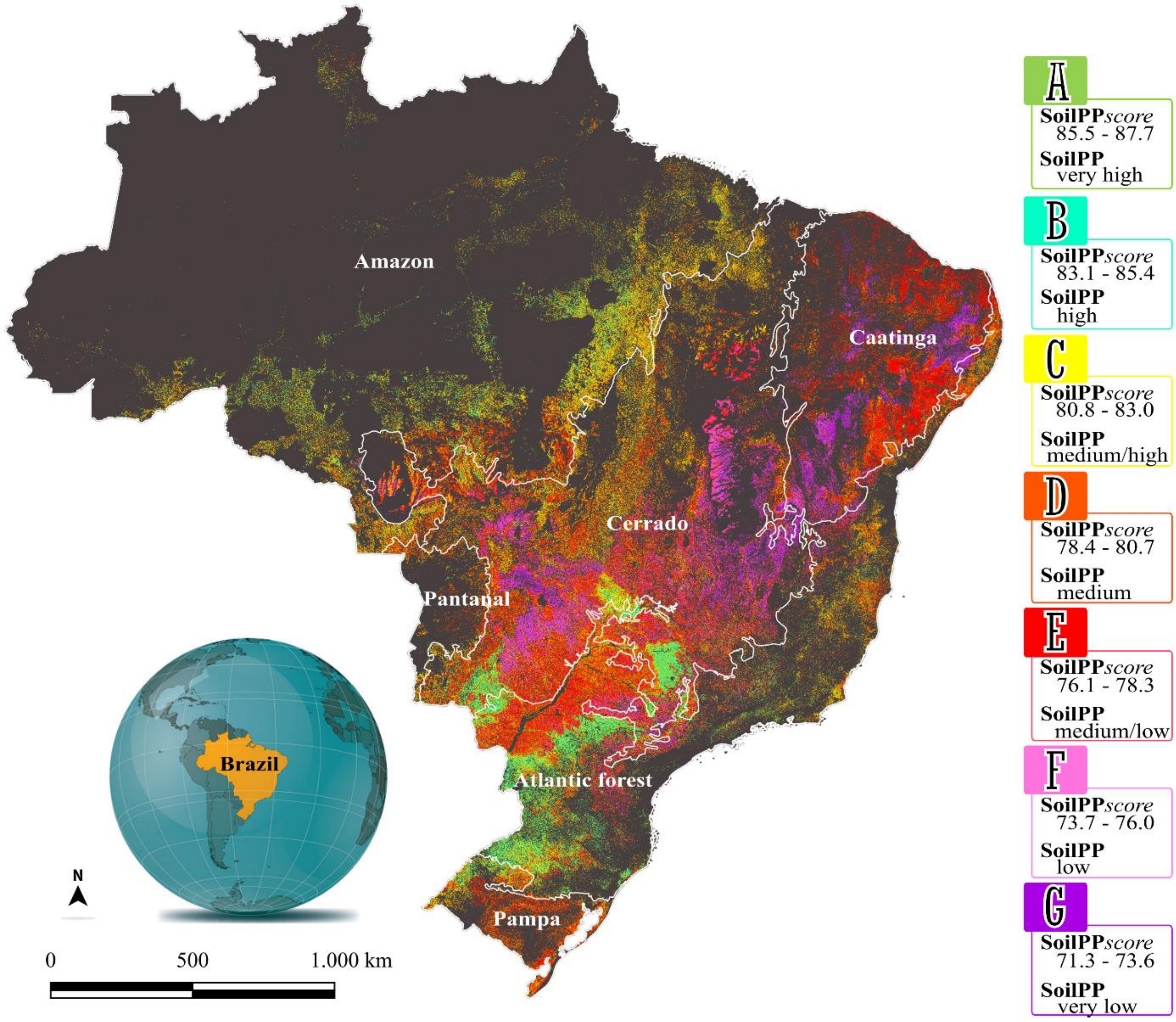

Fig. 2: The Soil Productive Potential map for Brazilian agricultural areas; SoilPP*score*: Soil Productive Potential score (pixel value); SoilPP: Soil Productive Potential.

Fig 3 shows the distribution of the areas of SoilPP levels across the different Brazilian biomes. Most agricultural soils located in the Amazon biome had SoilPP ranging from very high to medium (Fig 3a). Through the use of the GEOS3 algorithm (Demattê et al., 2018) and the agricultural mask (MapBiomas, 2022), about 35.5 Mha are destined for agricultural use in this biome. The very high, high, medium/high, and medium SoilPP corresponds to an area of 3.2, 4.4, 15.5, and 8.8 Mha respectively. The smaller part of the agricultural areas of the Amazon (9.8%) had a low SoilPP. The Cerrado biome is the one with the largest territorial use for agricultural activities in Brazil, occupying around 81.9 Mha (Fig 3b). Most agricultural soils located in the Cerrado showed a medium/high to medium/low SoilPP with an area of 52.8 Mha (64.5% of the territory). The SoilPP considered low and very low represent an area of 24.7 Mha and high and very high with an area of 4.4 Mha.

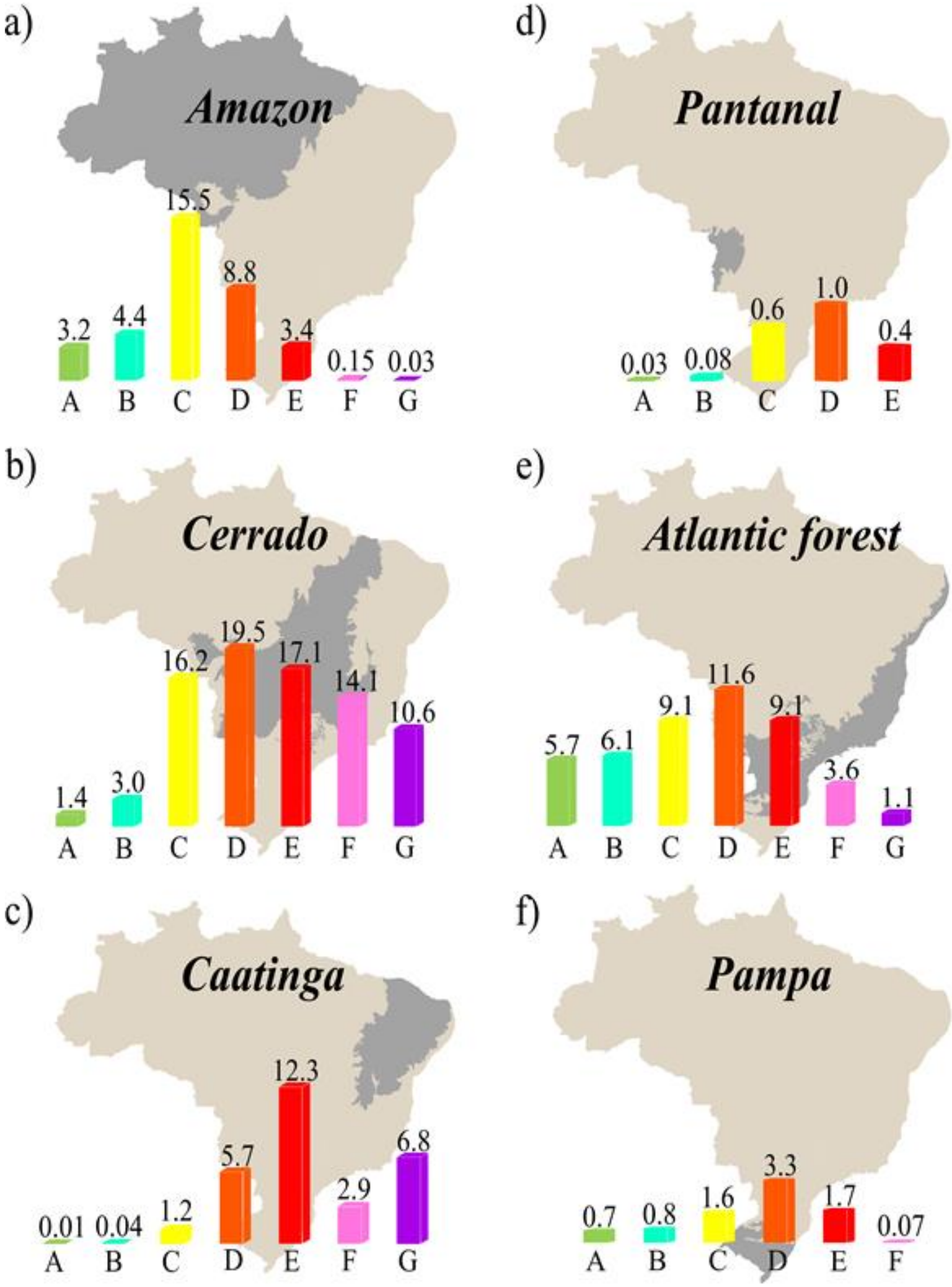

Fig. 3: Area of productive potential of agricultural soils quantified by Brazilian biomes in Millions of hectares (Mha). a) SoilPP for the Amazon biome in Mha. b) SoilPP for the Cerrado biome in Mha. c) SoilPP for the Caatinga biome in Mha. d) SoilPP for the Pantanal biome in Mha. e) SoilPP for the Atlantic Forest biome in Mha. f) SoilPP for the Pampa biome in Mha.

The Caatinga biome had an estimated 29 Mha occupied for agricultural use (Fig 3c). Most of the agricultural soils in the Caatinga show a medium to very low SoilPP (95.5% of the agricultural soils in the Caatinga). The wetland biome had most agricultural soils ranging from medium/high to medium/low classes, with 3 Mha (Fig 3d). The soils belonging to the low and very low classes in this biome correspond to an amount of less than 10 thousand hectares. The Atlantic Forest (Fig 3e) is the second-largest Brazilian biome in terms of land use for agricultural activities, occupying an area of about 46.6 Mha. Most agricultural soils located in the Atlantic Forest had a SoilPP ranging from very high to medium (69.7% of the Atlantic Forest territory). Soils with low and very low SoilPP are at 4.7 Mha (10%). Finally, the pampa biome shows most of its soils with medium SoilPP (3.3 Mha), however, a significant part of the areas has soils with very high to medium/high potential.

3.3 Assessment of croplands (national extent)

We evaluated the yield level for soybean (*Glycine max*) and sugarcane (*Saccharum officinarum*), both of which are of great agricultural and economic importance to Brazil and other countries. This assessment allows us to understand whether Brazil is making the most of SoilPP for its crops, which will result in the need to expand agricultural areas in the future. If not, yields for these crops can be increased where they are already being grown. This analysis was carried out at the municipal level, using real average values of productivity corresponding to 5 harvests (2016 - 2020). Municipalities were categorized into seven different classes (Fig. 2 a and d) according to average grain yield (Table 2). The SoilPPmap (Fig. 2) was also averaged from all the pixels that make up the Brazilian municipalities, resulting in an average value of SoilPP for each municipality (Fig 3b and e), with seven different classes (Table 2). Afterwards the municipal yield level was calculated using equation (6), where municipalities with positive values indicate that SoilPP is being exploited to the fullest. On the other hand, municipalities with negative values indicate that productivity can be increased at some level.

About 2304 Brazilian municipalities with average soybean productivity above 1500 kg ha^{-1} (Fig. 4a) were analyzed. Approximately 1051 soy-producing municipalities showed positive values. Therefore, the SoilPP is being well explored in these areas due to agricultural management and its yield level is high for the soybean, resulting in the highest average municipal yields (Fig. 4c). About 155 municipalities had a yield level equal to 0, i.e, the level of the yield of soybean crops is at adequate levels, but their yield can be increased in these municipalities. Finally, about 896 municipalities showed negative yield level values. Therefore, the yield for the soybean crop is below adequate levels, showing a difference in yield when compared to the municipalities that had a positive yield level, i.e, soybean productivity can be increased in these municipalities through management.

For sugarcane, approximately 2468 municipalities with average productivity above 35 tonnes ha^{-1} (Fig. 4d) were analyzed. Around 1070 sugarcane municipalities showed positive yield level values, that is, the productivity of sugarcane plantations in these municipalities is above SoilPP (Fig. 4f). Therefore, the soil potential is being well explored for the production of biomass in these areas due to agricultural management and its yield level is high for the crop in question. Around 342 municipalities had a yield level equal to 0, that is, the yield level of sugarcane crops is at adequate levels, but their yield can still be increased in these municipalities (Fig. 4f). Finally, 1056 municipalities presented negative yield level values. So, the yield for sugarcane is below adequate levels, showing a difference in yield when compared to the municipalities that had a positive yield level, that is, the productivity of sugarcane can be increased in these municipalities through management.

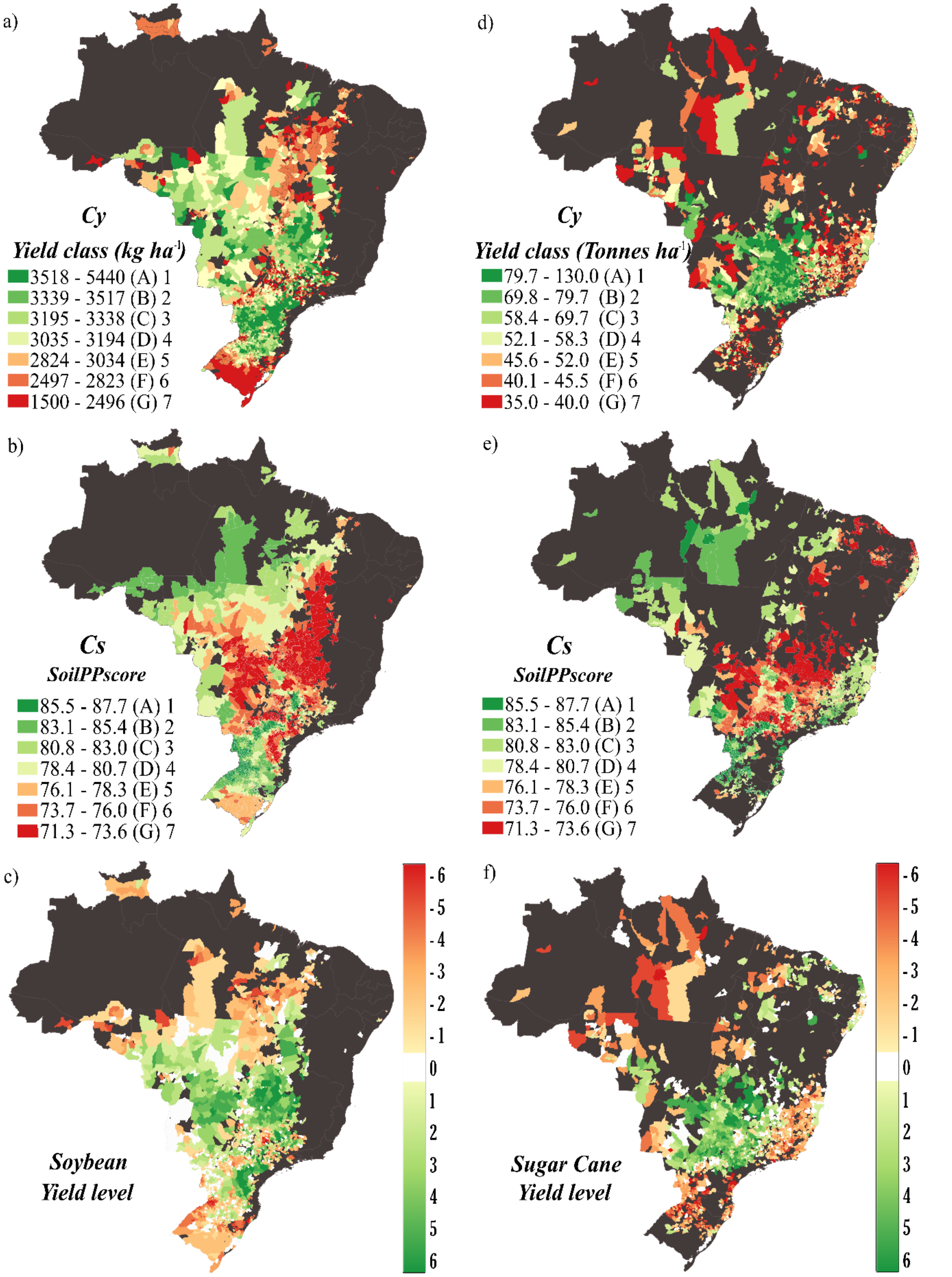

Fig. 4: analysis of the exploitation of SoilPP at the municipal level for the cultivation of sugar cane and soy. a) main Brazilian soy-producing municipalities. b) Average SoilPP of Brazilian municipalities. c) $^{(6)}$ yield level for Brazilian soybean-producing municipalities (calculated using eq. 6). d) main Brazilian sugarcane-producing municipalities. e) Average SoilPP of Brazilian municipalities. f) $^{(6)}$ yield level for Brazilian sugarcane-producing municipalities (calculated using eq. 6). Cy $_{(6)}$: Yield class; Cs $_{(6)}$: Score class.

3.3.1 Farm extent

A comparative analysis was carried out between SoilPP*map* and the Production Environments System (PES) developed by Demattê and Demattê (2009) at the farm level. The PES was developed after some empirical studies initiated by Coopersucar (1994), which observed the strong relationship between soil types and sugarcane productivity. The basis of this system is the soil pedological map, where the pillars that support plant productivity are soil classification, texture, fertility, depth, water retention (from surface to subsoil, 1 m), and climate. The PES is categorized into seven different classes (Fig. 5) due to the sugarcane productivity in each mapping unit, which corresponds to a soil class, represented by a specific color.

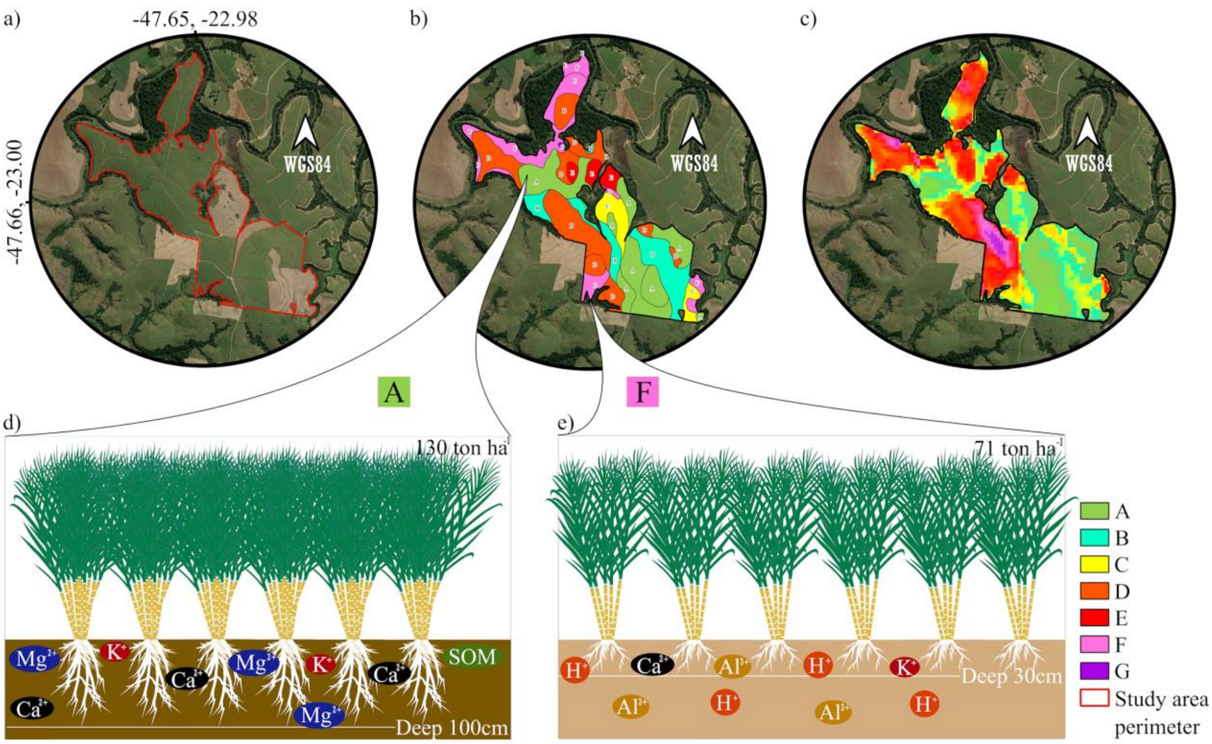

Fig. 5: a) Limit of the study area (Rafard Farm); b) Empirical production environment for sugarcane; c) SoilPP map. d) Yield of sugarcane in environment A. e) Yield of sugarcane in environment F. Sugarcane productivity categories: A: greater than 100 tons ha^{-1}; B: 90-100 tons ha^{-1}; C:86-90 tons ha^{-1}; D: 81-85 tons ha^{-1}; E: 76-80 tons ha^{-1}; F: 71-75 tons ha^{-1}; G: less than 70 tons ha^{-1}.

The farm's area is 182 ha (located in the city of Rafard, São Paulo state) (Fig. 5a). Figure 5b presents the PES for the sugarcane farm. The SoilPPmap (Fig. 2) was cut along the perimeter of the Rafard farm (Fig. 5c), making it possible to compare its similarities and differences in relation to the PES (Fig. 5b) for the sugarcane crop. Therefore, the correspondence between areas with the highest SoilPP (A) and PES mapping units with the highest yields (A) of sugarcane (Fig 5b and c) is remarkable, i.e, the most productive environments classified in the field present similarities when related to the best SoilPP classes. Fig 5d exemplifies soil conditions for environment 'A' which has a higher production of plant biomass where the soil has a clayey texture and base saturation (V%) above 50%. Figure 5e demonstrates the soil and plant conditions of an 'F' environment, where this soil has a very sandy texture, base saturation below 50%, and the presence of Al^{3+}.

3.4 Correspondence analysis between SoilPP vs. Production Environment System (PES)

Multiple Correspondence Analysis (MCA) was performed with approximately 21000 soil samples evaluated in the field via PES and used in SoilPPmap prediction, to verify the association of categorical groups between both systems. The spatial association between SoilPP and PES classes (Demattê and Demattê, 2009) is confirmed ($p<0,01$) by correspondence analysis (Fig. 6). The soil samples that obtained the highest scores (class A) via the soil productive potential index "SoilPPI", express a significant association with class A via "PES" determined in the field. Therefore, the soils that showed the highest productive potential presented high similarity with the best categories (A, B and C) of production environments. Soil samples that obtained the lowest scores (classes F and G) by SoilPPI presented greater correspondence with the worst classes of PES (Classes F and G) evaluated in the field. The category with low correspondence between the two methods was D (Fig. 6). Classes B, C, and E also demonstrate the correspondence between the different methods.

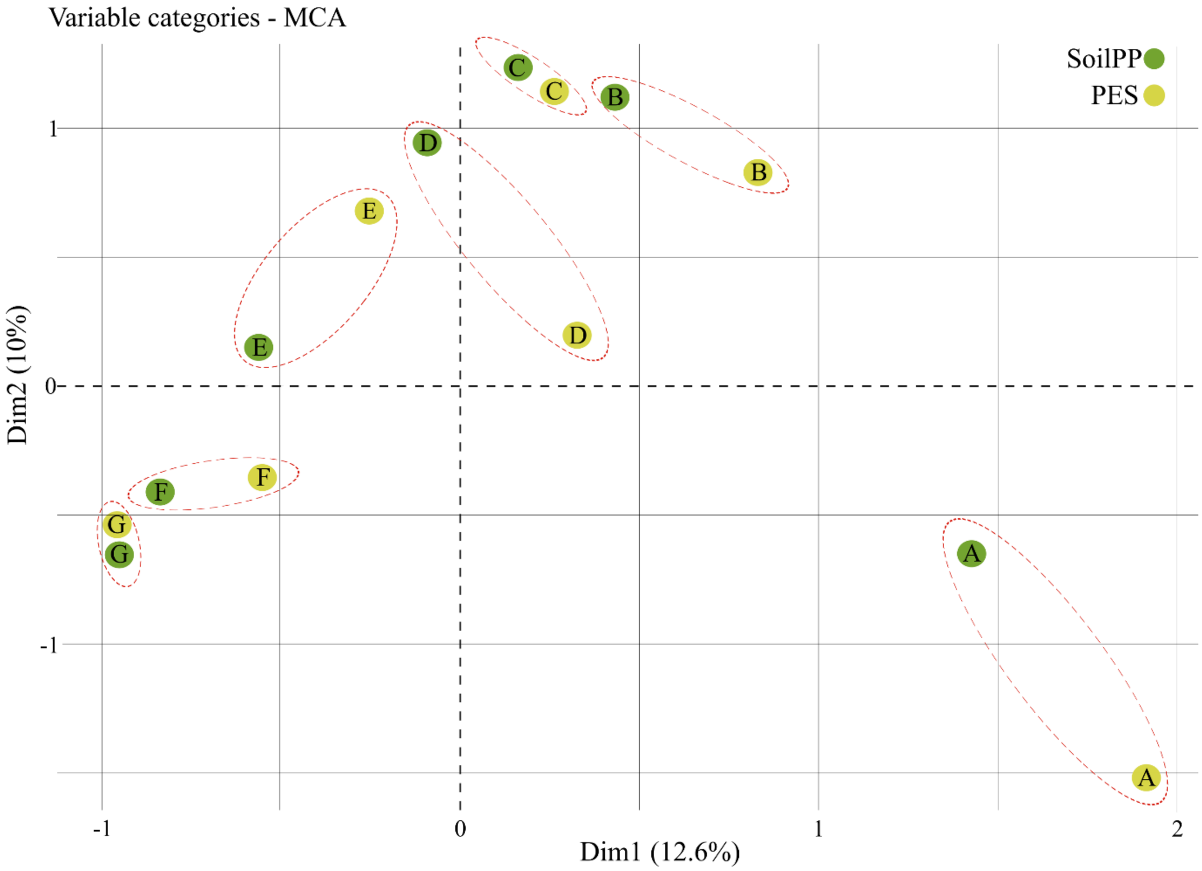

Fig 6. Correspondence analysis between Production Environment System 'PES' and SoilPP.

4. DISCUSSION

4.1 Prediction of soil productive potential (SoilPP)

The low prediction accuracy of SoilPPmap (Table 3) is possibly associated with the evaluation of soil chemical attributes in SoilPPI (Table 1). This problem is possibly linked to the variability of these attributes caused by crop management, such as liming and fertilization of agricultural lands. Safanelli et al. (2021) performed poorly for the prediction of soil pH in water (R^2 0.07) and CEC mmolc kg $^{-1}$ (R^2 0.27) for Brazilian agricultural areas using approximately 4500 soil samples. Mendes et al. (2019) performed the prediction of chemical attributes such as CEC (R^2 0.35 to 0.02), V% (R^2 0.12 to 0.05), m% (R^2 0.27 to 0.17), and Al (R^2 0.36 to 0.07), using different covariates and prediction methods, but the model accuracy was also low. Poor chemical property prediction performance also occurred in other studies (Poppiel et al., 2019; Rizzo et al., 2020). Therefore, the insertion of chemical attributes (Table 1) in the SoilPPI to generate the SoilPPmap was probably the main factor that reduced the accuracy of the prediction, since the chemical attributes had a greater influence on the index result (Table A1) and present high variability in the landscape.

4.2 SoilPPmap for Brazilian agricultural areas

First, the correspondence of the different Soil categories with the Brazilian geological map with a scale of 1:5.000.000 was observed (Gómez et al., 2019). Agricultural soils that present higher SoilPP (Fig. 2) correspond to areas of source material from volcanic rocks, mainly in territories with the presence of basalt (Fig. A1a and A2c) (Gómez et al., 2019). Areas with medium/high to very low SoilPP correspond mostly to agricultural soils with parental material from sedimentary and metamorphic rocks (Fig. A1a and A2c).

In agreement with Beerling et al., (2018) basalt is widely recognized source material for as a producer of productive soils because it weathers quickly, releasing essential elements for plant growth, including P, K, Ca, Mg and Fe. This statement justifies why the Brazilian soils from agricultural areas with the highest SoilPP (Fig. 2) are located in territories originating from basalt. The main soil classes that are located in the areas with the highest SoilPP (Fig. 2) are mainly Nitosols (Nitosols, Lixisols, and Alisols), Latosols (Ferralsols), and Ultisols (Acrisols, Lixisols, and Aisols) (Fig. A1b and A2e).

SoilPPmap showed high spatial similarity with the soil organic carbon stocks (SOC) map for Brazilian agricultural areas produced by Safanelli et al., (2021) (Fig. A2a and b). The areas that obtained the highest scores through SoilPPI (very high and high) (Fig. 2) are the same areas that demonstrate the highest concentrations of SOC. In agreement with Huang et al., (2021), SOC has the potential to improve soil structure (such as aggregate stability) and retention of water and nutrients (such as nitrogen) in soils. Other studies indicate that increasing SOCs can increase soil quality (improvement in the performance of certain soil functions, such as water retention, aggregation, and cation exchange capacity) and improve crop yields (Lal, 2006; Williams et al., 2016; Yost and Hartemink, 2019). Therefore, the territories with the highest concentration of SOC in Brazilian agricultural areas (Safanelli et al., 2021) correspond with the highest values of SoilPP, since SOC has a direct influence on the production of biomass of cultivated plants (Huang et al., 2021).

This factor can encourage farmers to reformulate their soil management for increasing the soil organic matter (SOM) content in agricultural areas. It can contribute to the global initiatives such as "Soil Carbon 4 per Mille", which proposed to increase SOC sequestration, mitigating the harmful effects of global anthropogenic greenhouse gas emissions (Minasny et al., 2017). In order to increase global food security for a rapidly growing population under a changing climate, best management practices should be adopted to improve soil structure and SOC stock that can increase water storage, nutrient retention and energy conservation (Huang et al., 2021).

The area quantification of the SoilPP*map* allowed seeing the spatial extension of the best and worst categories of agricultural soils in each Brazilian biome, based on the evaluation of the physical, chemical, and biological properties of these soils (Fig 3). The best agricultural soils in Brazil due to their territorial extension that belong to the SoilPP categories "very high and high" are found in the Atlantic Forest (11.8 M ha), Amazon (7.6 M ha), and Cerrado (4.4 M ha) (Fig. 3a, b, and e). On the other hand, soils ranging from medium/high to very low potential are found in the Cerrado (77.9 M ha), Caatinga (28.9 M ha), and Atlantic Forest (34.5 M ha) biomes (Fig. 3a, b, and e). A possible explanation for the lower values of SoilPP in the Cerrado and Caatinga biomes can be observed through the pasture quality map provided by MapBiomas (https://mapbiomas.org) on Google Earth Engine platform. Most of the pastures in the respective biomes are in a severe or moderate degree of degradation, corresponding to the worst SoilPP classes "medium/low, low and very low" (Strassburg et al., 2014).

4.3 Assessment of croplands (national extent)

Through the Yield level analysis of 2304 Brazilian soybean producing municipalities, around 896 municipalities were identified that could increase their production levels (Fig. 4c). For sugarcane, 2468 producing municipalities were evaluated, where 1056 municipalities could increase crop productivity (Fig. 4f). Some Brazilian municipalities had high average yields for both crops even at low SoilPP values, such as the Southeast, Midwest, and Northeast regions (Fig. 4a, b, d, and e). This is possibly associated with the level of technological investment of the municipalities in the respective regions. For example, the regions with the adoption of irrigation practices have the highest average productivity per municipality, which can be verified through the map of irrigated Brazilian agricultural areas (https://mapbiomas.org; ANA - National Water and Sanitation Agency).

Another factor is the large Brazilian reservoirs of residual phosphorus (P) accumulated over decades of cultivation, called “P legacy”, where Pavinato et al. (2020) mapped the Spatio-temporal distribution of the legacy of P in the last 50 years in Brazil. Some regions with the highest accumulation of P legacy in Brazilian soil (greater than 300 kg ha^{-1}) demonstrate the intense technological investment in municipalities with low SoilPP to leverage production levels for soybeans and sugarcane (Pavinato et al., 2020).

The highest values of Gross Domestic Product (GDP) per capita in 2019 (IBGE, 2019) correspond to regions of irrigated areas that belong to some agricultural frontiers of strong expansion in Brazil (ANA - National Water and Sanitation Agency). Notably, in central Mato Grosso, southern Goiás and eastern Mato Grosso do Sul, western Bahia, and the upper reaches of the Parnaíba River, where there was a high share of agricultural activities in the GDP associated with a relatively small population (IBGE, 2019). Therefore, high municipal productivity may be associated with a high technological investment in the respective regions, such as fertilization (Pavinato et al., 2020) and irrigation (ANA). Thus, the development of public policies to increase technological investment in municipalities with a negative Yield level (Fig. 4c and f), could result in an increase in their average productivity without the exploitation of new agricultural areas, contributing to food security (Huang et al. al., 2021; McBratney et al., 2014), forecasting of supply and support services (Adhikari and Hartemink, 2016; Dominati et al., 2014; Kopittke et al., 2019; Robinson et al., 2012; Tóth et al., 2013).

4.3.1 In farm extent

Sugarcane, cultivated extensively in highly weathered soils, generates approximately US$43 billion per year for the Brazilian economy (Beerling et al., 2018). The soil classes with the highest productive potential on the farm located in the municipality of Rafard (state of São Paulo) (Fig. 5c) were mainly in the mapping units (Fig. A3d) occupied by the Red Nitosol (NV) and Argilluvic Chernossolo (MT) (Bazaglia Filho et al., 2013). These units have a fine texture and base saturation (V%) greater than fifty (eutrophic soils) (Bazaglia Filho et al., 2013; Santos et al., 2018). Confirmation of the best soil classes can be observed by the PES performed from field information (Fig. A3a) and by observing its visual correspondence with SoilPPmap (Fig. A3b). The worst soils in the area classified by PES and SoilPPmap have a coarse texture (sandy) with a V% lower than 50 (dystrophic soils) (Bazaglia Filho et al., 2013). Acidification of agricultural soils is a worldwide problem and reversing it improves nutrient uptake, root growth, and crop yields (Beerling et al., 2018; Gillman, 1980). Neutralization of acidic soils also reduces metal toxicity (e.g. aluminum and manganese levels) and increases P availability, especially in highly weathered tropical acidic soils where metal oxides bind strongly to remaining P reserves (Beerling et al., 2018).

The locations with the highest SoilPP (very high and high) on the farm have diabase (mafic igneous rock) as their parental material (Nanni and Demattê, 2006) (Fig. A3e). Mello et al. (2021) evaluated the magnetic susceptibility (κ) on the same farm and the highest values of κ are found in soil classes where clay was formed from source materials rich in ferrimagnetic minerals, where NV, MT, and CX occur. (Fig. A3d) and consequently at the sites originating from diabase (Fig. A3e). The soil classes that presented the lowest SoilPP are PVA and PA (Fig. A3d), where Mello et al. (2021) observed the lowest values of κ due to the presence of sandy texture in this class, coming from siltstone (Fig. A3e). Therefore, the SoilPPmap prediction assigned the highest SoilPP values in exposed soil pixels with the lowest reflectance intensities (Demattê et al., 2018).

4.4 Correspondence analysis between SoilPP vs. Production Environment System (PES)

Through correspondence analysis (Fig. 6), it was possible to confirm the similarity between two different categorical systems, such as SoilPP and PES (Demattê and Demattê, 2009). The categories that showed the greatest difference were 'B', 'D', and 'E' when comparing the two systems. On the other hand, categories 'A', 'C', 'F', and 'G' were the ones that showed the greatest correspondence between the two categorical systems. Therefore, it is notable that the numerical system that evaluates SoilPP through the scoring of soil properties corresponds to an empirical method of PES developed by Demattê and Demattê, (2009), where the categories of both systems showed similarities.

4.5 Contributions and limitations of SoilPP*map* for understanding soils from agricultural areas in Brazil

Legacy soil maps strongly supported Brazilian agricultural expansion. However, most soil maps in Brazil date from 1970 to 1980 (RADAM BRASIL) and have low cartographic detail (1:500.000 to 1:5.000.000). This information contributed to agricultural development in several areas of Brazil. Currently, there is great interest in the elaboration of detailed maps, promoting more efficient and sustainable use of land (Safanelli et al., 2021). And these techniques can help increase global food security (Huang et al., 2021; McBratney et al., 2014), through the aid of national-scale policy formulations, such as "Brazil's National Soil Program" – PronaSolos (Polidoro et al., 2016) and "Computing the Iowa Corn Suitability Rating for Your Farm | Ag Decision Maker," (2013). They can also contribute to agricultural resource management and soil security (Lehmann et al., 2020; McBratney et al., 2014; Mulder et al., 2011).

The main limitations of SoilPPmap (Fig. 2) are related to the low accuracy of the prediction model and to agricultural areas without information on exposed soil reflectance which allowed us to assess only 80% (205 M ha) of the total Brazilian agricultural areas (263 M ha) (MapBiomas, 2020). The low precision performances are possibly associated with the evaluation of chemical attributes in SoilPPI, due to the low correlation of the chemistry with the reflectance of bare soils (Mendes et al., 2019; Poppiel et al., 2019; Rizzo et al., 2016; Safanelli et al., 2019; al., 2021) and high with SoilPPI (Table A1). Techniques for soil management and plant nutrition, such as limestone and fertilizer application, no-tillage, and biological nitrogen fixation, justify the high dynamics of soil chemical attributes (Beerling et al., 2018).

Agricultural areas that did not show reflectance of exposed soil are possibly linked to two main reasons, the incidence of clouds at some time when the soil was exposed (Demattê et al., 2018), or the decreasing rate of soil exposure, as more soils were covered due to the adoption of conservation agriculture practices (no-till), which can reduce soil degradation (Demattê et al., 2020b). With this, the complexity of working with the mapping of large territories, such as the case of Brazil, is perceptible. Despite the difficulties, mapping was carried out with a spatial resolution of thirty meters, evaluating soil properties up to one-meter-deep for 80% of Brazilian agricultural areas.

5. CONCLUSION

The proposed method for obtaining a SoilPPmap for agricultural Brazilian areas based on digital mapping of soil attributes integrated by scoring functions achieved a satisfactory spatial correspondence with external data of crop productivity at national and farm geographic extents. The indexing strategy generated through PCA to evaluate the soil attributes of the respective samples showed good categorical correspondence with the Production Environment System (PES). Therefore, this technique can assist in the development of soil products to assist global food and soil security policies and be replicated in other countries, regions, municipalities, or farms. The SoilPP*map* for Brazilian agricultural areas had the following main limitations: 1) the sample representativeness for all types of agricultural soils in Brazil; 2) the low accuracy of the prediction model, which is possibly related to the evaluation of soil chemical attributes. Despite the limitations and challenges, about 205 Mha (80% of the total agricultural areas in Brazil) were mapped with soil information up to 1 m deep and a spatial resolution of 30 m.

The area quantification of SoilPP made it possible to see the best and worst agricultural soils in each Brazilian biome. The best agricultural soils in Brazil are found in the Atlantic Forest (11.8 Mha), Amazon (7.6 Mha), and Cerrado (4.4 Mha) biomes. On the other hand, soils with medium/high to very low potential are found in the Cerrado (77.9 Mha), Caatinga (28.9 Mha), and Atlantic Forest (34.5 Mha) biomes. Through the evaluation of the average municipal productivity level, it was observed that of 2304 soybean-producing municipalities, 896 could increase the average soybean productivity to a certain level. For sugarcane, out of 2468 municipalities evaluated, 1056 can increase their average productivity. Therefore, increasing agricultural technology in some municipalities through the development of public policies could increase average productivity for the respective crops in areas where they are already cultivated. This can make it possible to reduce the opening of new areas, minimize the negative impacts of deforestation, mitigate the effects of climate change and contribute to global food security.

REFERENCES

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

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Leonardo, J., Gonçalves, M., Sparovek, G., 2013. Köppen's climate classification map for Brazil. https://doi.org/10.1127/0941-2948/2013/0507

ANA - Agência Nacional de Águas e Saneamento Básico, 2021. Atlas Irrigação. Uso da água na agricultura irrigada, Agência Nacional de Águas e Saneamento Básico.

Bazaglia Filho, O., Rizzo, R., Lepsch, I.F., Prado, H. do, Gomes, F.H., Mazza, J.A., Demattê, J.A.M., 2013. Comparison between detailed digital and conventional soil maps of an area with complex geology. Rev. Bras. Ciência do Solo 37, 1136–1148. https://doi.org/10.1590/s0100-06832013000500003

Beerling, D.J., Leake, J.R., Long, S.P., Scholes, J.D., Ton, J., Nelson, P.N., Bird, M., Kantzas, E., Taylor, L.L., Sarkar, B., Kelland, M., DeLucia, E., Kantola, I., Müller, C., Rau, G., Hansen, J., 2018. Farming with crops and rocks to address global climate, food and soil security. Nat. Plants 2018 43 4, 138–147. https://doi.org/10.1038/s41477-018-0108-y

Beerling, D.J., Leake, J.R., Long, S.P., Scholes, J.D., Ton, J., Nelson, P.N., Bird, M., Kantzas, E., Taylor, L.L., Sarkar, B., Kelland, M., DeLucia, E., Kantola, I., Müller, C., Rau, G., Hansen, J., 2018. Farming with crops and rocks to address global climate, food and soil security. Nat. Plants 2018 43 4, 138–147. https://doi.org/10.1038/s41477-018-0108-y

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

Bishopp, A., Lynch, J.P., 2015. The hidden half of crop yields. Nat. Plants 2015 18 1, 1–2. https://doi.org/10.1038/nplants.2015.117

Bordonal, R. de O., Carvalho, J.L.N., Lal, R., de Figueiredo, E.B., de Oliveira, B.G., La Scala, N., 2018. Sustainability of sugarcane production in Brazil. A review. Agron. Sustain. Dev. 2018 382 38, 1–23. https://doi.org/10.1007/S13593-018-0490-X

Breiman, L., 2001. RANDOM FORESTS. Int. J. Adv. Comput. Sci. Appl. https://doi.org/https://doi.org/10.1023/A:1010933404324

Cherubin, M.R., Karlen, D.L., Cerri, C.E.P., Franco, A.L.C., Tormena, C.A., Davies, C.A., Cerri, C.C., 2016. Soil Quality Indexing Strategies for Evaluating Sugarcane Expansion in Brazil. PLoS One 11, e0150860. https://doi.org/10.1371/journal.pone.0150860

Computing the Iowa Corn Suitability Rating for Your Farm | Ag Decision Maker [WWW Document], 2013. IOWA. URL https://www.extension.iastate.edu/agdm/wholefarm/html/c2-87.html (accessed 3.11.22).

COPERSUCAR. Potencial de produção de cana de açúcar. In: SEMINÁRIO DE TECNOLOGIA AGRONÔMICA, 7., 1997, Piracicaba. Anais. Piracicaba: Centro de Tecnologia Copersucar, 1997.

de Araújo, F.M., Ferreira, L.G., Arantes, A.E., 2012. Distribution Patterns of Burned Areas in the Brazilian Biomes: An Analysis Based on Satellite Data for the 2002–2010 Period. Remote Sens. 2012, Vol. 4, Pages 1929-1946 4, 1929–1946. https://doi.org/10.3390/RS4071929

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

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., Mendes, W. de S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., Mello, F.A. de O., Paiva, A.F. da S., Souza, A.B., Santos, N.V. dos, Maria Nascimento, C., Mello, D.C. de, Bellinaso, H., Gonzaga Neto, L., Amorim, M.T.A., Resende, M.E.B. de, Vieira, J. da S., Queiroz, L.G. de, Gallo, B.C., Sayão, V.M., Lisboa, C.J. da S., 2020a. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. Sci. Reports 2020 101 10, 1–11. https://doi.org/10.1038/s41598-020-61408-1

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., Mendes, W. de S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., Mello, F.A. de O., Paiva, A.F. da S., Souza, A.B., Santos, N.V. dos, Maria Nascimento, C., Mello, D.C. de, Bellinaso, H., Gonzaga Neto, L., Amorim, M.T.A., Resende, M.E.B. de, Vieira, J. da S., Queiroz, L.G. de, Gallo, B.C., Sayão, V.M., Lisboa, C.J. da S., 2020b. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. Sci. Reports 2020 101 10, 1–11. https://doi.org/10.1038/s41598-020-61408-1

Demattê, J.L.I., Demattê, J.A.M., 2009. Ambientes de produção como estratégia de manejo na cultura da cana-de-açúcar. Informações Agronômicas 10–18.

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

Embrapa Territorial, 2022. Síntese - Portal Embrapa [WWW Document]. URL https://www.embrapa.br/car/sintese (accessed 6.9.22).

FAO & ITPS, 2015. Intergovernmental Technical Panel on Soils. Status of the World's Soil Resources., Intergovernmental Technical Panel on Soils.

FAOSTAT, 2022. FAOSTAT [WWW Document]. URL https://www.fao.org/faostat/en/#data/QC/visualize (accessed 6.9.22).

Ferreira, M.D.P., Féres, J.G., 2020. Farm size and Land use efficiency in the Brazilian Amazon. Land use policy 99, 104901. https://doi.org/10.1016/J.LANDUSEPOL.2020.104901

Fróna, D., Szenderák, J., Harangi-Rákos, M., 2019. The Challenge of Feeding the World. Sustain. 2019, Vol. 11, Page 5816 11, 5816. https://doi.org/10.3390/SU11205816

Gerland, P., Raftery, A.E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G.K., Wilmoth, J., 2014. World population stabilization unlikely this century. Science. 346, 234–237. https://doi.org/10.1126/SCIENCE.1257469/SUPPL_FILE/GERLAND.SM.REVISION.1.PDF

Gillman, G.P., 1980. The Effect of Crushed Basalt Scoria on the Cation Exchange Properties of a Highly Weathered Soil. Soil Sci. Soc. Am. J. 44, 465–468. https://doi.org/10.2136/SSSAJ1980.03615995004400030005X

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/J.GEODERMA.2019.01.007

Gómez, J., Schobbenhaus, C., Montes, N.E., 2019. Geological Map of South America 2019. Scale 1:5 000 000. https://doi.org/https://doi.org/10.32685/10.143.2019.929

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

Guerra, A., Reis, L.K., Borges, F.L.G., Ojeda, P.T.A., Pineda, D.A.M., Miranda, C.O., Maidana, D.P.F. de L., Santos, T.M.R. dos, Shibuya, P.S., Marques, M.C.M., Laurance, S.G.W., Garcia, L.C., 2020. Ecological restoration in Brazilian biomes: Identifying advances and gaps. For. Ecol. Manage. 458, 117802. https://doi.org/10.1016/J.FORECO.2019.117802

Huang, J., Hartemink, A.E., Kucharik, C.J., 2021. Soil-dependent responses of US crop yields to climate variability and depth to groundwater. Agric. Syst. 190, 103085. https://doi.org/10.1016/J.AGSY.2021.103085

IBGE, 2019. Produto Interno Bruto dos Estados. Inst. Bras. Estatística e Geogr. 57, 1–2.

Johnston, A.E., Poulton, P.R., 2018. The importance of long-term experiments in agriculture: their management to ensure continued crop production and soil fertility; the Rothamsted experience. Eur. J. Soil Sci. 69, 113–125. https://doi.org/10.1111/EJSS.12521

Kopittke, P.M., Menzies, N.W., Wang, P., McKenna, B.A., Lombi, E., 2019. Soil and the intensification of agriculture for global food security. Environ. Int. 132, 105078. https://doi.org/10.1016/J.ENVINT.2019.105078

Kumar, U., Kumar, N., Mishra, V.N., Jena, R.K., 2019. Soil Quality Assessment Using Analytic Hierarchy Process (AHP) 1–18. https://doi.org/10.4018/978-1-5225-7784-3.ch001

Lal, R., 2006. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. L. Degrad. Dev. 17, 197–209. https://doi.org/10.1002/LDR.696

Landell, M.G. de A., do Prado, H., de Vasconcelos, A.C.M., Perecin, D., Rossetto, R., Bidoia, M.A.P., Silva, M. de A., Xavier, M.A., 2003. Oxisol subsurface chemical attributes related to sugarcane productivity. Sci. Agric. 60, 741–745. https://doi.org/10.1590/S0103-90162003000400020

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020a. The concept and future prospects of soil health. Nat. Rev. Earth Environ. 1, 544–553. https://doi.org/10.1038/s43017-020-0080-8

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020b. The concept and future prospects of soil health. Nat. Rev. Earth Environ. 2020 110 1, 544–553. https://doi.org/10.1038/s43017-020-0080-8

Mapbiomas, 2022. Mapbiomas Brasil [WWW Document]. URL https://mapbiomas.org/colecoes-mapbiomas-1?cama_set_language=pt-BR (accessed 6.9.22).

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

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

Mello, D., Alexandre Melo Demattê, J., Alcantara de Oliveira Mello, F., Roberto Poppiel, R., ElizabetQuiñonez Silvero, N., Lucas Safanelli, J., Barros e Souza, A., Augusto Di Loreto Di Raimo, L., Rizzo, R., Eduarda Bispo Resende, M., Ernesto Gonçalves Reynaud Schaefer, C., 2021. Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. Geoderma 381. https://doi.org/10.1016/j.geoderma.2020.114736

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

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

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

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

Nanni, M.R., Demattê, J.A.M., 2006. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. Soil Sci. Soc. Am. J. 70, 393–407. https://doi.org/10.2136/SSSAJ2003.0285

Nyle Brady, Ray R. Weil, 2012. Elementos da Natureza e Propriedades dos Solos, 3a edição. ed. Bookman.

Pavinato, P.S., Cherubin, M.R., Soltangheisi, A., Rocha, G.C., Chadwick, D.R., Jones, D.L., 2020. Revealing soil legacy phosphorus to promote sustainable agriculture in Brazil. Sci. Reports 2020 101 10, 1–11. https://doi.org/10.1038/s41598-020-72302-1

Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, andÉdouard, Duchesnay EDOUARDDUCHESNAY, Fré., 2011. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. J. Mach. Learn. Res. 12, 2825–2830.

Polidoro, J.C., 2016. Programa Nacional de Solos do Brasil (PronaSolos)., Programa Nacional de Solos do Brasil (PronaSolos).

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

Rizzo, R., Demattê, J.A.M., Lepsch, I.F., Gallo, B.C., Fongaro, C.T., 2016. Digital soil mapping at local scale using a multi-depth Vis–NIR spectral library and terrain attributes. Geoderma 274, 18–27. https://doi.org/10.1016/J.GEODERMA.2016.03.019

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

The R Project for Statistical Computing | CiNii Research [WWW Document], n.d. URL https://cir.nii.ac.jp/crid/1570572701250545792 (accessed 6.9.22).

Robinson, D.A., Hockley, N., Dominati, E., Lebron, I., Scow, K.M., Reynolds, B., Emmett, B.A., Keith, A.M., de Jonge, L.W., Schjønning, P., Moldrup, P., Jones, S.B., Tuller, M., 2012. Natural Capital, Ecosystem Services, and Soil Change: Why Soil Science Must Embrace an Ecosystems Approach. Vadose Zo. J. 11. https://doi.org/10.2136/VZJ2011.0051/111525

Runyan, C.W., Stehm, J., 2019. Land Use Change, Deforestation and Competition for Land Due to Food Production. Encycl. Food Secur. Sustain. 21–26. https://doi.org/10.1016/B978-0-08-100596-5.21995-1

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

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

Santos et al., 2018. Sistema brasileiro de classificação de solos, Embrapa Solos.

Strassburg, B.B.N., Latawiec, A.E., Barioni, L.G., Nobre, C.A., da Silva, V.P., Valentim, J.F., Vianna, M., Assad, E.D., 2014. When enough should be enough: Improving the use of current agricultural lands could meet production demands and spare natural habitats in Brazil. Glob. Environ. Chang. 28, 84–97. https://doi.org/10.1016/J.GLOENVCHA.2014.06.001

Tóth, G., Gardi, C., Bódis, K., Ivits, É., Aksoy, E., Jones, A., Jeffrey, S., Petursdottir, T., Montanarella, L., 2013. Continental-scale assessment of provisioning soil functions in Europe. Ecol. Process. 2, 1–18. https://doi.org/10.1186/2192-1709-2-32/FIGURES/7

Vogel, H.J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., Wiesmeier, M., Wollschläger, U., 2019. Quantitative Evaluation of Soil Functions: Potential and State. Front. Environ. Sci. 7, 164. https://doi.org/10.3389/FENVS.2019.00164/BIBTEX

Williams, A., Hunter, M.C., Kammerer, M., Kane, D.A., Jordan, N.R., Mortensen, D.A., Smith, R.G., Snapp, S., Davis, A.S., 2016. Soil Water Holding Capacity Mitigates Downside Risk and Volatility in US Rainfed Maize: Time to Invest in Soil Organic Matter? PLoS One 11, e0160974. https://doi.org/10.1371/JOURNAL.PONE.0160974

Yang, J., Hammer, R.D., Thompson, A.L., Blanchar, R.W., 2003. Predicting soybean yield in a dry and wet year using a soil productivity index. Plant Soil 250, 175–182. https://doi.org/10.1023/A:1022801322245

Yost, J.L., Hartemink, A.E., 2019. Effects of carbon on moisture storage in soils of the Wisconsin Central Sands, USA. Eur. J. Soil Sci. 70, 565–577. https://doi.org/10.1111/EJSS.12776

APPENDIX

Table A1. Principal component analysis (PCA) for soil attributes.

	PC1	PC2	PC3	PC4	PC5	PC6
Sand	-0.217	0.344	-0.084	-0.159	0.169	-0.029
Silt	0.212	-0.142	-0.095	0.241	-0.232	0.049
Clay	0.180	-0.368	0.146	0.095	-0.112	0.015
Bulk Density	-0.215	0.361	-0.146	0.036	-0.029	-0.019
SOC	0.221	-0.209	0.092	-0.346	0.365	0.020
SOM	0.221	-0.208	0.093	-0.345	0.365	0.021
pH in water	0.204	0.229	0.208	0.055	0.022	0.163
Ca^{+2}	0.322	0.109	-0.021	0.061	0.001	0.004
Mg^{+2}	0.302	0.108	-0.083	0.122	-0.014	-0.046
K^{+}	0.176	0.059	-0.048	0.098	0.204	-0.267
Al^{+3}	-0.041	-0.166	-0.456	0.000	-0.129	-0.048
$H^{+}+Al^{+3}$	0.077	-0.267	-0.373	-0.139	-0.072	-0.020
CEC pH7	0.285	-0.096	-0.278	-0.030	-0.039	-0.036
SB	0.331	0.114	-0.044	0.088	0.012	-0.033
V%	0.249	0.281	0.134	0.101	-0.003	-0.023
m%	-0.219	-0.191	-0.261	0.023	0.003	-0.030
Delta pH	0.011	0.075	-0.127	0.492	0.589	0.163
Ki	0.042	0.101	-0.038	0.156	-0.131	-0.716
Clay activ.	0.072	0.245	-0.354	-0.343	0.022	0.121
Slope	0.053	0.016	-0.106	0.234	-0.238	0.568

Table A2. Importance of the Principal Component Analysis.

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	2.81	2.00	1.78	1.18	1.12	1.04
Proportion of variance	0.34	0.17	0.14	0.06	0.05	0.05
Cumulative proportion	0.34	0.52	0.65	0.71	0.77	0.81

***** The proportion of variance (PCn) of each selected principal component is one of the variables for calculation through the SoilPPI equation (5).

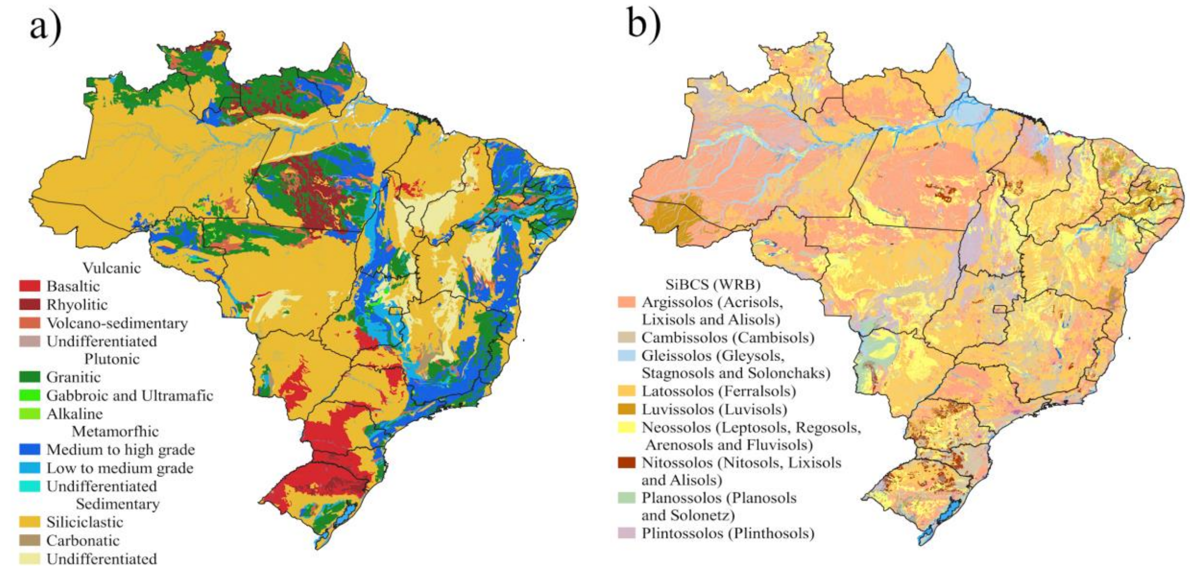

Fig. A1: a) Brazilian geological map "scale 1:5.000.000" (Gómez et al., 2019). b) Brazilian pedological classification (scale 1:5.000.000).

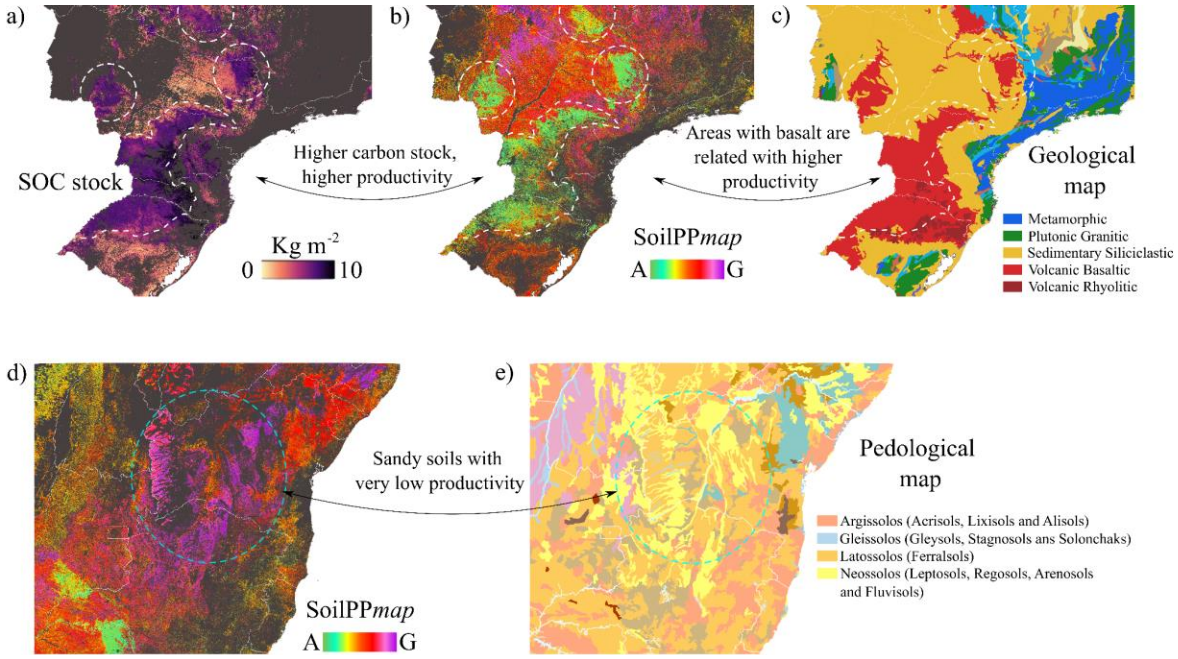

Fig. A2. a) SOC stock Brazilian map (Safanelli et al., 2021); b) SoilPP*map*; c) Brazilian geological map (Gómez et al., 2019); d) SoilPP*map*; e) Brazilian pedological classification (Santos et al., 2018).

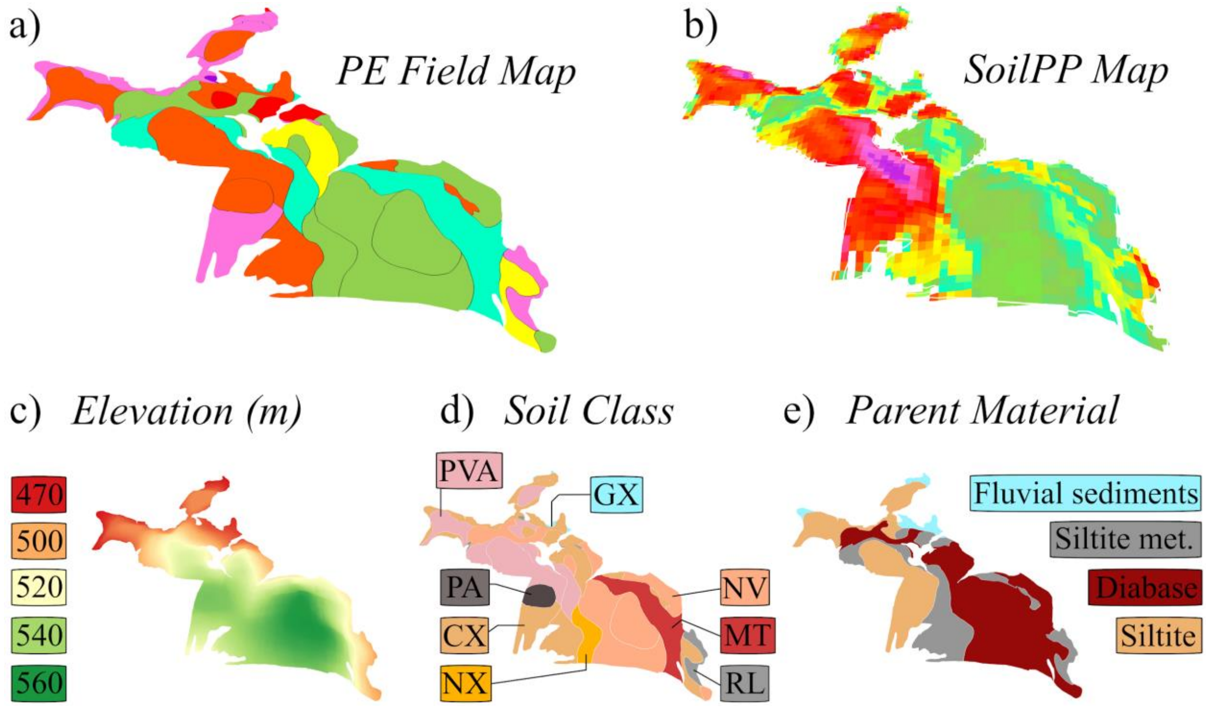

Fig. A3: a) PES*map* based on pedological classification and sugarcane productivity; b) SoilPP*map* carried out from the evaluation of soil properties (equation 1, 2, 3, 4 and 5) (Table A1 and A2); c) Elevation of the terrain, (meters); d) Pedological classification map represented by soil classes: PVA: Argissolo vermelho amarelo; PA: Argissolo amarelo; CX: Cambissolo háplico; NX: Nitossolo háplico; NV: Nitossolo vermelho; MT: Chernossolo argilúvico; RL: Neossolo Litólico; e) Parent material.