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

Synergy between digital soil mapping and crop modeling: influence of soil data on sugarcane attainable yield

Natasha Valadares dos Santos

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

Piracicaba
2021

Natasha Valadares dos Santos
Agronomist

Synergy between digital soil mapping and crop modeling: influence of soil data on sugarcane attainable yield

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
2021

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

Santos, Natasha Valadares dos

Synergy between digital soil mapping and crop modeling: influence of soil data on sugarcane attainable yield / Natasha Valadares dos Santos. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

53 p.

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

1. DSSAT 2. Qualidade de dados de solo 3. Variabilidade espacial de solos 4. Previsão de produtividade em grade I. Título

To the Author of life, who gives intuition to the heart
and instinct to the mind (Job 38:36).

ACKNOWLEDGMENT

I thank the One who breathed life into me. I am thankful for all his blessings and for having dreams to me that exceed my understanding.

I thank my father for being my safe port and a wonderful friend for all the time. Mom, thank you for all the care and adventures by your side, you bring joy to life. My sister and best friend, thanks for all the shared moments, the days are lighter and more fun with you. Ilza, my second mother, thank you for always encouraging me and for all the support.

I thank to my eternal roommate Camila. Thank you for receiving me so well. I will always remember all our conversations, laughs, cries and adventures together. Anna (Cup), I am grateful for your friendship over the years. Thank you for every encouraging word and shared moments. You are part of every single achievement. Ana (Crocs) and Thaís (D-Celva), thank you for always adding a bit of craziness to every moment, life is so much better with you by my side.

To Matheus, Philip and Wesley. Thank you for all the incredible and unforgettable moments. I will never forget our trips, our nights together laughing, eating and having a lot of fun. May the world prepare for our journeys together!

To all my colleagues at the GeoCis laboratory, for their help, company, coffees and chats. My special thanks to Rodnei, for the help and to guide the development of this work. I am grateful for his patience, support and friendship. Thank you to my advisor, José Alexandre, for all the mentoring over these years. You always encouraging me to be a better professional!

I also thank the "Luiz de Queiroz" College of Agriculture (ESALQ/ USP) for the opportunity to attend a master course. The Department of Biosystems Engineering, in particular Prof. Sentelhas, for all the support for the development of this project.

My thanks to Coordination for the Improvement of Higher Education Personnel (CAPES) for the granting of the scholarship. I also thank the joint project of Raízen Company with the Luiz de Queiroz Agricultural Studies Foundation (Grant nº 87017).

Finally, I would like to thank to all the people that I was not able to mention in those few words. Every single person touched me in a special way and will be remembered for going through this journey with me. My sincere thanks to everyone!

“Wisdom is a tree of life to those who embrace her;
happy are those who hold her tightly.”.

Proverbs 3:18

CONTENTS

RESUMO 7

ABSTRACT 8

1. INTRODUCTION 9

2. MATERIAL AND METHODS 12

2.1. Study area and soil database 12
2.2. Digital soil mapping of soil attributes 13
2.3. Pedotransfer functions and soil hydrological attributes 15
2.4. The DSSAT/CANEGRO model 16
2.5. Impact of different soil data sources in yield predictions 17
2.6. Influence of the soil attributes' uncertainties on the estimation of attainable sugarcane yield 18
2.7. Sugarcane yield assessment 19

3. RESULTS 20

3.1. Spatial predictions of soil attributes and uncertainty 20
3.2. Site-specific evaluation of DSSAT/CANEGRO: Influence of soil data sources in yield predictions 24
3.3. MCS 25
3.4. Spatially explicit simulation of sugarcane attainable yield 25

4. DISCUSSION 30

4.1. Performance of digital soil mapping 30
4.2. Simulated sugarcane yield 31
4.3. Spatial sugarcane attainable yield 32

5. CONCLUSIONS 35

REFERENCES 36

APPENDIX A. 49

RESUMO

Sinergia entre mapeamento digital de solos e modelagem de culturas: influência dos dados do solo na produtividade atingível da cana-de-açúcar

Os modelos de produção agrícola desempenham um papel fundamental na segurança alimentar, prevendo futuros desafios agrícolas e apoiando o estabelecimento de políticas públicas e práticas de gestão sustentável. Entretanto, devido à falta de informações confiáveis, especialmente nos países em desenvolvimento, eles apresentaram desempenho limitado e restrições para análises espacialmente explícitas. Assim, o objetivo deste estudo foi avaliar o MDS (Mapeamento Digital de Solos) como uma alternativa para preencher a ausência de dados do solo. A região considerada nesse estudo está situada no sudeste do Brasil em uma área de 4.815 km² que detém enorme heterogeneidade quanto a sua geologia e tipos de solo. Para a realização do estudo as seguintes estapas foram realizadas: (i) Foi usado um conjunto de dados de solo, obtidos a partir de 1.125 tradagens e 27 perfis que foram pradronizados em profundidades por meio de equações de interpolação; (ii) Um algoritmo de aprendizado de máquina (AM) foi usado para predição dos atributos de solo e suas incertezas (iii) Funções de pedo-transferência foram realizadas para obter as propriedades hidrológicas do solo (iv) DSSAT/CANEGRO foi simulado em uma grade de 250 m para cana-de-açúcar, com plantio em outubro e colheita completando 12 meses (v) Três níveis de fonte de dados do solo foram comparados: um mapa de solo (MS) (escala 1:100.000), SoilGrids (SG) e o mapa de atributos (MA) derivado de nosso AM. A argila foi o atributo que obteve o melhor desempenho em superfície e subsuperfície (R^2=0,70 e 0,59, RMSE= 88,87 e 141 g kg^{-1}) e baixa incerteza (40 e 110%). Em profundidade, os atributos obtiveram uma redução em seu teor e aumento da incerteza. Portanto, o MA foi a fonte de dados de solo mais confiável, sendo a que mais se assemelha aos dados de campo, apresentando o melhor índice de concordância (d= 0,8) e coeficiente de confiança (c=0,74). Além disso, uma grade de 250 m permitiu a avaliação da variabilidade espacial da produtividade atingível da cana-de-açúcar em nível regional. Os nitossolos alcançaram maior produtividade e os solos rasos não excederam 100 t ha^{-1}. Sendo assim, este trabalho mostrou a aplicabilidade do mapeamento digital para uso na modelagem de culturas. Esta metodologia pode ser replicada no planejamento agrícola em nível regional e aplicações de manejo na agricultura.

Palavras-chave: DSSAT, Variabilidade espacial de solos, Qualidade de dados de solo, Previsão de produtividade em grade

ABSTRACT

Synergy between digital soil mapping and crop modeling: influence of soil data on sugarcane attainable yield

Models of crop production play a key role in food security, predicting future agriculture challenges and supporting the establishment of public policies and sustainable management practices. However, due to the lack of reliable information, especially in developing countries, they have presented limited performance and restrictions for spatially explicit analyses. Thus, the objective of this study was to evaluate the DSM (Digital Soil Mapping) as an alternative to fill the gap of soil data. Our study site is in Southwest of Brazil in a 4,815 km² area heterogeneous in geology and soil classes. The study were conducted with the following framework: (i) We used a soil survey data, containing 1,125 collected points with auger and 27 profiles and applied equal-spline equations to standardized the soil dataset into depth; (ii) A machine learning (ML) algorithm were used to predict soil attributes and their uncertainties (iii) Pedotransfer functions were performed to obtain soil hydrological properties (iv) DSSAT-Canegro was simulated in a 250m grid to sugarcane planted in October with harvest completing 12 months (v) We compared three levels of soil data source: a soil map (SM) (1:100,000 scale), SoilGrids (SG) and the map of attributes (MA) derived from our ML. Clay was the attribute that obtained the best performance to surface and subsurface (R^2=0.70 and 0.59, RMSE= 88.87 and 141 g kg^{-1}) and low uncertainty (40 and 110%). In depth the attributes were reduced in their content and increased uncertainty. Therefore, the MA to be the most reliable source of data, being the one that most resembles field data, presents the best index of agreement (d= 0.8) and confidence coefficient (c=0.74). In addition, a 250m grid allowed the evaluation of the spatial variability of the attainable yield of sugarcane at a regional level. Nitisols achieved higher productivity and shallow soils did not exceed 100 t ha^{-1}. Thus, this work showed the applicability of digital mapping for application in crop modeling. This methodology can be replicated for decision-making at a regional level and also to improve management strategies for agriculture.

Keywords: DSSAT, Spatial soil variability, Soil data quality, Grid yield forecast

1. INTRODUCTION

The sugarcane sector plays an important role in the production of biofuels, energy, and sugar. Brazil is the biggest world producer of sugarcane, with its south-central region contributing to almost 90% of production, which represents 8.5 million hectares cultivated in 2020 (CONAB, 2020). Although Brazil is considered a key supplier for future bioenergy demands, the sector is not prepared for the rising demand and the production is currently stagnant (CONAB, 2020). Once sugarcane expansion is not a viable option (Bordonal et al., 2018), new strategies must be designed to improve yields where the crop is already cultivated, which will allow optimizing the use of natural resources (Gerland et al., 2014; Runyan and Stehm, 2019). Therefore, identifying the potential of sugarcane production is an important step towards to make this crop and its entire sector more sustainable.

Currently, few tools are capable of indicating the crop potential production at farm or regional levels (Antle et al., 2017). A method extensively used in sugarcane fields is the Production Environments (PE) approach. In this technique, fields are divided in zones (named PEs) with different yield potentials. The zoning process considers basically the soil and normal climate conditions (Demattê and Demattê, 2009; Landell et al., 2003). Although PE has been widely used by the sugarcane sector to assess yield variability in crop fields, the method is partially empirical and is based on a limited number of attributes from soil and climate. Besides, for more detailed zoning, PE definition requires detailed soil maps (Sanches et al., 2019), being resource intensive and time consuming.

An alternative approach to assess yield in sugarcane fields is the use of crop simulation models (Dias and Sentelhas, 2017; Scarpare et al., 2016). Process-based crop models simulate plant growth, development, and yield, according to the genotype, weather, soil, and crop management (Jones et al., 2003, 2017). These models have been widely applied to understand the impacts of climate variability and change in agricultural production (Karthikeyan et al., 2020; Liang et al., 2020; Sachin et al., 2019), to rationalize resources in precision agriculture (Thorp et al., 2008) and to support improvements in sustainable agricultural practices (Miao et al., 2006; Sentelhas et al., 2015). They are usually employed in site-specific evaluations (i.e., simulations considering only a pair of geographic coordinates) at a field-scale. However, there is an increasing interest to apply the crop models spatially, which requires detailed weather and soil data for high resolution simulations both at local and regional levels (Machwitz et al.,

2019; Zhao et al., 2013). A spatially-explicit approach could explain the yield variability across the field and to guide a more rational crop planning and decision-making to improve crop yield (Basso et al., 2001).

Among the existing crop modeling frameworks, the most widely used is the Decision Support for Agrotechnology Transfer (DSSAT) which is a platform that encompasses several crop simulation models (Hoogenboom et al., 2019). Recently, DSSAT was coupled to a spatial analysis platform named CRAFT (Crop Forecasting Toolkit), which allows users to perform grid-based simulations with 5 or 30 arc-minutes resolution (Shelia et al., 2019). In this case, soil data required by the model can be retrieved from conventional soil maps and/or databases like SoilGrids (Hengl et al., 2017). However, there are limitations when using large-scale soil data, mostly due to insufficient representation of the spatial variability in small areas (Baigorria et al., 2007; Launay and Guerif, 2005; Pagani et al., 2017; Sadler et al., 2000). In fact, the lack of detailed information has been recognized as one of the greatest challenges when using crop models in spatially explicit analyses (Guo et al., 2018; Nendel et al., 2013). Furthermore, soil data uncertainties could propagate throughout the model simulations, and it may have an impact even greater than any other input (Folberth et al., 2016).

In this context, Digital Soil Mapping (DSM) is an alternative to produce updated and finer scale soil data. In DSM, maps of soil types and attributes are produced by a computer-assisted method, which uses spatial and non-spatial soil inference systems coupled with field and laboratory observational methods (McBratney et al., 2003; Scull et al., 2003). The great advantage of DSM is the rapid production and standardization of soil data, which can be promptly used as input in crop simulation models. DSM has proven to be a reliable soil data source for crop modeling, reaching similar results to point field data (Tewes et al., 2020). The potential for developing and disseminating soil data, including the spatially explicit ones, has raised the interest of crop modelers around the world (Han et al., 2019; Lagacherie et al., 2000; Wimalasiri et al., 2020).

To the best of our knowledge, Han et al. (2019) and Shelia et al. (2019) were the only ones to standardized soil spatial data and later used them in a grid-based DSSAT modeling. Therefore, this study aims to evaluate if DSM techniques could aid in crop modeling, and demonstrate the advantages of coupling such methods to regional-scale predictions of sugarcane attainable yield. By accomplishing our goals, we expect to raise even more attention to the importance of a synergistic use between DSM and crop simulation models, as well as

demonstrate the importance of soil scientists to standardize and share their soil datasets with the most diversified group of users as possible.

2. MATERIAL AND METHODS

2.1. Study area and soil database

The study site has 4,815 km² and is located in southeastern Brazil, at São Paulo State (Figure 1). The region presents a complex geology with soils being developed from different parent materials, such as sandstones, siltstones, shales, schists, basalts, diabases, limestones and conglomerates (Mezzalira, 1965). According to Köppen system, the climate is characterized by hot summer, comprising the months of January to March, and dry winter from June to August. The mean temperature ranges from 18 to 22 °C, and the average annual precipitation corresponds to 1200 mm (Alvares et al., 2013).

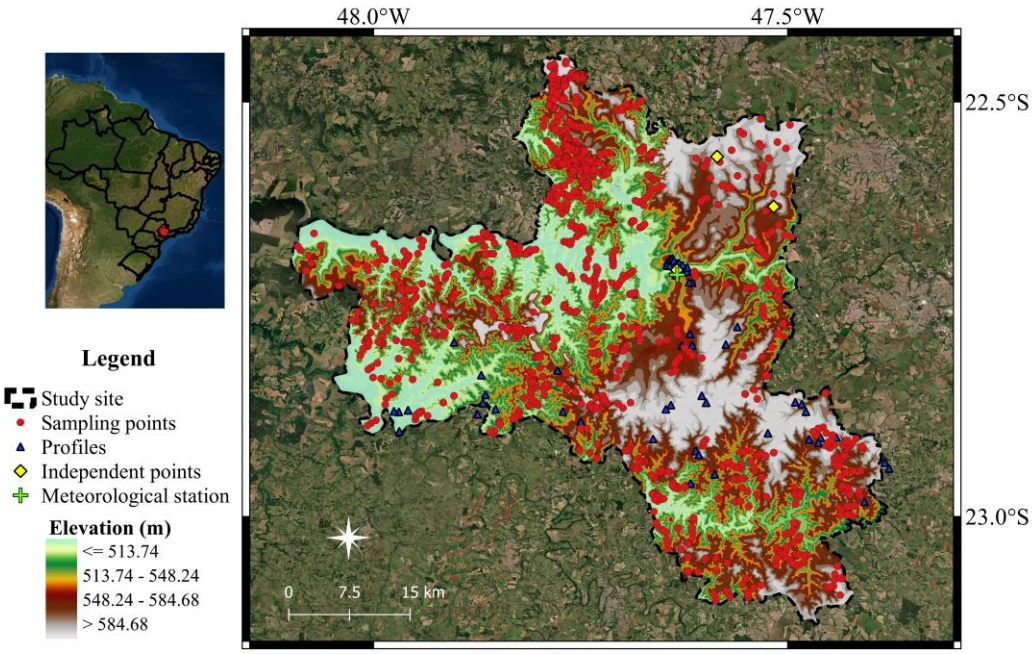


Figure 1. Study area and distribution of the soil dataset.

A soil survey was conducted in the site, where 27 soil profiles were characterized and 1125 soil samples were collected with an auger, at three depths: 0-20; 40-60; and 80-100 cm. All samples were air-dried at 45 º C for 24 h, grounded and sieved through 2 mm mesh. Chemical analyses were performed to quantify pH (water and KCl), exchangeable bases (Ca^{2+}+,

Mg^{2+} and K^{+}), exchangeable aluminum (Al^{3+}), potential acidity ($H^{+}+Al^{3+}$), organic matter (OM) and total iron concentration (Fe_2O_3) by sulfuric acid digestion (Camargo et al., 2009). We calculated the sum of bases (SB), cation exchange capacity (CEC), percentage of base saturation (Bsat %), percentage of aluminum saturation (Alsat %) and clay activity. Clay content was determined by the densimeter method, sand content by sieving and silt content by difference between the other fractions (Camargo et al., 2009).

The resulting dataset had samples at mismatching depths due to different sampling methods, i.e., collected from auger pits and soil profiles. Therefore, the results were interpolated between the depths, with 10 cm intervals until 100 cm and in 50 cm to 200 cm. We used an equal-area splines technique, which assumes that soil attributes vary smoothly in the profile. The method uses a polynomial function to estimate the attribute values at the standardized depths (Bishop et al., 1999; Malone et al., 2009). If the original value was already sampled in a standard depth, no prediction is performed by the function. The package "GSIF" in R environment (R Core Team, 2020) was used for depth harmonization (Hengl, 2020). The lambda (λ) parameter was set to 0.1, which gives the model flexibility to generate for the model in terms of smoothly and fidelity of the observed data.

2.2. Digital soil mapping of soil attributes

Herein we employed a methodology similar to the one used by Fongaro et al. (2018), where a machine learning algorithm (ML) was calibrated using as covariates a Landsat Bare Surface Composite Images (BSCI) and terrain attributes. Briefly, the BSCI is a per-pixel mosaic representing the bare surface of landscape (i.e., a synthetic satellite image representing the exposed soil surface), which is retrieved by the GEOS3 methodology (Demattê et al., 2018). Considering that our study site is predominantly an agricultural area, it's bare surface will be directly related to the soil superficial layer. More information about the method and data inputs can be found elsewhere (Demattê et al., 2020, 2018).

A Digital Elevation Model (DEM) of 5 m was derived from the topographic maps from IGC (Institute of Geography and Cartography of the state of São Paulo) at 1:1000 scale. Elevation and five terrain attributes were derived from DEM: slope, horizontal and vertical

curvature, shape index and aspect. These terrain covariates were generated using the terrain analysis function in SAGA GIS (Conrad et al., 2015).

Later, we calibrated prediction models with a ML algorithm called gradient boosting trees (GBT), which constructs an ensemble of multiple decision trees and returns their mean prediction (Breiman, 2001). Analyses were performed in Python 3.7 (Python Software Foundation, 2018) and models were calibrated to predict clay, sand, silt, organic matter (OM), cation exchange capacity (CEC) and pH in water. Soil depth model followed the same calibration steps, but in this case we also added conventional soil maps as covariates. During the processing, the soil samples database was randomly divided in calibration (80% for model training) and validation subsets (20% for test). We used the scikit-learn Python library to obtain the quantile regression from the gradient boosting trees (QGT). The hyperparameters settings were tested for a varying number of samples leaf (10, 20, 30, 40, 50, 100, 200, 500), number of random predictors (3, 5, 8, 11, 13) and forest size (30, 60, 100, 200). The model performance was evaluated by the following parameters: the root mean square (RMSE, Equation 1) to assess the model's inaccuracy, the variance of the model by the coefficient of determination (R^2, Equation 2), and the ratio of performance to interquartile range (RPIQ, Equation 3) to evaluate the consistency between the predicted values and the variability of the test dataset (Khaledian and Miller, 2020).

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

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

$$RPIQ = \frac{(P_{75}-P_{25})}{RMSE} \quad (3)$$

where $\hat{y}$ is the predicted values, y is the measured values, $\bar{y}$ is the mean of y. P75 is the 75% quantile of the prediction (%) and P25 is the 25% quantile of the prediction (%).

Predictions of uncertainty were defined by calculating the difference between the 5th and 95th percentiles using bootstrapped samples with GBT. As a result of the difference of the quartiles, it was obtained the 90% prediction interval (PI), which later were converted to a percent scale by dividing the prediction interval by the median estimate (Safanelli et al., 2020). High percent values are indicative of a high degree of uncertainty. As uncertainty is a standardized parameter, it allows comparing the different attributes maps of attributes and

their spatial uncertainties. Finally, all the predicted maps were later resampled to 250 m to facilitate the further processing steps (Figure 2).

2.3. Pedotransfer functions and soil hydrological attributes

Soil hydrological properties are extremely important for DSSAT simulations (Singels and Bezuidenhout, 2002), once they are inputs in the water balance model, which defines the water availability for the crops. This information is very scarce in Brazil and the only dataset available has a few hundreds of samples spread across the country (Ottoni et al., 2018). In this sense, pedotransfer functions developed for tropical soils were applied to obtain the soil water content at wilting point (WP), field capacity (FC), saturation (SAT) and also bulk density (BD) (Tomasella et al., 2000). Besides these attributes, it was also necessary to define drainage rate (DR) and runoff (RO). The DR represents the capacity of percolating water along the profile, and it is usually divided in seven permeability classes (Ritchie et al., 1990): Very slow, slow, moderated slow, moderate, moderate rapid, rapid and very rapid. We assessed the permeability in agreement with these classes considering the clay content and slope in percent (%). The RO was calculated according to Gijsman et al. (2007) which classifies groups by the letter's "A" to "D". Groups with shallow soils and high values of slope are considered to have high runoff potential (A), and deep soils at flat areas are classified as low runoff potential (D).

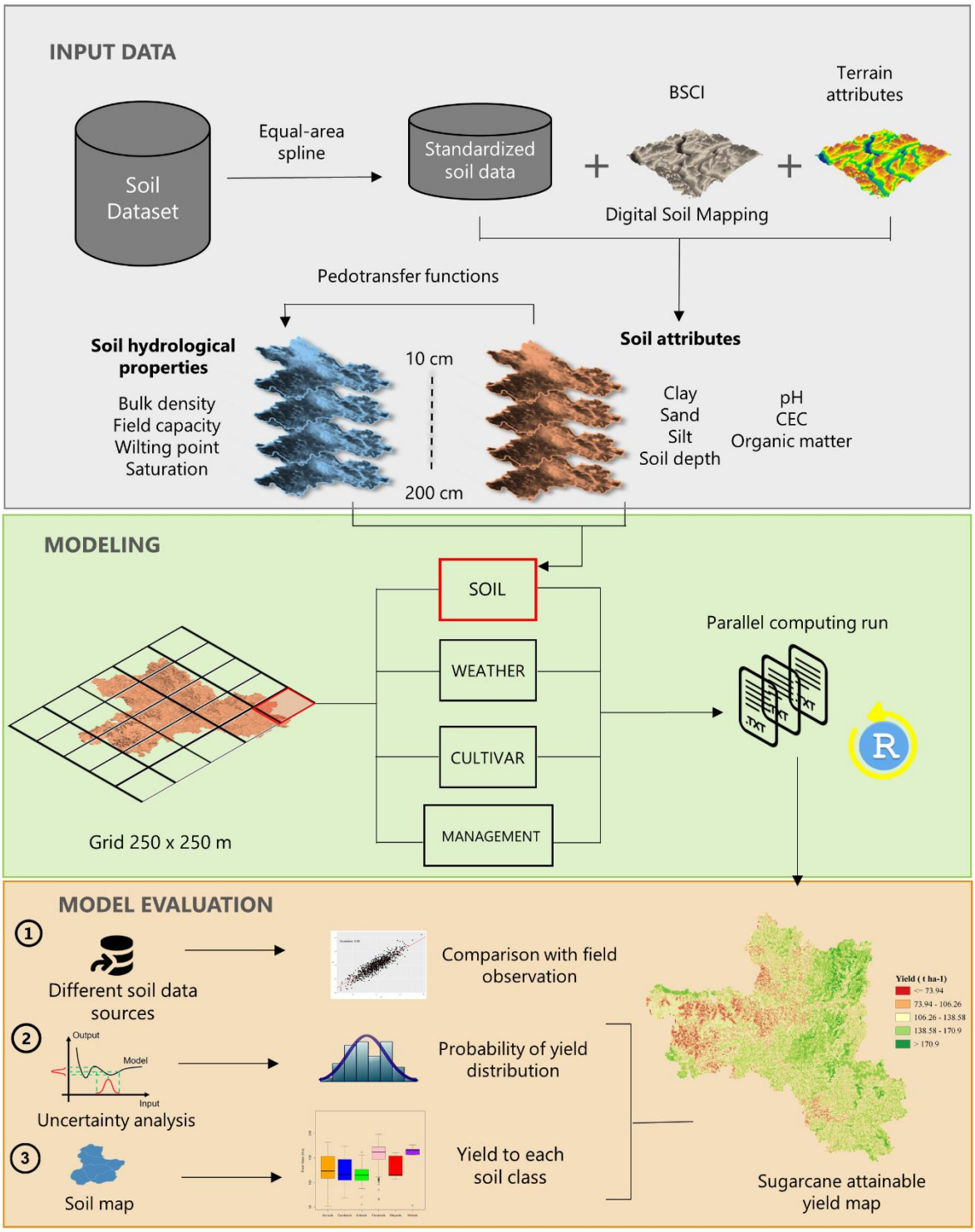


Figure 2. Flowchart representing the dataset and the approach adopted in this study.

2.4. The DSSAT/CANEGRO model

The DSSAT/CANEGRO is a crop model for simulating daily sugarcane growth, development and yield. The minimum input required to run the model are daily weather data, soil physical attributes, genotype, and crop management data (Jones et al., 2016). In order to enable a spatial approach, a code was created to run DSSAT on each individual cell from our raster grid (250 x 250 m). Therefore, our code had to perform the following steps: (i) Harmonize the spatially explicit input data in a standard raster grid; (ii) create txt files for every pixel of the grid, following the DSSAT specifications (running batch file mode); (iv) Execute the model in loop to every txt file; and (v) perform the back-conversion of DSSAT outputs to a spatial grid. The input manipulation was supported by the R package named "DSSAT", which provides cross-platform functions to read and write input files, run DSSAT and read output files (Alderman, 2020).

Sugarcane growth was simulated to the RB86-7515 variety, which is one of the most cultivated varieties in Brazil (Marin et al., 2013). The cultivar coefficients were adjusted for Brazilian conditions according to Dias and Sentelhas (2017). The DSSAT experimental file (SCX) considered common sugarcane management practices in simulations, such as planting density of 15 plants per meter, 150 cm row spacing and no irrigation. The simulations considered planting season in October and harvest period within 12 months. Weather data included solar radiation (SRAD), precipitation (RAIN), as well as maximum and minimum air temperature (TMAX and TMIN, respectively). Daily weather data were acquired from the automatic meteorological station from Luiz de Queiroz College of Agriculture (Figure 1), for the period between 1997 to 2015. This weather station has consistent information and adequately represents the regional weather. Once that our focus was to evaluate the influence of soil data in crop modeling, we did not use spatially explicit weather data. Adding spatial weather data could have great influence in predictions and compromise the evaluation regarding the soil datasets.

2.5. Impact of different soil data sources in yield predictions

To evaluate the impacts of different soil data sources in the sugarcane simulation by DSSAT/CANEGRO model, locations (points) were selected, from where independent soil profiles descriptions were available (Figure 1) (Batjes et al.,2009; Battie Laclau and Laclau, 2009; RadamBrasil, 1981). Soil data retrieved from these points were considered our reference

dataset, and they were all measured by laboratory methods, and have a small level of uncertainty. To each point, we performed the DSSAT simulation according to the conditions defined in Section 2.4. The process was repeated for different years (1997-2015) to avoid selecting only years with climatic adversities. The sugarcane stalk fresh mass, from daily output data, was used to define the differences between the reference estimate (predicted with reference soil data) and soil information acquired from different soil maps.

The first soil data source considered was a legacy map (LM) representing soil types with several class associations at a 1:100,000 scale from the Agronomic Institute of Campinas (Oliveira and Prado, 1987). In this case, only the predominant soil type was used in the analysis. Another data source was the 250 m soil attribute maps from the SoilGrids project (SG; Hengl et al., 2017). Soil attributes information was only extracted from the grid points corresponding to our reference profiles. The third source considered was the soil attribute maps (MA) from our regional ML algorithm.

Later, we performed an analysis to identify the influence of these different data sources on the prediction of sugarcane attainable yield. The model's performance was described by the R^2, RMSE, bias, index of agreement (d) and confidence coefficient (c) (Camargo and Sentelhas, 1997; Willmott et al., 1985). The index of agreement varies from 0 to 1, with higher index values indicating that the predicted values have better agreement with the observations. The index c evaluates the performance of the model, also varying from 0 to 1.0, where higher values indicate higher performance.

2.6. Influence of the soil attributes' uncertainties on the estimation of attainable sugarcane yield

Another important step corresponds to the definition of how uncertainties propagate from predicted soil attributes to the DSSAT/CANEGRO and consequently to yield predictions. A Monte Carlo Simulation (MCS) was conducted to evaluate if the uncertainty in soil attributes, predicted by ML, could have a considerable influence in CANEGRO predictions. The MCS randomly selected attribute values from a predetermined distribution, which has mean and standard deviations defined previously, during the DSM process. Simulation was repeated 100 times and at each repetition, the value of the attribute estimates was adjusted. Once that all soil attributes predictions were considered the same environmental covariates, and in some cases, such attributes are naturally correlated, uncertainties were assumed to be dependent.

Therefore, we addressed a joint multivariate normal distribution, which was performed with the "mc2d" package in R environment (Pouillot and Delignette-Muller, 2010).

Since MSC is a time-consuming process due to the number of repetitions and, added to this, the study area is extensive, representative points were selected to perform the MSC analysis. Therefore, we used the Latin Hypercube sampling method to indicate the number of samples that ensures full coverage of every parameter's range (Malone et al., 2019). Evaluation of the data was based on the coefficient of variation results from the output sets of 100 simulations, which resulted in a probability of distribution of sugarcane yield.

2.7. Sugarcane yield assessment

The comparison of yield with the soil classes derived from a detailed soil map allows the comparison with the PE. The PE through the soil classes defines the class that presents the higher productive potential. This approach considers soil texture, depth, water availability, soil fertility and evapotranspiration (Demattê and Demattê, 2009). The evaluation of yield by soil class is a parameter to assess the coherence of the model output with a methodology already used in sugarcane fields.

Therefore, spatially explicit simulation of attainable yield was performed from October of 2003 to September 2004 (i.e., with a 12 months' cycle), following the specifications in Section 2.4 (Figure 2). Thus, in order to evaluate the sugarcane yield by soil class we extracted from the map generated the yield and evapotranspiration to each soil borderline. Detailed pedological maps (1:20,000 scale) were delineated, following the traditional survey method (Vasquez et al., 2014). The soil classes were classified according to the IUSS Working Group (IUSS Working Group WRB, 2006). A Tukey's HSD (honest significant difference) test was performed to evaluate the variation in productivity for each soil class.

3. RESULTS

3.1. Spatial predictions of soil attributes and uncertainty

The soil attributes predicted by ML algorithm will be briefly described regarding their spatial distribution and uncertainties. Clay content maps presented good results for both surperficial and subsurficial layers (R^2= 0.70 and 0.59, RMSE= 88.87 and 141 g kg^{-1} and RPIQ= 1.62 and 1.06). Similarly, sand content also had good results, with R^2, RMSE and RPIQ values in surface corresponding to 0.66, 124 g kg^{-1}, 1.55, respectively. Prediction os sand content in subsuperficial layers had slighty lower performance, with R^2, RMSE and RPIQ values of 0.62, 200 g kg^{-1} and 1.14. Although silt content did not perform as good as the other elements, predictions were satisfactory (R^2=0.41 and 0.29, RMSE= 73.3 and 200 g kg^{-1} and RPIQ= 1.08 and 0.73). Chemical attributes had regular-unsatisfactory performance, especially in deeper layers. In this case, subsuperficial OM, CEC and pH presented R^2 of 0.42, 0.17 and 0.11, respectively. Soil depth had R^2 of 0. 12 and RMSE of 149.11 cm. Thus, our models were able to predict attributes more accurately in superficial layers than in deeper layers (Details in Table 1 from Appendix A).

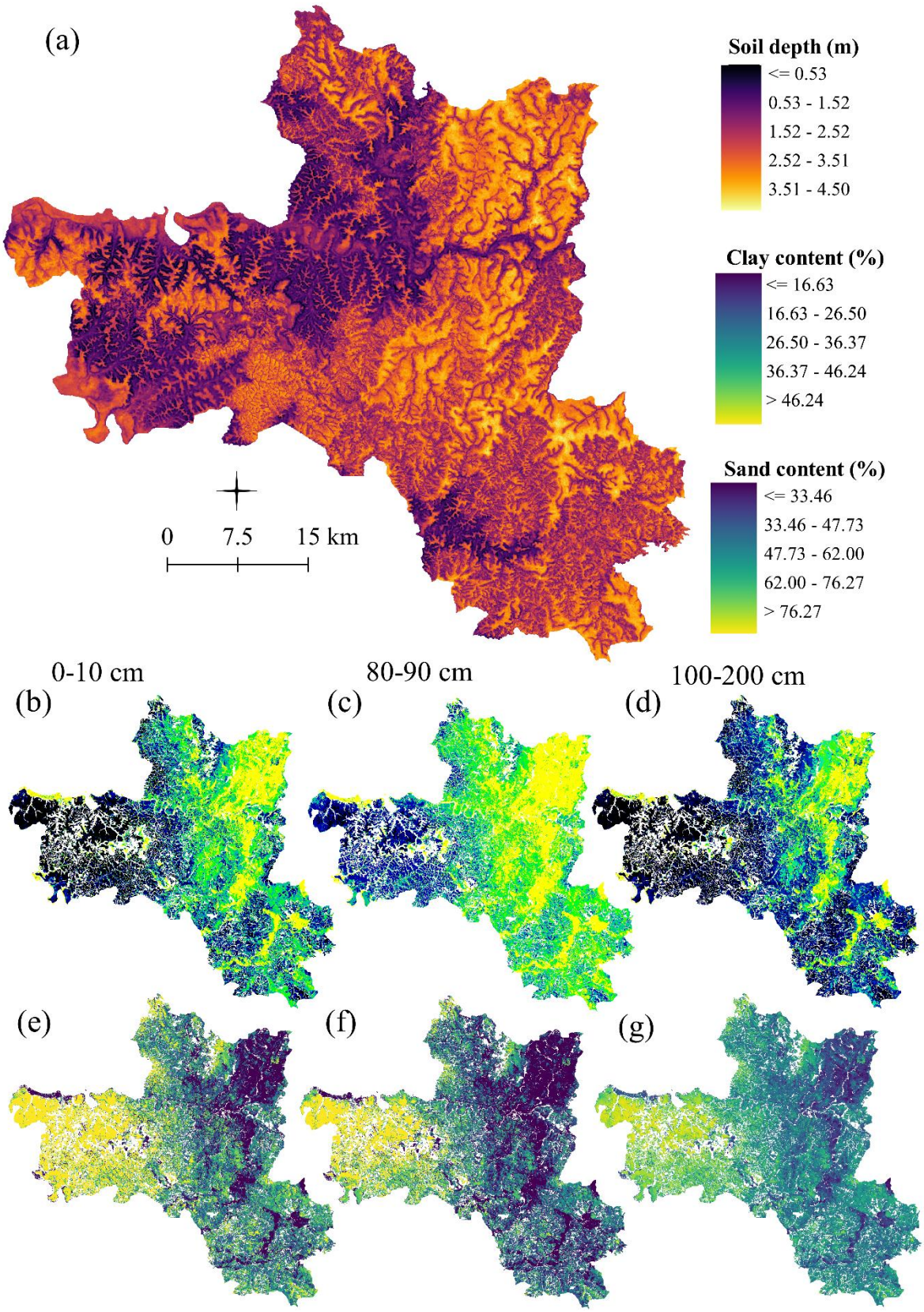


Figure 3. (a) Soil depth; Spatial distribution of predicted clay (b, c, d) and sand (e, f, g) contents in the studied region for 0-10, 80-90, and 150-200 cm soil layers.

The evaluation of DSM maps indicates a great spatial variability of soil attributes in the studied region (Figure 3). The soil depth map reinforces the variability that occurs in the region, varying from 0.50 to 4.50 meters (Figure 3a). Deep soils occur in the northwest of the study area, with average over 300 cm. Clay predictions had high variation, ranging from 9 to 38 % in topsoil's and mean value of 22% in deeper layers (Figure 3b, c, d). Clay was the predominant fraction in the surface layer, with increasing values from eastern to western regions (Figure 3b). In the site's northwestern region, there was a slightly higher clay content at 60-80 cm (Figure 3c), with mean values of 44% in the northwest. As expected, sand content was inversely related

to clay (Figure 3.e, f, g), with a gradual decrease from west to east. Topsoil's sand content had an average variation of 44 to 84%, with decreasing mean values in deeper layers.

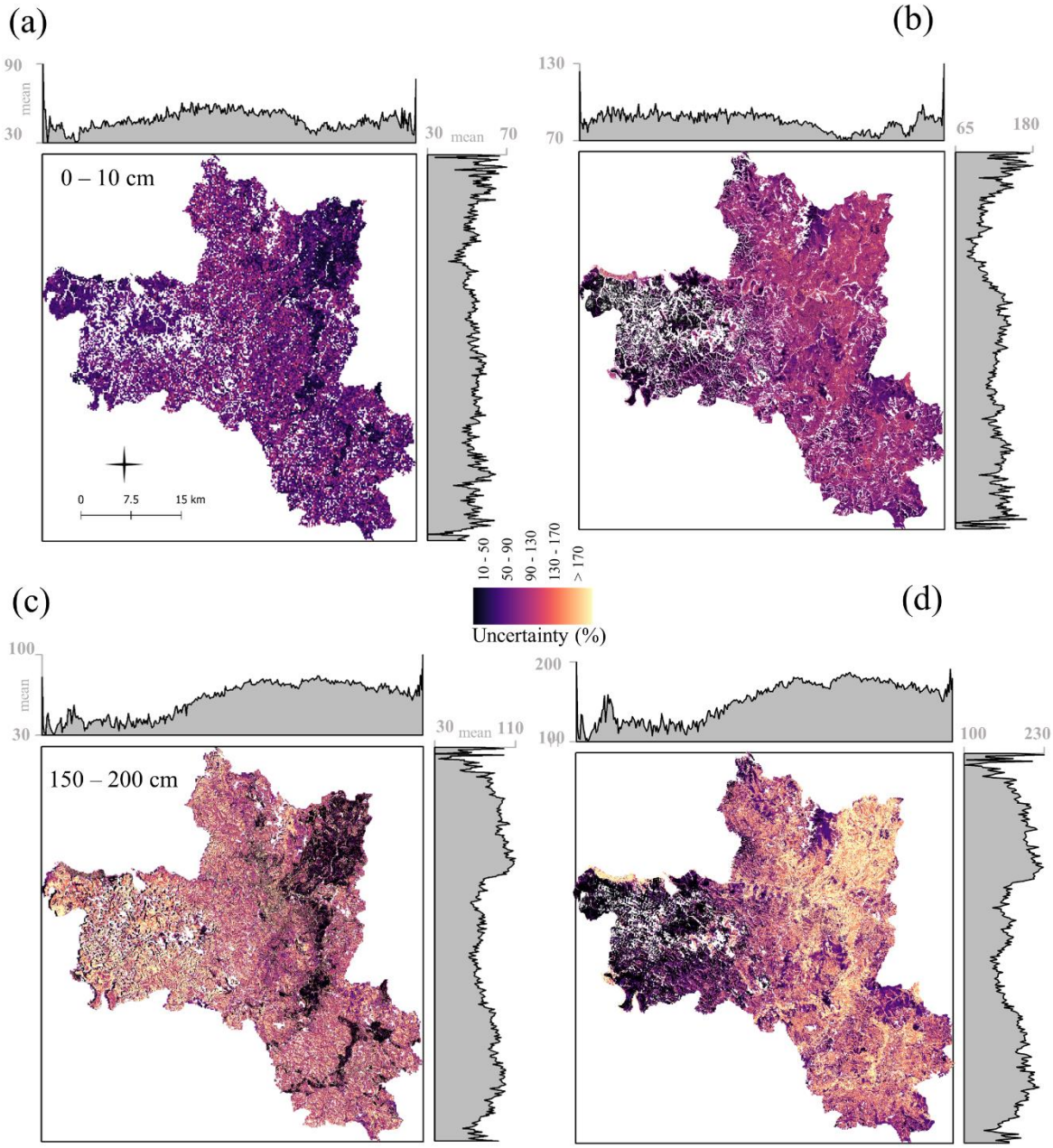


Figure 4. Clay and sand uncertainty maps of 0 - 10 cm with the mean value in latitude/longitude (a, b); (c, d) clay and sand uncertainty maps at 150-200 cm depth.

Uncertainty maps allow the identification of locations with variable prediction accuracies. In this study, clay uncertainty varied between 10% to 90% upper layers (0- 10 cm), with the highest values located in the western part of the studied region (Figure 4a). Sand uncertainty varied between 40 and 110% (Figure 4b), with an increasing spatial trend from west to east. Thus, it was observed that uncertainty was strictly related to the content of the mapped

attribute, indicating that areas with high clay and sand content presented lower uncertainties (Figure 3a, d). Besides that, decreasing values of soil attributes in depth are related to increasing uncertainties, ranging from 100 to 230%. Such results indicate that the prediction uncertainties of soil attributes tend to be greater in deeper layers (Figure 4c, d).

3.2. Site-specific evaluation of DSSAT/CANEGRO: Influence of soil data sources in yield predictions

The prediction of sugarcane attainable yield was impacted by the soil data source. Through the coefficient of determination (R^2) it was possible to identify the degree of proportion between cane stalk fresh mass estimated by DSSAT/CANEGRO model when using soil data from an observed profile and different alternative sources (SA, LM, and SG). From these analyses, the LM soil data was the one that resulted in the worst performance when comparing the yields estimated with observed and alternative soil data, with R^2 closer to 0.7 (Table 1) As a consequence of the poor correlation, the LM soil data source had the highest RMSE and Bias, ranging from 33.50 to 44.43 t ha^{-1}, and -21.91 to 23.82 t ha^{-1}, respectively. As expected, this data source had the lowest performance indices, with d between 0.33 and 0.41 and c between 0.22 and 0.31. The best cane yield estimates were obtained by using the MA soil data, which resulted in the highest R^2 (0.90 to 0.96), lowest RMSE (22.85 to 35.38 t ha^{-1}), lowest Bias (-6.31 to 5.8 t ha^{-1}), and highest performance indices, with d between 0.7 and 0,8 and c between 0.63 and 0.74. The use of SG data source remained in an intermediary position but with poor final performance when analyzing d and c indices for all assessed conditions.

Table 1. Comparison between soil data sources (MA, SG and LM) used as inputs for sugarcane yield simulation.

Independent points	Soil Source	R^2	RMSE (t ha^{-1})	Bias (t ha^{-1})	d	c
Brasil (1981)	MA[1]	0.90	22.85	4.25	0.80	0.72
	LM[2]	0.71	36.59	-21.91	0.41	0.29
	SG[3]	0.89	23.72	17.7	0.55	0.48
Battie Laclau and Laclau (2009)	MA[1]	0.96	35.38	-6.31	0.77	0.74
	LM[2]	0.68	44.43	23.82	0.33	0.22
	SG[3]	0.72	40.52	15.23	0.38	0.27
Batjes et al. (2009)	MA[1]	0.91	22.90	5.31	0.70	0.63

LM[2]	0.85	33.50	-19.11	0.37	0.31
SG[3]	0.90	38.12	-10.75	0.34	0.31

[1]Map of attributes derived from ML; [2]Legacy maps; [3] SoilGrids.

3.3. MCS

The soil parameters produced by DSM resulted in uncertainty that was assumed to influence the estimates of soil water parameters and ultimately the sugarcane attainable yield. The MSC resulted in a coefficient of variation (CV) from 0.2 to 15 (%) (Figure 5a). Repeated simulations showed that sandy soils had a distribution from 30 to 70 t ha^{-1} (Figure 5b), as a consequence, the coefficient of variation of 15%. On the other hand, the variation of clayey soils is considerably smaller when compared to sand's, ranged from 116 to 120 t ha^{-1}, with CV closer to 1%. In this sense, it was observed a higher susceptibility of sandy soils as a function of the uncertainties of the soil attributes, since the predicted sand maps presented a higher uncertainty (Figure 3).

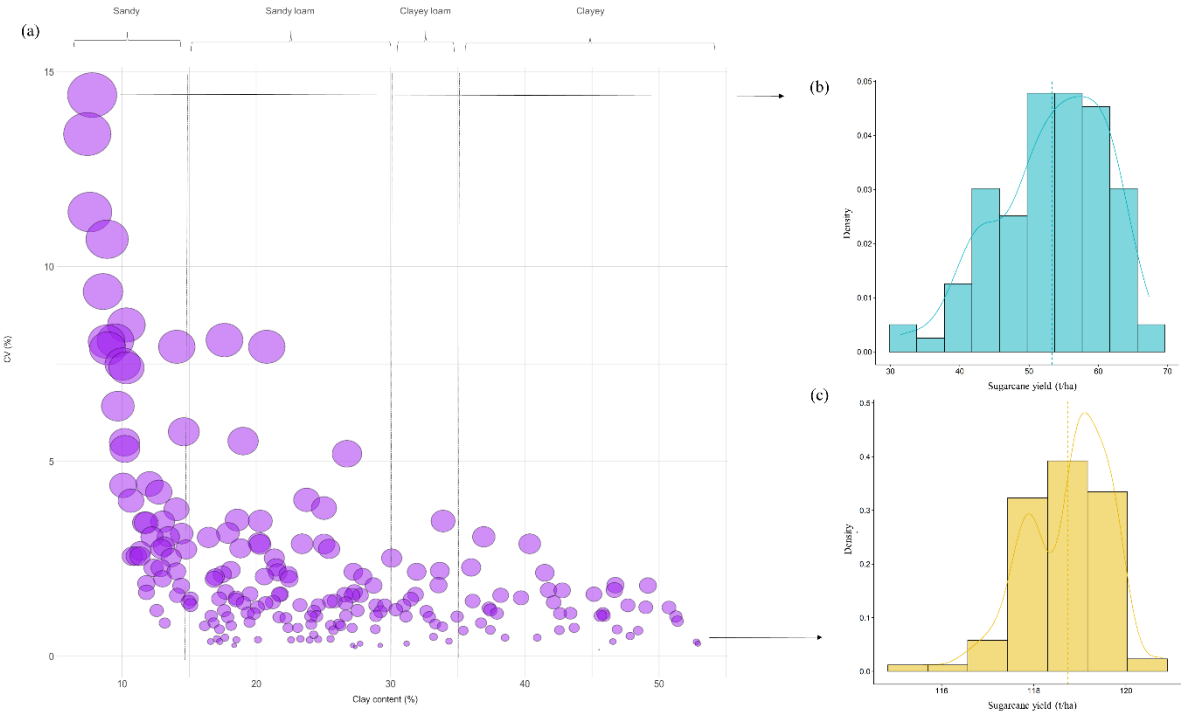


Figure 5. (a) Coefficient of variation (%) for points using the MSC depending on clay content; distribution of sugarcane attainable yield for the MSC simulations considered a sandy (a) and clayey soil (c).

3.4. Spatially explicit simulation of sugarcane attainable yield

A wide spatial variability of sugarcane attainable yield was observed in the studied area, being strictly related to the particle size distribution and soil depth of the study area. Therefore, observing soil depth and sugarcane attainable yield maps together, it is clear that areas with lower yields are those with shallow soils (Figure 3a and Figure 6a). Estimated sugarcane attainable yields ranged from 73.94 to 170.90 t ha^{-1} (Figure 6a). Higher values were reached in the north of the study area, corresponding to deeper soils (Figure 6a) and with

higher clay content (Figure 3). On the other hand, sandy-textured areas (located at northwest) had the lowest yields, never exceeding 100 t ha^{-1} (Figure 6a).

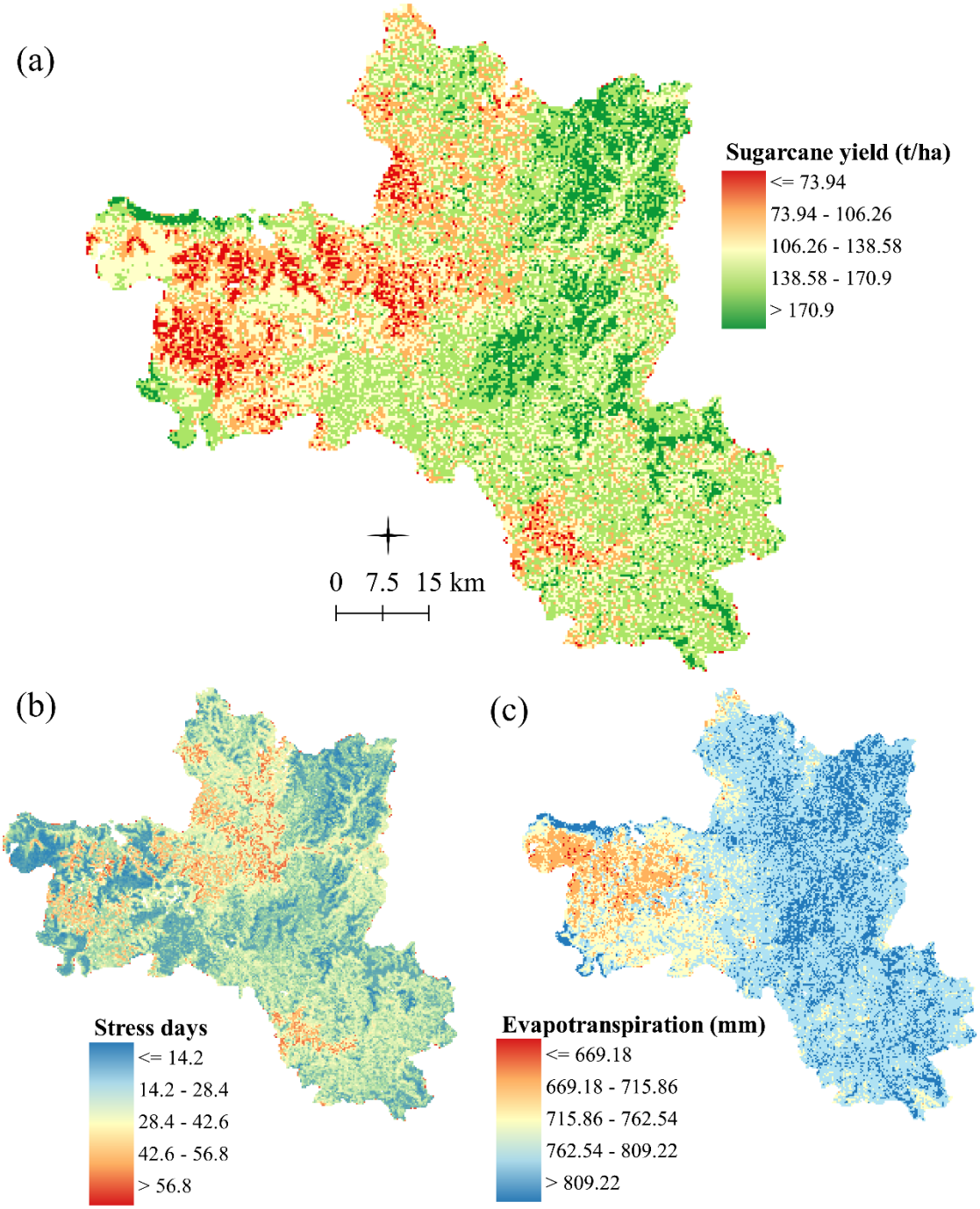


Figure 6. Sugarcane attainable yield estimated by DSSAT/CANEGRO with soil data from ML database (a); total days for the simulated period with water stress for sugarcane growth (b); and total actual evapotranspiration throughout the sugarcane cycle (c).

The map in Figure 6b shows the number of days that the sugarcane crop suffered with water restrictions, which affected its development. The areas that most suffered with this kind of stress were those where the soil is shallower (Figure 6c). Thus, through the actual

evapotranspiration map (Figure 6c), the impact of water stress on sugarcane yield is reinforced. Actual evapotranspiration represents the water consumed by the crop according to the weather conditions and soil water availability. Higher actual evapotranspiration values accumulated throughout the sugarcane cycle indicate that the crop suffered less stress, once considering that the maximum crop evapotranspiration was the same for the entire region. These results emphasize the connection between soil condition and profile, water balance and the crop, which in summary is expressed by the yield. The highest actual evapotranspiration values were observed over the east portion of the studied area, where the deeper soils predominate (Figure 6c).

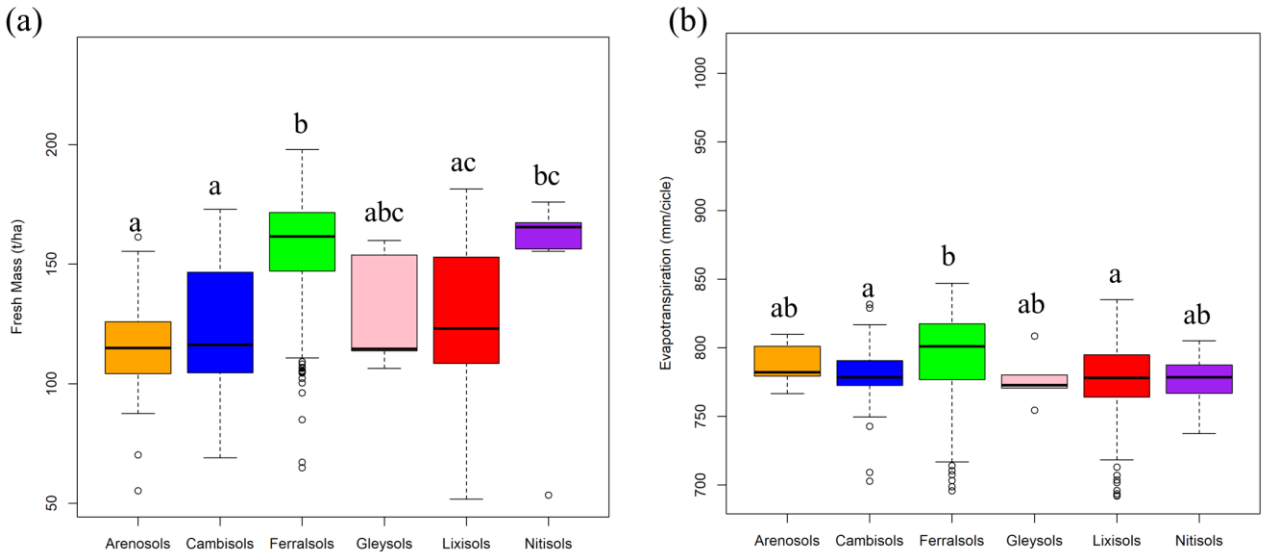


Figure 7. Boxplot considering the soil classes with the maps of DSSAT/CANEGRO output files, (a) Sugarcane yield by soil classes; (b) total evapotranspiration.

Comparing the sugarcane yield estimated for each soil class present in the studied area (Figure 7), interesting relationships were observed. Nitisols allowed to have higher attainable yields and with smaller interannual variability than for the other soils (Figure 7a). Nitisols are fairly deep, clayey and present good drainage, which are characteristics that favor sugarcane growth and development even under short to medium periods without rainfall. Nevertheless, considering the group comparison test, no differences of sugarcane yields were observed between Nitisols and Ferralsols (Figure 7a). These soil classes reached the highest yields. Shallow soils, such as Arenosols and Cambisols, were statistically different from the other soil classes, presenting the lowest sugarcane yields. Gleysols were the soils with the

greatest yield interannual variability. Regarding the actual evapotranspiration, the simulations for the Lixisols and Cambisols resulted in similar values, which in fact did not differ from actual evapotranspiration simulated with Arenosols, Gleysols and Nitisols. The simulations with Ferralsols resulted in the highest variability for actual evapotranspiration, which was statistically different from the values estimated with the other soil classes (Figure 7b).

4. DISCUSSION

4.1. Performance of digital soil mapping

The attribute maps derived from the digital mapping presented a satisfactory performance to explain the variability of the region. Clay was the attribute that obtained the best performance using GBT, obtaining results similar to studies previously conducted in the same study area (Fongaro et al., 2018; Gallo et al., 2018; Mendes et al., 2019; Rizzo et al., 2020). On the other hand, soil chemical attributes (e.g., OM, CEC and pH) had unsatisfactory performance (Table 1 from Appendix A). These elements are very dynamic throughout the soil profile and undergo changes for agricultural purposes, mainly in the top layer. As the BSMI is a composite image considering time-series images, the quality of the mapping is influenced by the dynamics of these elements over time (Chen et al., 2008; Miller and Kissel, 2010). Thus, the period that occurred the soil sampling differs from the satellite reflectance image and therefore is recommended to use a single image (Angelopoulou et al., 2019; Sulaeman et al., 2012). Because of that difference, adding terrain variables in digital mapping overcomes the limitation of using a BSCI (McBratney et al., 2003). Chemical attributes in depth are less affected by agricultural changes and may perform better at DSM. Although the influence of the time-series in the modeling, our results were similar to previous studies (Hengl et al., 2017; Mendes et al., 2019).

Uncertainty maps reveal that where a higher concentration of the attribute occurs the uncertainty tends to be lower (Figure 4). Malone et al. (2011) assessed how the uncertainties of soil attributes changed spatiality and along with the soil profile to DSM applications. The authors report two main sources of uncertainty, one related to sample density and the second to relief. The first is related to the distribution of the soil samples in the database, the samples must be able to capture the variations of the attributes at the site and also through the profile. The authors report that a larger variation of attributes values leads to higher predictive errors. Thus, the higher content of the mapped attribute to the location, the lower the values of uncertainty. The decrease in data with depth may reflect an increase of uncertainty in deeper layers. Besides the sample density, the relief also impacts the uncertainty in the attribute's prediction. Areas with irregular topography, causing depression lines or places with high slopes, such as to the northeast and southwest in our study area (Figure 1), presents increasing values

of uncertainty (Stumpf et al., 2017). The variation of attributes in these locations occurs more dynamically, which makes the performance of the model difficult.

According to the results presented in this study, the ML algorithm was fairly capable of capturing soil variability. Our results reinforce the potential of DSM to map soil attributes in medium to large areas, especially in countries such as Brazil, where soil data is scarce and imposes a challenge to soil mapping processes (Nolasco de Carvalho et al., 2015).

4.2. Simulated sugarcane yield

The sugarcane model employed in the present study was calibrated and validated by Dias and Sentelhas (2017) using Brazilian cultivars, with an error of less than 15 t ha^{-1} for estimated stalk fresh mass. Thus, the model was capable of representing the sugarcane crops for the studied area. Therefore, considering all model parameters and inputs as constant, except for the soil input data, the yield variation caused by different soil classes was estimated. The use of legacy soil map at the 1:1,00.000 scale with several associations resulted in the most erroneous and inconsistent source of data (Table 1). On the other hand, SoilGrids proved to be a satisfactory source of data, despite the fact that our predicted maps were the most consistent to estimate sugarcane yield (Table 1).

Other studies compared the impact of the spatial resolution of the soil data source in yield prediction. Guo et al. (2018) evaluated for the WheatGrow model what factors would impact the wheat yield. For this, it was used two soil data sources, a soil map from Germany (1:1,00.000 scale) and a soil grid (100 x 100 m). The author used a crop growth model to simulate wheat yield at the spatial scale based on partitioned zones and these were divided according to their soil properties. Simulation zone partitioning is usually employ to divide an area in similar zones with similar characteristics of environment and crop growth. Thus, the author reported that the grid soil data obtained a fundamental factor to explain the variability of production. Tewes et al. (2020) used the SIMPLACE (Scientific Impact Assessment and Modelling Platform for Advanced Crop and Ecosystem Management) model to evaluated the contribution of SoilGrid and field data sources in wheat productivity prediction. In a similar approach, Wimalasiri et al. (2020) used data from SoilGrid and OpenMap to estimate rice productivity in APSIM (Agricultural Production Systems Simulator) model. For both studies SoilGrid proved to be the soil data source that most resembled field data, proving to be a reliable source. For both studies SoilGrid proved to be the soil data source that most closely

resembled field data, proving to be a reliable and open source for use in crop modeling. However, the authors observed an inaccurate estimating of particle size distribution, mainly in deeper layers, reaching uncertainty above 250% at 100 cm. The SoilGrid could be combined with a regional dataset, instead of just using global data in order to improve the accuracy of the model and consequently decrease uncertainty (Hengl et al., 2017; Tifafi et al., 2018). In our study area, we used a local database, and consequently, we had lower uncertainty in soil data than SG (Table 1). However, our model can still be improved, especially in deeper layers (Figure 4).

As a consequence of uncertainty in soil attributes, the pedotransfer functions which use these inputs were affected, impacting the calculation of water availability (Botula et al. 2012). Our results indicate the impact of DSM on the water content available in the soil, since the error in the modeling presents positive values, indicating an overestimation, which is agreement with Ryczek et al. (2017) (Table 1). Moreover, we did not compare different pedotransfer equations on soil hydrological properties, thus this choice could also impact on yield prediction.

Furthermore, although the MA reached similar results when compared to field data, there are some limitations and opportunities for improvement in DSM, since their uncertainties impacted the prediction of sugarcane yield (Figure 6). An aspect that must be faced in DSM is to obtain the soil depth, which is controlled by complex interactions between soil properties and crop root system characteristics. Therefore, more studies are required for selecting the best variables to determine the soil depth (Dharumarajan et al., 2020; Tesfa et al., 2009). In our case, the highest error was obtained when the estimates were compared to the experimental data from Battie Laclau and Laclau (2009) (Table 1). In the condition of this experiment, these authors observed that sugarcane roots reached a depth of 3 m into the soil profile, as also reported by Inman-Bamber and Smith (2005), which resulted in very high yields, since little water deficit was observed along the crop cycle. On the other hand, the simulations with our approach were conducted for a soil with roots reaching 2 m in the maximum, resulting in lower yields than simulated with observed soil data (Table 1).

4.3. Spatial sugarcane attainable yield

The approach used in the present study allowed to map sugarcane attainable yield in a 250 m resolution. This map makes possible to identify the areas in the region with the highest

potential for cane production, as well as to choose the most appropriate strategies for improving yields in the area under restrictions (Singels et al. 2010). The results presented here show that the main limiting factor for sugarcane in the assessed region is soil water availability, which is influenced by soil water holding capacity, as well as by genotype drought tolerance, rainfall, and crop evapotranspiration, which is basically controlled by weather conditions (Van Ittersum et al., 2013). In this context, soil plays an important role to define attainable sugarcane yield, since the soil water holding capacity in function of soil texture and structure (Tomasella et al., 2000; Folberth et al., 2016). Another aspect to be considered in the present analysis, is that DSSAT/CANEGRO does not consider soil nutrients availability, making only soil physical properties influence cane yield (Jones et al.,2016). Considering that, our results were in line with what was expected, with a positive relationship between cane yields and clay content (Figure 6a).

Concerning soil classes, each one has a variation of attributes, and consequently of sugarcane yield (Figure 7). These results were in agreement with those from Demattê and Demattê (2009), which reported that the soil was one of the pillars to assist in sugarcane yield zoning. Deeper soils, with clay content varying from medium to clay texture (clay content over 30%) reached yields greater than 140 t ha^{-1}. Nitisols and Ferralsols soil allowed sugarcane yield to reach more than 130 t ha^{-1} (Figure 7a), highlighting the concordance of results with the PEs. On the other hand, Landell et al. (2003) also reported that soils with low water holding capacity (Entisols and Cambisols) had cane yields below 70 t ha^{-1}, which was evidenced in the map presented here (Figure 6a and 7a).

The variation within soil classes is also an important aspect to be considered in crop modeling since the models require soils attributes in the profile, making only soil class maps less useful than detailed soil attributes per soil layer (Varella et al., 2010). Especially to sugarcane cultivars management, soil attributes, such as clay and OM, are important to classify the areas suitable for genotypes with different drought tolerance (Landell et al., 2003). In summary, clayey soils with high OM present a higher yield potential than sandy ones. OM affects soil aggregation, impacting soil physical attributes, such as BD and water availability, which in turn define yield level (Carvalho et al., 2017). As addressed by Sanches et al. (2019), which conducted experiments in sugarcane fields in Brazil, clay, OM, CEC are the main soil attributes that most affect sugarcane yield.

Hence, the attainable yield map (Figure 6a) reinforces what studies report on sugarcane production for the Piracicaba region. As indicated by the study conducted by Dias and Sentelhas (2018), the main cause of the yield gap in the state of São Paulo sugarcane fields is the water stress that normally occurs between April and September. Thus, the potential and attainable cane yields have been used for studies related to the potential of irrigation as the water deficit yield gap can indicate the suitable areas for higher yields using this technique (Pènè et al. 2012). In fact, the regions with the lowest evapotranspiration are those with the highest values of stress days. Thus, the results highlight the importance of delimited strategies to the sugarcane sector in order to reduce the water stress and consequently achieved production.

5. CONCLUSIONS

The present study developed a reliable soil data source through DSM that can be successfully used in crop modeling. Among the sources of soil data at different levels the MA has proved to be the most reliable for yield estimation by crop simulation models. The implications of this result reinforces the holistic vision of DSM, which goes beyond obtaining soil information, where it can be applied to crop production (i.e. food and fiber) and enhances the applicability of soil data.

Therefore, the approach presented here, with a yield estimation in a 250 m grid, enabled us to investigate how soil-related aspects influence the attainable yield of sugarcane and, probably, other crops, being the texture and depth of the soil key attributes for that. Finally, the spatial and, also, temporal vision that the proposed approach allows to use that for crop planning and decision making, improving the sustainability of sugarcane production.

REFERENCES

Alderman, P.D., 2020. A comprehensive R interface for the DSSAT Cropping Systems Model. Comput. Electron. Agric. 172, 105325. https://doi.org/10.1016/j.compag.2020.105325

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

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D., 2019. Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review. Remote Sens. 11, 676. https://doi.org/10.3390/rs11060676

Antle, J.M., Basso, B., Conant, R.T., Godfray, H.C.J., Jones, J.W., Herrero, M., Howitt, R.E., Keating, B.A., Munoz-Carpena, R., Rosenzweig, C., Tittonell, P., Wheeler, T.R., 2017. Towards a new generation of agricultural system data, models and knowledge products: Design and improvement. Agric. Syst. 155, 255–268. https://doi.org/10.1016/j.agsy.2016.10.002

Baigorria, G., Jones, J., Shin, D., Mishra, A., O'Brien, J., 2007. Assessing uncertainties in crop model simulations using daily bias-corrected Regional Circulation Model outputs. Clim. Res. 34, 211–222. https://doi.org/10.3354/cr00703

Batjes, N.H., 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use Manag. 25, 124–127. https://doi.org/10.1111/j.1475-2743.2009.00202.x

Basso, B., Ritchie, J.T., Pierce, F.J., Braga, R.P., Jones, J.W., 2001. Spatial validation of crop models for precision agriculture. Agric. Syst. 68, 97–112. https://doi.org/10.1016/S0308-521X(00)00063-9

Battie Laclau, P., Laclau, J.P., 2009. Growth of the whole root system for a plant crop of sugarcane under rainfed and irrigated environments in Brazil. F. Crop. Res. 114, 351–360. https://doi.org/10.1016/j.fcr.2009.09.004

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

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. https://doi.org/10.1007/s13593-018-0490-x

Brasil. Ministério das Minas e Energia. Secretaria Geral. Projeto RADAMBRASIL. Folha SD.22-Goiás, SD.23- Brasília, SE.22-Goiânia. (Levantamento de Recursos Naturais, 25,29,31). Rio de Janeiro.1981.

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

Camargo, A.P, Sentelhas, P.C., 1997. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, 5, Rev. Bras. Agrometeorol, Brasil (1997), pp. 89-97

Camargo, O.A, Moniz, A.C., Jorge, J.A., Valadares J.M.A.S. Métodos de Análise Química Mineralógica e Física de Solos do Instituto Agronômico de Campinas, Campinas (2009).

Carvalho, J.L.N., Nogueirol, R.C., Menandro, L.M.S., Bordonal, R. de O., Borges, C.D., Cantarella, H., Franco, H.C.J., 2017. Agronomic and environmental implications of sugarcane straw removal: a major review. GCB Bioenergy. https://doi.org/10.1111/gcbb.12410

Chen, F., Kissel, D.E., West, L.T., Adkins, W., Rickman, D., Luvall, J.C., 2008. Mapping Soil Organic Carbon Concentration for Multiple Fields with Image Similarity Analysis. Soil Sci. Soc. Am. J. 72, 186–193. https://doi.org/10.2136/sssaj2007.0028

CONAB (2020). Acompanhamento da Safra Brasileira: Cana-de-açúcar. v.4 Safra 2020/19 n.3 - Terceiro levantamento. Dezembro 2020. Companhia Nacional de Abastecimento.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Demattê, J. L. I., & Demattê, J. A. M. (2009). Ambientes de produção como estratégia de manejo na cultura de cana-de-açúcar. Informações Agronômicas, (127), 10-18.

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

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

Dharumarajan, S., Vasundhara, R., Suputhra, A., Lalitha, M., Hegde, R., 2020. Prediction of Soil Depth in Karnataka Using Digital Soil Mapping Approach. J. Indian Soc. Remote Sens. 48, 1593–1600. https://doi.org/10.1007/s12524-020-01184-7

Dias, H.B., Sentelhas, P.C., 2018. Sugarcane yield gap analysis in Brazil – A multi-model approach for determining magnitudes and causes. Sci. Total Environ. 637–638, 1127–1136. https://doi.org/10.1016/j.scitotenv.2018.05.017

Dias, H.B., Sentelhas, P.C., 2017. Evaluation of three sugarcane simulation models and their ensemble for yield estimation in commercially managed fields. F. Crop. Res. 213, 174–185. https://doi.org/10.1016/j.fcr.2017.07.022

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L.B., Obersteiner, M., van der Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. Nat. Commun. 7, 11872. https://doi.org/10.1038/ncomms11872

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

Gallo, B., Demattê, J., Rizzo, R., Safanelli, J., Mendes, W., Lepsch, I., Sato, M., Romero, D., Lacerda, M., 2018. Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology. Remote Sens. 10, 1571. https://doi.org/10.3390/rs10101571

Gerland, P., Raftery, A.E., Sevčíková, 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, 234–7. https://doi.org/10.1126/science.1257469

Gijsman, A.J., Thornton, P.K., Hoogenboom, G., 2007. Using the WISE database to parameterize soil inputs for crop simulation models. Comput. Electron. Agric. 56, 85–100. https://doi.org/10.1016/j.compag.2007.01.001

Guo, C., Zhang, L., Zhou, X., Zhu, Y., Cao, W., Qiu, X., Cheng, T., Tian, Y., 2018. Integrating remote sensing information with crop model to monitor wheat growth and yield based on simulation zone partitioning. Precis. Agric. 19, 55–78. https://doi.org/10.1007/s11119-017-9498-5

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

Hengl, T., 2020. Global Soil Information Facilities [R package GSIF version 0.5-5.1]. Available online: https://CRAN.R-project.org/package=GSIF (accessed on 11 March 2019).

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

Hoogenboom, G., C.H. Porter, K.J. Boote, V. Shelia, P.W. Wilkens, U. Singh, J.W. White, S. Asseng, J.I. Lizaso, L.P. Moreno, W. Pavan, R. Ogoshi, L.A. Hunt, G.Y. Tsuji, and J.W. Jones. 2019. The DSSAT crop modeling ecosystem. In: p.173-216 [K.J. Boote, editor] Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, United Kingdom. (http://dx.doi.org/10.19103/AS.2019.0061.10).

IUSS Working Group, 2006. WRB World reference base for soil resources World Soil Resources Report, 103

Inman-Bamber, N.G., Smith, D.M., 2005. Water relations in sugarcane and response to water deficits, in: Field Crops Research. Elsevier, pp. 185–202. https://doi.org/10.1016/j.fcr.2005.01.023

Jones, J.., Hoogenboom, G., Porter, C.., Boote, K.., Batchelor, W.., Hunt, L.., Wilkens, P.., Singh, U., Gijsman, A.., Ritchie, J.., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. https://doi.org/10.1016/J.AGSY.2016.05.014

Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. J. Hydrol. 586, 124905. https://doi.org/10.1016/j.jhydrol.2020.124905

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

Lagacherie, P., Cazemier, D.R., Martin-Clouaire, R., Wassenaar, T., 2000. A spatial approach using imprecise soil data for modelling crop yields over vast areas. Agric. Ecosyst. Environ. 81, 5–16. https://doi.org/10.1016/S0167-8809(00)00164-X

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. Atributos químicos de subsuperfície de latossolos e produtividades da cana-de-açúcar. Sci. Agric. 60, 741–745. https://doi.org/10.1590/S0103-90162003000400020

Launay, M., Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agric. Ecosyst. Environ. 111, 321–339. https://doi.org/10.1016/j.agee.2005.06.005

Liang, Y., Yu, S., Zhen, Z., Zhao, Y., Deng, J., Jiang, W., 2020. Climatic change impacts on Chinese sugarcane planting: Benefits and risks. Phys. Chem. Earth 116, 102856.

https://doi.org/10.1016/j.pce.2020.102856

Machwitz, M., Hass, E., Junk, J., Udelhoven, T., Schlerf, M., 2019. CropGIS – A web application for the spatial and temporal visualization of past, present and future crop biomass development. Comput. Electron. Agric. 161, 185–193. https://doi.org/10.1016/J.COMPAG.2018.04.026

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

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152. https://doi.org/10.1016/j.geoderma.2009.10.007

Marin, F.R., Jones, J.W., Singels, A., Royce, F., Assad, E.D., Pellegrino, G.Q., Justino, F., 2013. Climate change impacts on sugarcane attainable yield in southern Brazil. Clim. Change 117, 227–239. https://doi.org/10.1007/s10584-012-0561-y

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

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

Mezzalira, S., 1965. Descrição geológica e geográfica das folhas de Piracicaba e São Carlos. São Paulo, Instituto Geográfico e Geológico, p. 37.

Miao, Y., Mulla, D.J., Batchelor, W.D., Paz, J.O., Robert, P.C., Wiebers, M., 2006. Evaluating Management Zone Optimal Nitrogen Rates with a Crop Growth Model. Agron. J. 98, 545–

553. https://doi.org/10.2134/agronj2005.0153

Miller, R.O., Kissel, D.E., 2010. Comparison of Soil pH Methods on Soils of North America. Soil Sci. Soc. Am. J. 74, 310–316. https://doi.org/10.2136/sssaj2008.0047

Nendel, C., Wieland, R., Mirschel, W., Specka, X., Guddat, C., Kersebaum, K.C., 2013. Simulating regional winter wheat yields using input data of different spatial resolution. F. Crop. Res. 145, 67–77. https://doi.org/10.1016/j.fcr.2013.02.014

Nolasco de Carvalho, C.C., Nunes, F.C., Homem Antunes, M.A., Nolasco, M.C., 2015. Soil Surveys in Brazil and Perspectives in Soil Digital Mapping. Soil Horizons 56, 0. https://doi.org/10.2136/sh14-01-0002

Oliveira, J.B., Prado, H., 1989. Carta Pedológica Semi-detalhada do Estado de São Paulo: Quadrícula de Piracicaba. Folha SF-23-Y-A-IV. Instituto Agronômico de Campinas, Campinas.

Ottoni, M.V., Ottoni Filho, T.B., Schaap, M.G., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2018. Hydrophysical Database for Brazilian Soils (HYBRAS) and Pedotransfer Functions for Water Retention. Vadose Zo. J. 17, 170095. https://doi.org/10.2136/vzj2017.05.0095

Pagani, V., Stella, T., Guarneri, T., Finotto, G., van den Berg, M., Marin, F.R., Acutis, M., Confalonieri, R., 2017. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. Agric. Syst. 154, 45–52. https://doi.org/10.1016/j.agsy.2017.03.002

Penè, C.B., S. N' Diaye, and C. N'Guessan-Konan. 2012. Sprinkler irrigation and soil tillage practices in sugarcane plantations as influence by soil texture and water storage in northern Ivory Coast. Journal of Appied Bioscience 54: 3916–3924

Python Software Foundation, 2018. Python Language Reference, Version 3.7. Available online: https://www.python.org/ (accessed on 11 March 2019).

R Core Team, 2020. R: A Language and Environment for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 01 March 2019).

Ritchie, J.T., Godwin, D.C., Singh, U., 1990. Soil and weather inputs for the IBSNAT crop models. International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) Project Proceedings of the IBSNAT Symposium: Decision Support System for Agrotechnology Transfer. Part I. Symposium Proceedings, Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, Hawaii, Las Vegas, NV, October 16–18, 1989

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

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

Ryczek, M., Kruk, E., Malec, M., Klatka, S., 2017. Comparison of pedotransfer functions for the determination of saturated hydraulic conductivity coefficient. Ochr. Sr. i Zasobow Nat. 28, 25–30. https://doi.org/10.1515/oszn-2017-0005

Sachin, G., Mohammed Ahamed, J., Nagajothi, K., Rana, M., Murugan, B.S., 2019. Automation of the DSSAT crop growth simulation model, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing, pp. 251–256. https://doi.org/10.5194/isprs-archives-XLII-3-W6-251-2019

Sadler, E.J., Gerwig, B.K., Evans, D.E., Busscher, W.J., Bauer, P.J., 2000. Site-specific modeling of corn yield in the SE coastal plain. Agric. Syst. 64, 189–207. https://doi.org/10.1016/S0308-521X(00)00022-6

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

Sanches, G.M., Graziano Magalhães, P.S., Junqueira Franco, H.C., 2019. Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. Geoderma 334, 90–98. https://doi.org/10.1016/j.geoderma.2018.07.051

Scarpare, F.V., Hernandes, T.A.D., Ruiz-Corrêa, S.T., Picoli, M.C.A., Scanlon, B.R., Chagas, M.F., Duft, D.G., Cardoso, T. de F., 2016. Sugarcane land use and water resources assessment in the expansion area in Brazil. J. Clean. Prod. 133, 1318–1327. https://doi.org/10.1016/j.jclepro.2016.06.074

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. Earth Environ. 27, 171–197. https://doi.org/10.1191/0309133303pp366ra

Sentelhas, P.C., Battisti, R., Câmara, G.M.S., Farias, J.R.B., Hampf, A.C., Nendel, C., 2015. The soybean yield gap in Brazil - Magnitude, causes and possible solutions for sustainable production. J. Agric. Sci. 153, 1394–1411. https://doi.org/10.1017/S0021859615000313

Shelia, V., Hansen, J., Sharda, V., Porter, C., Aggarwal, P., Wilkerson, C.J., Hoogenboom, G., 2019. A multi-scale and multi-model gridded framework for forecasting crop production, risk analysis, and climate change impact studies. Environ. Model. Softw. 115, 144–154.

https://doi.org/10.1016/J.ENVSOFT.2019.02.006

Singels, A., Bezuidenhout, C.N., 2002. A new method of simulating dry matter partitioning in the Canegro sugarcane model. F. Crop. Res. 78, 151–164. https://doi.org/10.1016/S0378-4290(02)00118-1

Singels, A., van den Berg, M., Smit, M.A., Jones, M.R., van Antwerpen, R., 2010. Modelling water uptake, growth and sucrose accumulation of sugarcane subjected to water stress. F. Crop. Res. 117, 59–69. https://doi.org/10.1016/j.fcr.2010.02.003

Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, S., Wadoux, A., Xiang, W., Scholten, T., 2017. Uncertainty-guided sampling to improve digital soil maps. Catena 153, 30–38. https://doi.org/10.1016/j.catena.2017.01.033

Sulaeman, Y., Sarwani, M., Minasny, B., McBratney, A.B., Sutandi, A., Barus, B., 2012. Soil-landscape models to predict soil pH variation in the Subang region of West Java, Indonesia, in: Digital Soil Assessments and Beyond - Proceedings of the Fifth Global Workshop on Digital Soil Mapping. Taylor and Francis - Balkema, pp. 317–323. https://doi.org/10.1201/b12728-63

Tesfa, T.K., Tarboton, D.G., Chandler, D.G., McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007474

Tewes, A., Hoffmann, H., Nolte, M., Krauss, G., Schäfer, F., Kerkhoff, C., Gaiser, T., 2020. How Do Methods Assimilating Sentinel-2-Derived LAI Combined with Two Different Sources of Soil Input Data Affect the Crop Model-Based Estimation of Wheat Biomass at Sub-Field Level? Remote Sens. 12, 925. https://doi.org/10.3390/rs12060925

Thorp, K.R., DeJonge, K.C., Kaleita, A.L., Batchelor, W.D., Paz, J.O., 2008. Methodology for the use of DSSAT models for precision agriculture decision support. Comput. Electron. Agric.

64, 276–285. https://doi.org/10.1016/j.compag.2008.05.022

Tifafi, M., Guenet, B., Hatté, C., 2018. Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on SoilGrids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data From USA, England, Wales, and France. Global Biogeochem. Cycles 32, 42–56. https://doi.org/10.1002/2017GB005678

Tomasella, J., Hodnett, M.G., Rossato, L., 2000. Pedotransfer Functions for the Estimation of Soil Water Retention in Brazilian Soils. Soil Sci. Soc. Am. J. 64, 327–338. https://doi.org/10.2136/sssaj2000.641327x

Van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance-A review. F. Crop. Res. 143, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

Vasques, G.M., Demattê, J.A.M., Viscarra Rossel, R.A., Ramírez-López, L., Terra, F.S., 2014. Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths. Geoderma 223–225, 73–78. https://doi.org/10.1016/j.geoderma.2014.01.019

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environ. Model. Softw. 25, 310–319. https://doi.org/10.1016/j.envsoft.2009.09.012

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. 90, 8995. https://doi.org/10.1029/jc090ic05p08995

Wimalasiri, E.M., Jahanshiri, E., Suhairi, T.A.S.T.M., Udayangani, H., Mapa, R.B., Karunaratne, A.S., Vidhanarachchi, L.P., Azam-Ali, S.N., 2020. Basic soil data requirements for process-based crop models as a basis for crop diversification. Sustain. 12, 7781. https://doi.org/10.3390/SU12187781

Zhao, G., Bryan, B.A., King, D., Luo, Z., Wang, E., Bende-Michl, U., Song, X., Yu, Q., 2013. Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. Environ. Model. Softw. 41, 231–238. https://doi.org/10.1016/j.envsoft.2012.08.007

APPENDIX

APPENDIX A.

Table 1. Performance evaluation for soil attributes prediction using ML algorithm.

Soil attribute[1]	Depth[2]	RMSE[3]	RPIQ[4]	[5]R²	FS[6]	nRP[7]	minSL[8]
Clay (g kg^{-1})	0-10	88.87	1.62	0.70	30	11	40
	0-20	86.3	1.67	0.69	200	11	50
	20-30	87.1	1.64	0.69	60	11	40
	30-40	88.1	1.67	0.67	100	13	100
	40-50	98.4	1.60	0.64	200	13	30
	50-60	95.2	1.65	0.65	200	13	40
	60-70	90.1	1.70	0.64	100	13	200
	70-80	96.5	1.57	0.63	30	11	20
	80-90	93.2	1.68	0.64	30	13	500
	90-100	95.1	1.65	0.63	30	5	500
	100-150	134	1.05	0.61	60	11	10
	150-200	141	1.06	0.59	60	11	11
Sand (g kg^{-1})	0-10	124	1.55	0.66	100	11	100
	0-20	124	1.54	0.67	200	8	50
	20-30	120	1.51	0.68	100	8	100
	30-40	127	1.46	0.64	200	3	100
	40-50	132.4	1.51	0.63	200	13	20
	50-60	125	1.58	0.65	100	11	50
	60-70	127.6	1.49	0.63	100	3	40
	70-80	121	1.56	0.65	200	13	50
	80-90	127.2	1.53	0.62	30	11	500
	90-100	127.2	1.53	0.62	30	11	500
	100-150	130.2	1.51	0.60	100	20	100
	150-200	255	1.14	0.62	200	8	10
Silt (g kg^{-1})	0-10	76.3	1.07	0.41	60	11	40
	0-20	73.3	1.08	0.41	30	11	100
	20-30	76.3	0.98	0.37	200	11	10
	30-40	70.5	1.07	0.39	60	5	50
	40-50	74.8	1.02	0.33	100	8	20
	50-60	71.9	1.05	0.35	200	5	40
	60-70	72.7	1.00	0.31	60	3	500
	70-80	70.3	1.04	0.33	60	11	30

	80-90	72.8	1.04	0.32	30	11	30
	90-100	75.1	1.02	0.31	60	3	200
	100-150	180.8	0.72	0.26	60	13	10
	150-200	172	0.73	0.29	200	3	20
	0-10	6.4	1.40	0.42	60	8	500
	0-20	5.9	1.37	0.41	200	13	40
	20-30	5.7	1.44	0.44	100	8	200
	30-40	4.6	1.40	0.43	200	8	100
	40-50	4.7	1.24	0.39	100	11	30
	50-60	4.3	1.27	0.38	100	5	100
OM (g kg^{-1})	60-70	4.1	1.21	0.35	30	8	30
	70-80	3.8	1.26	0.39	60	8	200
	80-90	3.9	1.04	0.36	60	11	500
	90-100	4.1	1.03	0.32	200	3	500
	100-150	4.8	1.17	0.34	200	3	10
	150-200	5.1	1.78	0.32	200	11	30
	0-10	34.2	0.81	0.17	100	3	30
	0-20	32.3	0.80	0.18	500	13	200
	20-30	31.1	0.74	0.13	50	8	200
	30-40	31.1	0.72	0.14	500	8	60
	40-50	32.1	0.68	0.12	200	13	60
CEC	50-60	32	0.65	0.11	200	8	200
(mmolc kg^{-1})	60-70	33	0.62	0.07	50	3	60
	70-80	40	0.60	0.09	100	5	100
	80-90	37	0.59	0.10	100	13	30
	90-100	35	0.58	0.07	50	3	30
	100-150	90	0.84	0.04	10	13	30
	150-200	140	0.82	0.04	10	11	60
	0-10	0.72	1.10	0.18	500	13	60
	0-20	0.67	1.70	0.17	500	13	60
	20-30	0.64	1.10	0.09	500	13	60
	30-40	0.64	1.10	0.04	500	13	60
	40-50	0.69	1.20	0.05	500	13	60
pH H2O	50-60	0.65	1.30	0.07	500	13	60
	60-70	0.60	1.30	0.08	500	13	60
	70-80	0.57	1.30	0.08	500	13	60
	80-90	0.58	1.40	0.08	500	13	30
	90-100	0.58	1.50	0.08	500	13	30
	100-150	1.40	0.76	0.05	10	5	60

	150-200	1.50	0.80	0.03	50	13	30

[1]Soil attributes: organic matter (OM), cation exchange capacity (CEC), soil pH determined in water solution (pH); [2]In centimeters (cm); [3]RMSE: root mean squared error; [4]RPIQ: ratio of performance to the interquartile range; [5]R^2: coefficient of determination; [6]FS: hyperparameter forest size; [7]nRP: hyperparameter number of random predictors; [8]minSL: hyperparameter minimum number of samples at leaves.

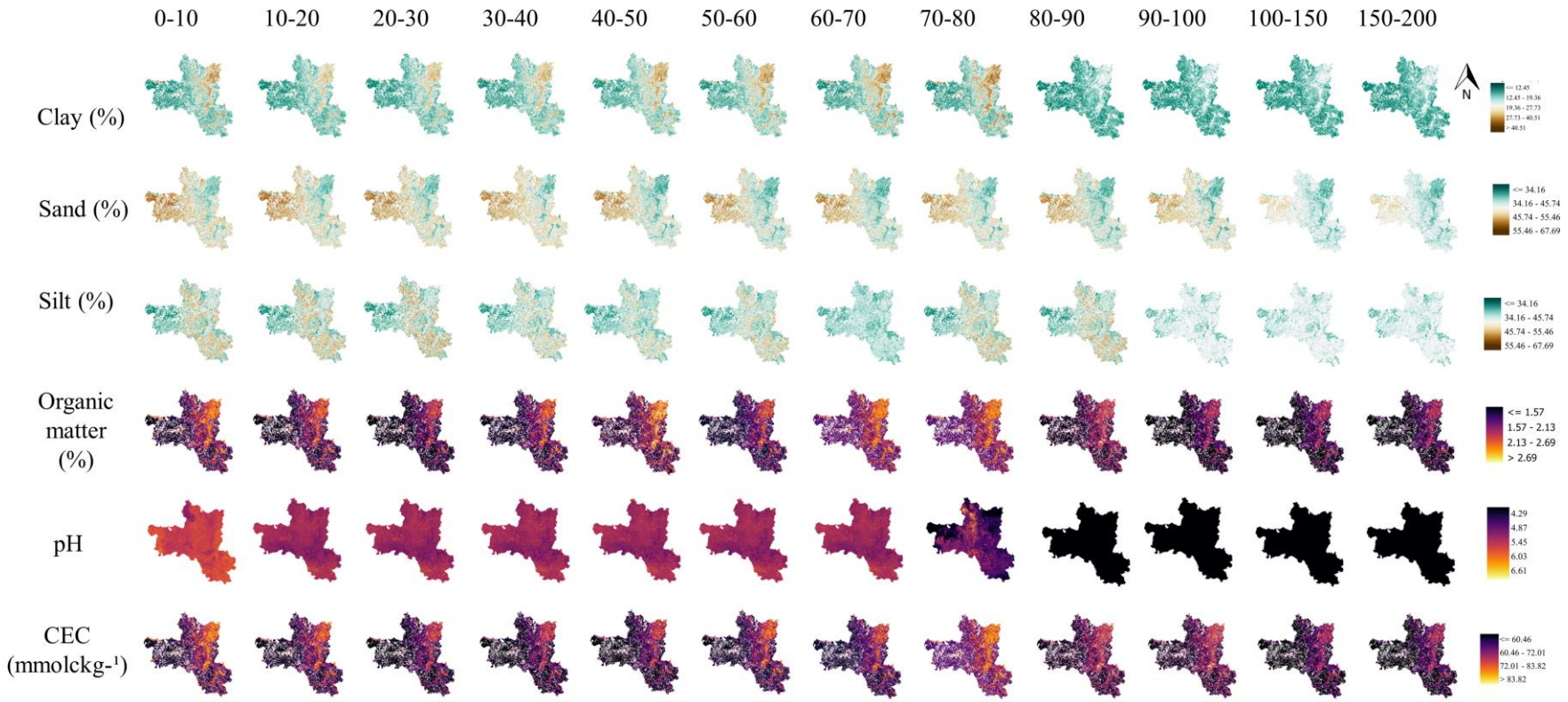
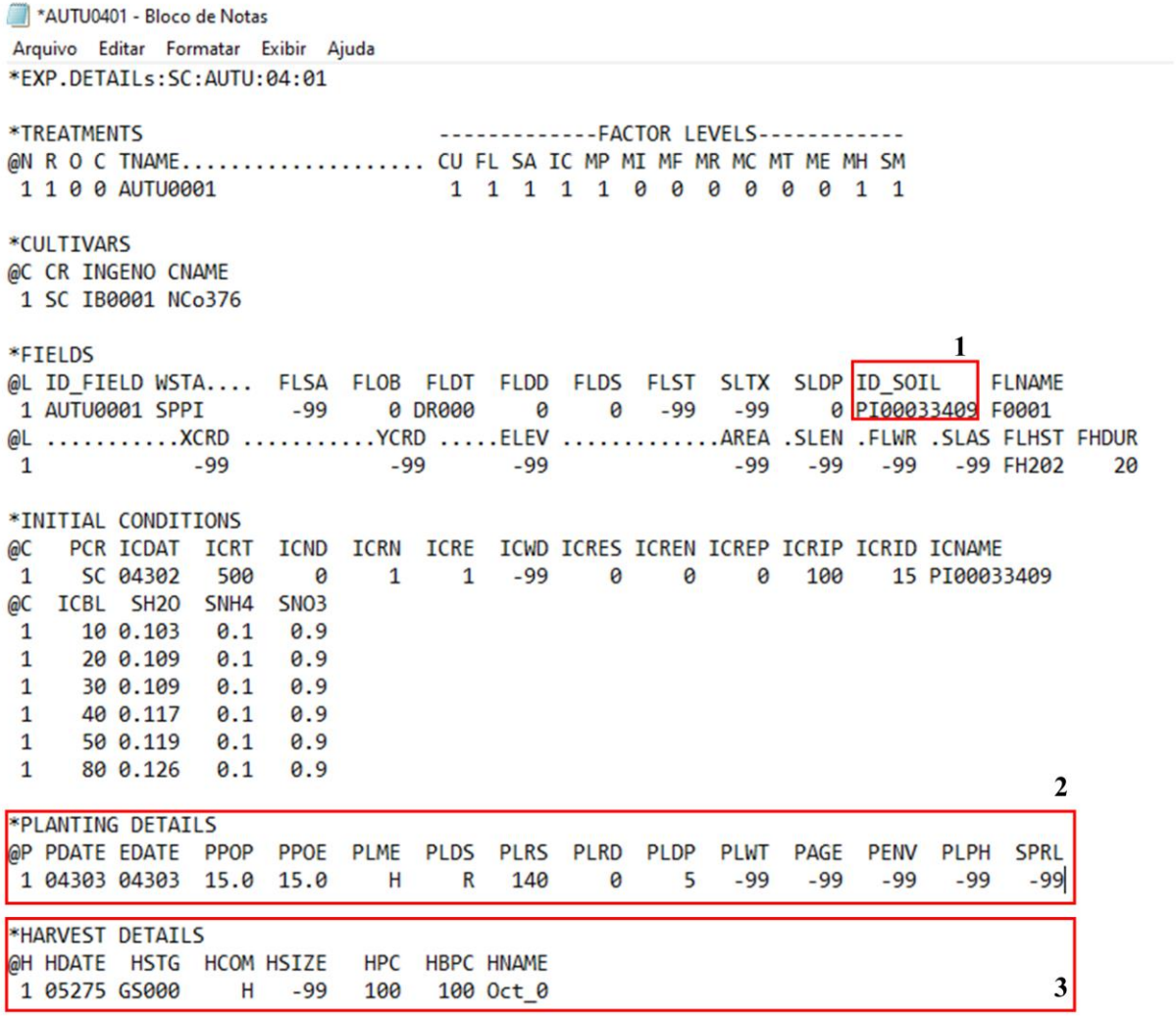


Figure 1. Maps of soil attributes at different depths where: clay, sand and silt content (%), organic matter (OM), pH determined in water solution (pH) and cation exchange capacity (CEC)



*AUTU0401 - Bloco de Notas

Arquivo Editar Formatar Exibir Ajuda

```
*EXP.DETAILs:SC:AUTU:04:01

*TREATMENTS                        -------------FACTOR LEVELS------------
@N R O C TNAME.................... CU FL SA IC MP MI MF MR MC MT ME MH SM
 1 1 0 0 AUTU0001                   1  1  1  1  1  0  0  0  0  0  0  1  1

*CULTIVARS
@C CR INGENO CNAME
 1 SC IB0001 NCo376

*FIELDS                                                                 1
@L ID_FIELD WSTA....  FLSA  FLOB  FLDT  FLDD  FLDS  FLST  SLTX  SLDP ID_SOIL    FLNAME
 1 AUTU0001 SPPI       -99     0 DR000     0     0   -99   -99     0 PI00033409 F0001
@L ...........XCRD ...........YCRD .....ELEV .............AREA .SLEN .FLWR .SLAS FLHST FHDUR
 1             -99             -99       -99               -99   -99   -99   -99 FH202    20

*INITIAL CONDITIONS
@C   PCR ICDAT  ICRT  ICND  ICRN  ICRE  ICWD ICRES ICREN ICREP ICRIP ICRID ICNAME
 1    SC 04302   500     0     1     1   -99     0     0     0   100    15 PI00033409
@C  ICBL  SH2O  SNH4  SNO3
 1    10 0.103   0.1   0.9
 1    20 0.109   0.1   0.9
 1    30 0.109   0.1   0.9
 1    40 0.117   0.1   0.9
 1    50 0.119   0.1   0.9
 1    80 0.126   0.1   0.9
                                                                        2
*PLANTING DETAILS
@P PDATE EDATE  PPOP  PPOE  PLME  PLDS  PLRS  PLRD  PLDP  PLWT  PAGE  PENV  PLPH  SPRL
 1 04303 04303  15.0  15.0     H     R   140     0     5   -99   -99   -99   -99   -99

*HARVEST DETAILS
@H HDATE  HSTG  HCOM HSIZE   HPC  HBPC HNAME
 1 05275 GS000     H   -99   100   100 Oct_0                                3
```

Figure 2. Example of the automatically generated experimental file (1) represents the grid that the simulation is being performed; (2) planting data and (3) harvest data.