A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes

Danilo César de Mello[1]; Gustavo Vieira Veloso[1]; Marcos Guedes de Lana[1]; Fellipe Alcantara de Oliveira Mello[2]; Raul Roberto Poppiel[2]; Diego Ribeiro Oquendo Cabrero[3]; Luis Augusto Di Loreto Di Raimo[4]; Carlos Ernesto Gonçalves Reynaud Schaefer[1]; Elpídio Inácio Fernandes-Filho[1]; Emilson Pereira Leite[4]; José Alexandre Melo Demattê[2]*

[1] Department of Soil Science, Federal University of Viçosa: daniloc.demello@gmail.com; gustavo.v.veloso@gmail.com; marcosguedeslana@gmail.com; carlos.schaefer@ufv.br; elpidio@ufv.br

[2] Department of Soil Science, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Av. Pádua Dias, 11, CP 9, Piracicaba, SP 13418-900, Brazil; emails: fellipeamello@usp.br; raulpoppiel@gmail.com; luis.diloreto@hotmail.com; jamdemat@usp.br*

[3] Geography Department of Federal University of Mato Grosso do Sul, Av. Ranulpho Marques Leal, nº 3484 - Distrito Industrial CEP 79610-100 Três Lagoas/MS: diego.cabrero@gmail.com

[4] Department of Geology and Natural Resources, Institute of Geosciences, University of Campinas, Rua Carlos Gomes, 250, Cidade Universitária, CEP 13083-855, Campinas/SP: emilson@ige.unicamp.br

Correspondence to: José Alexandre Melo Demattê (jamdemat@usp.br)

Abstract. Geophysical sensors combined with machine learning algorithms were used to understand the pedosphere system, landscape processes and to model soil attributes. In this research, we used parent material, terrain attributes, and data from geophysical sensors in different combinations, to test and compare different and novel machine learning algorithms to model soil attributes. We also analyzed the importance of pedo-environmental variables in predictive models. For that, we collected soil physico-chemical and geophysical data (gamma-ray emission from uranium, thorium and potassium, magnetic susceptibility and apparent electric conductivity) by three sensors (gamma-ray spectrometer - RS 230, susceptibilimeter KT10 – Terraplus and Conductivimeter – EM38 Geonics) at 75 points and analyzed the data. The models with the best performance (R^2 0.48, 0.36, 0.44, 0.36, 0.25 and 0.31) varied for clay, sand, Fe_2O_3, TiO_2, SiO_2 and Cation Exchange Capacity prediction, respectively. Modeling with the selection of covariates at three phases (variance close to zero, removal by correction, and removal by importance) was adequate to increase the parsimony. The results were validated using the method "nested leave one-out cross validation". The prediction of soil attributes by machine learning algorithms yielded adequate values for field-collected data, without any sample preparation, for most of the tested predictors (R^2 values ranging from 0.20 to 0.50). Also, the use of four regression algorithms proved to be important since at least one of the predictors used one of the tested algorithms. The performances values of the best algorithms for each predictor were higher than those obtained with the use of a mean value for the entire area comparing the values of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The best combination of sensors that reached the highest model performance was that of the gamma-ray spectrometer and the susceptibilimeter. The most important variables were parent material, digital elevation, standardized height, and magnetic

susceptibility for most predictions. We concluded that soil attributes can be efficiently modelled by geophysical data using machine learning techniques and geophysical sensor combinations. This approach can facilitate future soil mapping in a more time-efficient and environmentally friendly manner.

1 Introduction

The pedosphere is composed of soils and their connections with the hydrosphere, lithosphere, atmosphere, and biosphere (Targulian et al, 2019). Soils are the result of several processes and factors and their interactions, resulting in specific soil types or horizons. The main soil processes are weathering and pedogenesis (Breemen and Buurman, 2003; Schaetzl and Anderson, 2005), and the soil-forming factors are parent material, relief, climate, organisms and time (Jenny, 1994). Their interactions during soil genesis results in different soil attributes such as texture, mineralogy, color, structure, base saturation, clay activity, among others.

In the last decades, there has been a growing demand for soil resource information worldwide (Amundson et al., 2015; Montanarella et al., 2015). Soils are recognized as having a key influence on global issues such as, water availability, food security, sustainable energy, climate change and environmental degradation (Amundson et al., 2015; Pozza and Field, 2020). Therefore, understanding the role of spatial variations in surface and subsurface soil is fundamental for its sustainable use as well as for other connected environmental resources and monitoring (Agbu et al., 1990). In this sense, it is necessary to increase the acquisition of information on the functional attributes of soils, and to achieve this, relevant and reliable soil information, applicable from local to global scales is required (Arrouays et al., 2014).

The acquisition of soil data and their attributes is generally achieved by traditional soil survey techniques. However, new geotechnologies have emerged in the last decades, allowing the acquisition of data at shorter times, with non-invasive and accurate methods such as reflectance spectroscopy, satellite imagery, and geophysical techniques (Mello et al., 2020; Demattê et al., 2017, 2007; Fioriob, 2013; Fongaro et al., 2018; Mello et al., 2021; Terra et al., 2018a, 2018b). Among these technologies, geophysical sensors have been recently used in pedology to understand pedogenesis and the relationship between these processes and soil attributes (Son et al., 2010; Schuler et al., 2011; Beamish, 2013; McFadden and Scott, 2013; Sarmast et al., 2017; Reinhardt and Herrmann, 2019). Among these geophysical techniques used, we highlight gamma-spectrometry, magnetic susceptibility (κ), and apparent electrical conductivity (ECa).

Gamma-ray spectrometry can be defined as the measurements of natural gamma radiation emission from natural emitters, such as K^{40}, the daughter radionuclides of U^{238} and Th^{232}, and total emissions from all elements in soils, rocks and sediments (Minty, 1988). Weathering and pedogenesis, concomitantly with the geochemical behavior of each radionuclide, determine their distribution and concentration in the pedosphere (Dickson and Scott, 1997; Wilford and Minty, 2006; Mello et al., 2021). Therefore, gamma-ray spectrometry can provide important information for the comprehension of soil processes and attributes (Reinhardt and Herrmann, 2019), soil texture (Taylor et al., 2002a), mineralogy (Wilford and Minty 2006; Barbuena et al. 2013), pH (Wong and Harper, 1999) and organic carbon (Priori et al., 2016).

Soil magnetic susceptibility (κ) can be defined as the degree to which soil particles can be magnetized (Rochette et al., 1992). The κ is related to several pedoenvironmental factors, such as soil mineralogy, lithology, and geochemistry of ferrimagnetic secondary minerals, such as magnetite and maghemite (Ayoubi et al., 2018). Also, the κ parameter can be related to other soil secondary minerals, like ferrihydrite and hematite (Valaee et al., 2016). The great potential of this technique is related to geological studies (Shenggao 2000; Correia et al. 2010), soil texture and organic carbon studies (Camargo et al., 2014; Jiménez et al., 2017), soil surveys (Grimley et al., 2004), and pedogenesis and pedogeomorphological processes (Viana et al., 2006; Sarmast et al., 2017; Mello et al., 2020).

Apparent electrical conductivity (ECa) is the ability of the soil to conduct an electrical current, expressed in *millisiemens* per meter. This soil property is related to the presence/amount of solutes in the soil solution, whose concentration in 1 dS/m is equivalent to 10 meq/L (Richards, 1954). Concerning the geophysical methods, the ECa is a geotechnology for identifying the soil physicochemical attributes and their spatial variation (Corwin et al., 2003). Various different soil attributes are related to the ECa, such as soil salinity (Narjary et al., 2019), soil texture (Domsch and Giebel, 2004), cation exchange capacity (Triantafilis et al., 2009), mineralogy, pore size distribution, temperature, soil moisture (McNeill, 1992; Rhoades et al., 1999; Bai et al., 2013; Farzamian et al., 2015; Cardoso and Dias, 2017).

As various sensors scan only the soil surface, disregarding the entire soil tridimensional profile (Xu et al., 2019), a single sensor may not be able or be the best solution to quantify multiple soil attributes. In this context, the concept and use of multi-sensor data acquisition and analysis is a complementary way to offer more robust and accurate estimations of a number of soil attributes (Xu et al., 2019; Javadi et al., 2021). The analysis of soil data acquired by multiple sensors requires a careful interpretation and a mathematical model, which can be considered the base of the observed variation and provides the basis for generalization, prediction and interpretation. (Heuvelink and Webster, 2001).

Recently, many models have been used to estimate soil attributes and their spatial distribution from geophysical data (gamma-ray, κ, and ECa) and soil attributes, including machine learning algorithms, such as Support Vector Machine-SVM (Priori et al., 2014; Heggemann et al., 2017; Li et al., 2017; Leng et al., 2018; Zare et al., 2020), Random Forests (Lacoste et al., 2011; Viscarra Rossel et al., 2014; Harris and Grunsky, 2015; Sousa et al., 2020), KNN and artificial neural network (ANN) (Dragovic and Onjia, 2007) and Cubist (Wilford and Thomas, 2012).

According to Batty and Torrens, (2001), the best models are those capable of explaining the same phenomena using the smallest number of variables without loss of performance, following the principle of parsimony - Occam's razor. Models that use fewer variables usually optimize the modelling process, making it easier to explain the influence of the variables on the modelling process and providing results that are easier to interpret. In addition, this facilitates the understanding and the faster computer processing of the data (Brungard et al., 2015). In this context, the Recursive Feature Elimination (RFE) algorithm may be used for the backward selection of optimal subsets of variables, while maintaining a satisfactory model performance (Vašát et al., 2017; Hounkpatin et al., 2018).

Some of geophysical sensors can detect soil attributes in the upper soil layers (0 – 0.50 m for gamma-ray by the RS230 model, 0.02 m for the magnetic susceptilimeter KT10 Terraplus model, and 1.5 m for the conductivimeter via the EM38 model, for

example), which are explained by naturally occurring soil processes and formation by soil factors (Mello et al., 2020; Mello et al., 2021). However, there is still a knowledge gap regarding the identification of the best covariables and their possible combinations to deepen our knowledge of soil weathering, genesis, and their relation to soil attributes. A standard approach to selecting the best input data to soil prediction models has yet to be developed (Levi and Rasmussen, 2014), mainly for geophysical sensors, which are little used in soil science. The identification of such covariates may improve the understanding of the interplays between soil processes and attributes, allowing an enhanced comprehension of soils from the punctual to the landscape scale, supporting digital soil mapping and better soil use and management.

In this context, this study aimed to: *i)* develop a new methodological framework on modelling soil attributes using combined data from three different geophysical sensors at five different sensor combinations; *ii)* assess the use of different machine learning algorithms and test the nested leave one out cross-validation method for prediction and selection of suitable models for each soil attribute evaluated; *iii)* evaluate the results and the importance of the variables and relate them to pedogeomorphological processes. Our main hypothesis is that the combined use of three geophysical sensor data enable a better prediction of soil attributes by different machine learning algorithms and better model performance. This study can provide an important background for geoscience studies and the improvement of geophysical and soil survey procedures.

2 Material and methods

2.1 Study area

The study area was located on a sugarcane farm covering 184 hectares, located in São Paulo State, Brazil (23º 0’ 31.37” to 22º 58’ 53.97” S and 53º 39’ 47.81” to 53º 37’ 25.65” W), in the Capivari River catchment, part of the Paulista Peripheric Depression geomorphological unit (Fig. 1). The lithology is mainly composed of Paleozoic sedimentary rocks, dominated by Itararé formation (siltites/meta-siltites) crossed by intrusive diabase dykes of the Serra Geral Formation. The lowlands are covered by Quaternary alluvial sediments deposited by the Capivari River in ancient fluvial terraces (Fig. 2a).

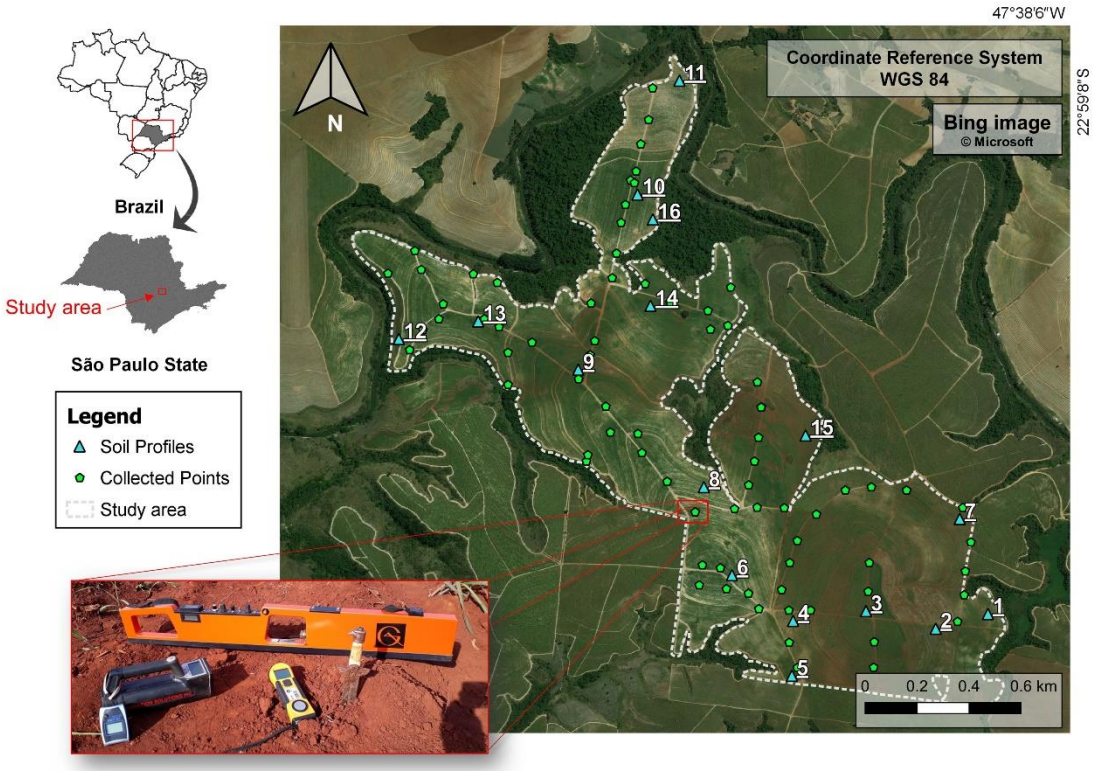


Figure 1. Study area, collection points, and geophysical sensors. A - Gamma-ray spectrometer (Radiation Solution - RS 230); B - Susceptibilimeter (KT-10 Terraplus); C - Geonics Ground Conductivity Meter (EM 38).

The heterogeneity of the landform and the parent materials drove the formation of several soil types (**Fig. 2b**). Previous soil surveys and mapping have been performed in the study area by expert pedologists (Bazaglia Filho et al., 2013; Nanni and Demattê, 2006), in which the main soil classes mapped were as follows: Cambisols, Phaeozems, Nitisols, Acrisols, and Lixisols (IUSS Working Group WRB, 2015). Besides the soil profiles, 75 subsamples from 75 points (0–20 cm layer) were collected with an auger for physicochemical analyses, according to **Figure 1**.

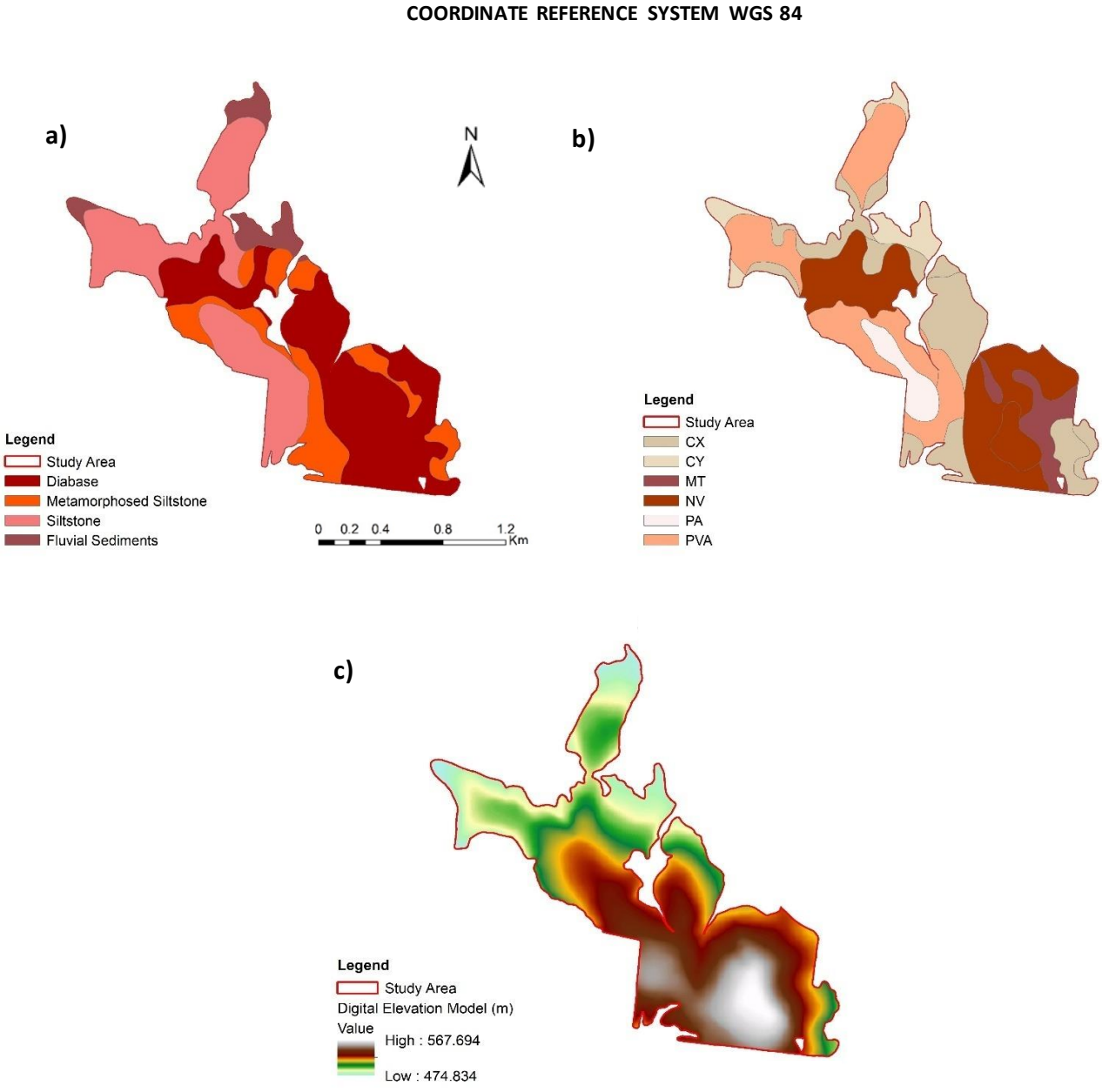


Figure 2. a) Geological compartments of the landscape. b) Soil classes: CX: Haplic Cambisols, CY: Fluvic Cambisols, MT: Luvic Phaozem, NV: Rhodic Nitisol: PA: Xanthic Acrisol, PVA: Rhodic Lixisol. The geological and Soil classes maps were adapted from Bazaglia Filho et al. (2012). d) Slope. c) Digital elevation model.

According to the Köppen classification the region's climate is subtropical, mesothermal (Cwa), with an average temperature from 18 °C (July-Winter) to 22 °C (February-Summer) and a, mean annual precipitation between 1,100 and 1,700 mm (Alvares et al., 2013).

2.2 Laboratory physico-chemical analysis

For soil physical analyses, the soil samples were first air-dried, ground, and sieved through a 2 mm mesh, followed by granulometric analysis. After that, clay, silt, and sand contents were determined by the densimeter method (Camargo et al., 1986). Using the granulometry data, the textural groups were determined following the EMBRAPA (2011) methodology.

The exchangeable cations aluminum, calcium, and magnesium (Al^{3+}, Ca^{2+}, and Mg^{2+}) were determined using KCl solution (1 mol L^{-1}) and quantified by titration (Teixeira et al., 2017). Mehlich-1 solution was used to extract K^{+}, which was quantified by flame photometry. Potential acidity (H^{+} + Al^{3}) was determined using calcium acetate solution (0.5 mol L^{-1}) at pH 7.0; for the pH in water determination, the soil: solution ratio of 1:2.5 was used (Teixeira et al., 2017). More details about the analysis methods can be found elsewhere (EMBRAPA, 2017). Soil organic carbon was determined using the Walkley Black method via oxidation with potassium (EMBRAPA, 2017; Pansu, M., Gautheyrou, J., 2006). The total iron content was determined using selective dissolution in sulfuric acid (EMBRAPA, 2017; Lim, C.H., Jackson, 1986). The resulting extract was used to determine the contents of silicon dioxide (SiO_2) and titanium dioxide (TiO_2), using the EMBRAPA methodology (2017). All other chemical parameters, such as Base Sum (BS) Cation Exchange Capacity (CEC), Base Saturation (V%), and Aluminum Saturation (m%), were determined using the analytical data obtained previously, following the methodology described elsewhere (EMBRAPA, 2017).

2.3.1 Radionuclides and gamma-ray spectrometry data

The total radionuclide K^{40} amount was measured by the absorption energy (1.46 MeV). Thorium (Th^{232}) and uranium (U^{238}) were quantified by absorption energy (approximately 2.62 and 1.76 MeV, respectively). This quantification was indirectly performed through thallium (Tl^{208}) and bismuth (Bi^{214}), derived by radioactive decay, respectively, for Th^{232} and U^{238}, which are expressed as eTh and eU (equivalent thorium and uranium, respectively).

For soil gamma spectrometric characterization, we used the near-gamma-ray spectrometer (GM) model Radiation Solution RS 230 – Radiation Solution INC – Ontario - Canada (**Fig. 1A**). The sensor can quantify the eTh and eU concentrations in parts per million (ppm), whereas K^{40} is quantified in % due to its major content in the pedosphere. Conventionally, radionuclides are expressed in mg kg^{-1} for eU and eTh, whereas for K^{40}, percentage is used. The GM detects the gamma-ray radiation emission down to a depth of 30–60 cm, which varies mainly with soil bulk density and moisture content (Wilford et al., 1997; Taylor et al., 2002; Beamish, 2015).

First, the GM was automatically calibrated by switching on and leaving the sensor on the ground surface for 5 minutes until readings of eU, eTh, and, K^{40} contents stabilized (Radiation Solutions, 2009). The measurements of radionuclides were taken in the "assay-mode" of the highest precision for quantification, in which the GM was kept at the soil surface for 2 minutes in

each sampling point (79 total collection points) (**Fig. 1)**. The geographic position was taken by a GPS coupled to the GM (GPS – Radiation Solution INC – Ontario – Canada – precision of 1 m). The data collected from all points were concatenated with their respective information from the soil physico-chemical analyses for later geoprocessing. The same methodology has been applied by Mello et al. (2021) for gamma-ray spectrometric data acquisition.

2.3.2 Magnetic susceptibility (κ)

For soil magnetic susceptibility (κ) characterization, surface readings were recorded at all 79 points, using a geophysical susceptibility meter sensor (*KT10 – Terraplus*) (**Fig. 1b**). This sensor can measure κ to a depth of 2 cm below the soil surface, with a precision of 10^{-6} *SI* units, expressed in $m^3\ kg^{-1}$. To perform the readings, the sensor was first calibrated by determining the frequency of the outdoor oscillator. Subsequently, we followed the sequence required to obtain the measurements performed in three steps: 1 - determining the frequency and amplitude of the oscillator in free air; 2 – measuring the frequency and amplitude of the oscillator with the coil placed directly on the soil surface (sample) outcrop; 3 – repeating step 1 and displaying the results. For more information about these procedures, see Sales (2021). We performed the readings at *scanner mode*, which uses the best geometric correlation to direct κ readings, providing fast and accurate quantification. We performed three readings in triangulation around each collection point and used the mean value of κ in all our analyses. This procedure was adopted to reduce noise. The same methodology for κ readings has been performed by Mello et al. (2020).

2.3.3. Apparent electrical conductivity (ECa)

The ECa measurements were performed using the conductivity meter Geonics EM38 (Geonics Ltd., Mississauga, Ontario, Canada) (McNeill, 1986) (**Fig. 1C**). The EM38 provides measurements of the quad-phase (conductivity) without any requirement for soil-to-instrument contact (Geonics, 2002); the unit is m Sm^{-1}.

First, the EM38 was calibrated following the instructions of Heil and Schmidhalter, (2019), Section 3.1.1. The values of ECa are a function of calibration, coil orientation, and coil separation (Heil and Schmidhalter, 2019). More details about the EM38 operation are provided in Hendrickx and Kachanoski (2002). After calibration, the ECa readings were performed at all 75collection points (**Fig. 1),** using the EM38 at vertical dipole orientation, which provided data from an effective soil depth at 1.5 m. Data were collected in the filed during the dry season, on bare soil, and at the same intervals to reduce the impacts of environmental variables. Also, all metal objects were kept away from the EM 38 to avoid reading interferences.

We developed our research and analysis by using three geophysical sensors (near-gamma-ray spectrometer RS 230, near-magnetic susceptibility sensor KT10, and conductivimeter Geonics EM38) due to the following reasons: these sensors are available in our institution and for our research partners, they are easy to operate, and the obtained data are highly accurate. In addition, the EM38 (conductivimeter) and RS 230 (gamma-ray spectrometer) provide information for the depth at which most of the pedogenetic processes occur. In addition, information obtained with EM38 and RS 230 can be associated with KT10

(susceptibilimeter) on the soil surface to provide additional information about some soil attributes related to soil subsurface horizons, which is also related to the other geophysical variables used (gamma-ray and apparent electrical conductivity).

Table 1. Terrain variables generated from the digital elevation model.

Terrain attributes	Abbreviations	Brief description
Convergence index	CI	Convergence/divergence index in relation to runoff
Cross-sectional curvature	CSC	Measures the curvature perpendicular to the down slope direction
Flow-line curvature	FLC	Represents the projection of a gradient line to a horizontal plane
General curvature	GC	Combination of both plan and profile curvatures
Hill	HI	Analytical hill shading
Hill index	HIINDEX	Analytical index hill shading
Longitudinal curvature	LC	Measures the curvature in the down-slope direction
Mass balance index	MBI	Balance index between erosion and deposition
Maximal curvature	MAXC	Maximum curvature in local normal section
Mid-slope position	MSP	Represents the distance from the top to the valley, ranging from 0 to 1
Minimal curvature	MINC	Minimum curvature for local normal section
Multiresolution index of ridge top flatness	MRRTF	Indicates flat positions in high-elevation areas
Multiresolution index of valley bottom flatness	MRVBF	Indicates flat surfaces at the bottom of the valley
Normalized height	NH	Vertical distance between base and ridge of normalized slope
Plan curvature	PLANC	Curvature of the hypothetical contour line passing through a specific cell
Profile curvature	PROC	Surface curvature in the direction of the steepest incline
Slope	S	Represents local angular slope
Slope height	SH	Vertical distance between base and ridge of slope
Standardized height	STANH	Vertical distance between base and standardized slope index
Surface specific points	SSP	Indicates differences among specific surface shift points
Tangential curvature	TANC	Measured in the normal plane in a direction perpendicular to the gradient
Terrain ruggedness index	TRI	Quantitative index of topography heterogeneity
Terrain surface convexity	TSC	Ratio of the number of cells that have positive curvature to the number of all valid cells within a specified search radius

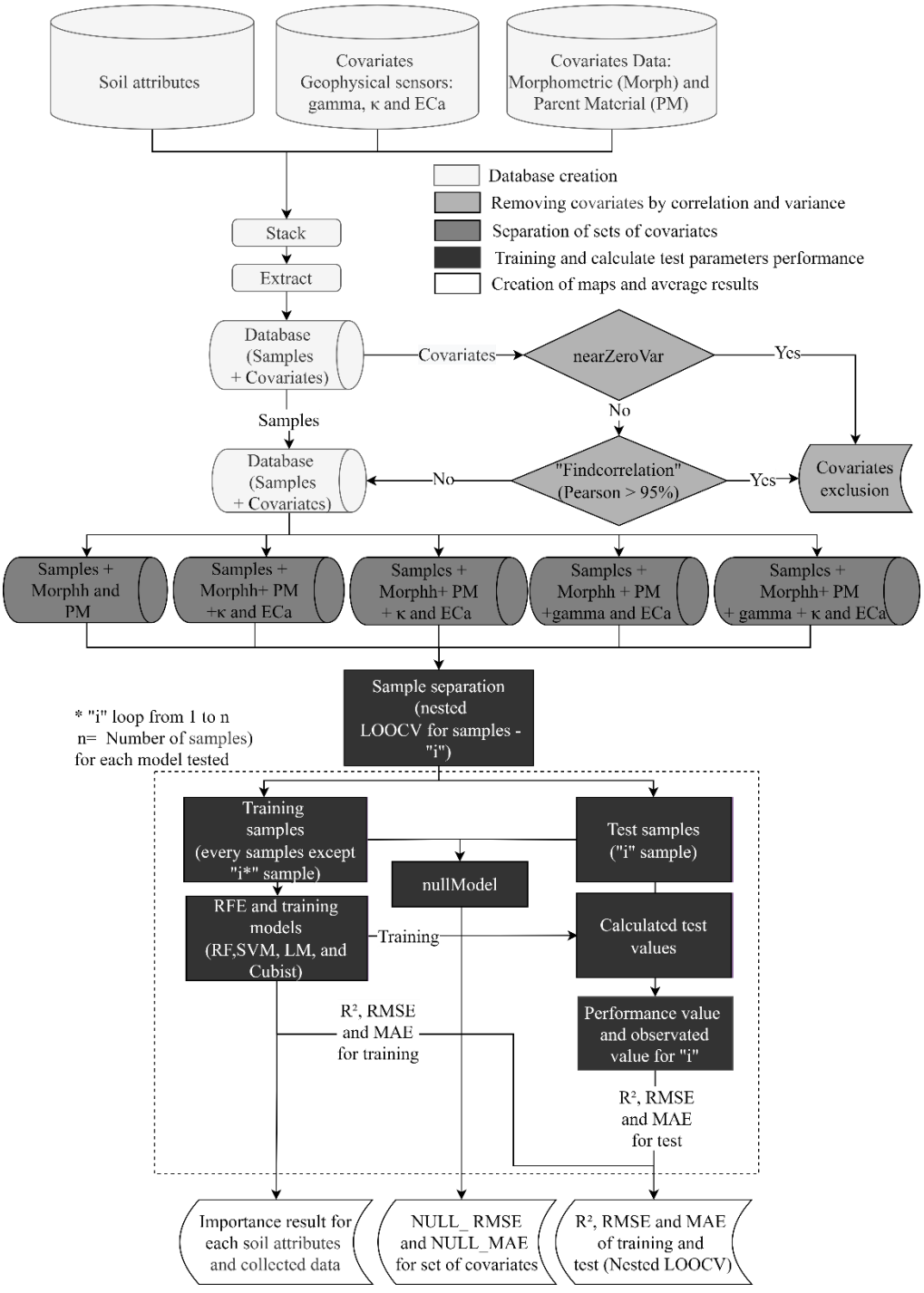
Cont..

Cont. Table 1

Terrain attributes	Abbreviations	Brief description
Terrain surface texture	TST	Splits surface texture into 8, 12, or 16 classes
Total curvature	TC	General measure of surface curvature
Topographic position index	TPI	Difference between a point elevation to the surrounding elevation
Valley depth	VD	Calculation of vertical distance at drainage base level
Valley	VA	Calculation of the fuzzy valley using the Top Hat approach
Valley Index	VAI	Calculation of the fuzzy valley index using the Top Hat approach
Topographic wetness index	TWI	Describes the tendency of each cell to accumulate water in relief

2.3.5. Modelling processing

The modeling process is demonstrated in the flowchart (**Fig. 3**) and can be divided into two parts: the selection of covariates and the training/testing of the data. In the selection phase, the algorithm tries to produce the ideal set of covariates, following the principle of parsimony. This is performed by removing highly correlated variables, evaluating the importance of covariables and removing variables that have a minor importance in training the model in the prediction process of each algorithm. Darst et al. (2018) considered the joint application of the methods for the selection of covariates by correlation and importance (RFE) since the use of RFE only reduces the effect of highly correlated covariates but does not eliminate it.



220 **Figure 3.** Methodological flowchart showing the sequence of methodologies applied for soil and geophysical attribute prediction. The most accurate model among Cubist, Random Forests (RF), Support Vector Machines (SVM), and Linear Models (LM) was selected to model and map the geophysical and soil attributes.

The correlation selection process was used to calculate the correlation of the set of covariates and covariables, which were evaluated with a correlation greater than the limit (Pearson test > 95%). The pairs that showed higher values were evaluated due to their correlation with the complete set of covariates, eliminating that with the highest value of the sum of the absolute correlation with the other covariables that started in this process. For this phase, we applied the *cor* and *find correlation* functions of the "*stats*" (Hothorn, 2021) and "*caret*" (Kuhn et al., 2020) packages, in the *R* software, respectively (Kuhn and Johnson, 2013). In this phase, the covariables: curv_cross_secational and curv_longitudinal were eliminated for all tested sensor sets. The set of covariables that passed this phase joined the samples followed by the separation of samples from training and testing.

The separation of training and testing was performed using the "nested" leave one out (nested LOOCV) method (Clevers et al., 2007; Honeyborne et al., 2016; Rytky et al., 2020). It is important to highlight that our number of soil samples and readings with geophysical sensors was small (75) due to several difficulties encountered in the field during data collection (high sugar cane size, sloping terrain, dense forest, etc.). In this sense, the nested LOOCV method is indicated for small sample sets (values near 100 samples) to which other validation/testing methods (as *holdout* validation) would not be viable due to the small sample set in the testing and/or training group (Ferreira et al., 2021). This is one of the main innovations of this research.

The nested LOOCV method is a double-loop process. In the first loop, the model is trained with a data set of size n-1, and the test is done in the second loop with the missing sample to validate the training performance (Jung et al., 2020; Neogi and Dauwels, 2019). The final results of the performance of the machine learning algorithm will be the mean performance indicators for all points (training/testing). This is a robust method to evaluate the performance of the algorithm and to detect possible samples with problems in the collections or outliers. The training set generated in each loop went through the process of selecting covariates for importance and subsequent training.

The selection of covariates by importance is performed using the *back forward* method, applying the Recursive Feature Elimination (RFE) function contained in the "caret" package (Kuhn and Johnson, 2013). The RFE is unique for each algorithm, with the result being the set of selected covariates used in the prediction of the final model in the same algorithm. The RFE is a selection method that eliminates the variables that least contribute to the model, based on a measure of importance for each algorithm (Kuhn and Johnson, 2013). The algorithm will be applied to complete sets of data (variable by the set of tested sensors) and 18 more subsets with 5,6,7, ... 19, 20 and 30 covariables. Reaching a set of fewer variables (more parsimonious), results in a better prediction performance. The optimization of the ideal covariate subset was based on leave one out (LOOCV), a repetition, and four values of each of the internal hype parameters of each tested algorithm (*tuneLength*). The hyperparameters of each algorithm are described in the caret package manual in chapter 6. "Models described" available at https://topepo.github.io/caret/train-models-by-tag.html. The metric for choosing the best subset for each model were R^2. For this work, five algorithms were tested: Random Forests (RF), Cubist (C), Support Vector Machines (SVM), Generalized Linear Models (lm). The choice was made with the use of families of different algorithms in mind, using linear and non-linear algorithms. The algorithms used are commonly applied in soil attribute mapping studies. At the end of the selection phase by importance, the most optimized set of covariates for training was generated for each algorithm.

Training was performed with the variables selected in the previous step and each tested algorithm by using LOOCV and 10 repetitions. Four values of each of the internal hype parameters of each tested algorithm were also tested (tuneLength). At the end of the training phase, a sample prediction was made that was not used in the training and the result was saved for the performance study. The performance of the prediction of the algorithms and the set of sensors was determined with a set of samples from the outer loop of the nested LOOCV method. Three evaluation parameters were used: R-square - R^2 (Eq. (1)), root mean squared error - RMSE (Eq. (2)), mean absolute error - MAE, (Eq. (3)).

$$R^2 = \frac{\left[\sum\left(Qpred - \overline{Qpred}\right)\times\left(Qobs - \overline{Qobs}\right)\right]^2}{\left[\sum(Qpred - \overline{Qpred})^2\right]\times\left[\sum(Qobs - \overline{Qobs})^2\right]} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\times\sum(Qobs - Qpred)^2} \quad (2)$$

$$MAE = \frac{1}{n}\times\sum|Qpred - Qobs| \quad (3)$$

Where:

Qpred = predicted samples

Qobs = observed samples

n = number of samples

For comparison purposes, null model values (NULL_RMSE and NULL_MAE) were also calculated. The null model considers using the average value quantified by the collected samples (EQ. 4 and EQ. 5). The null model (NULL_RMSE and NULL_MAE) emulates other model-building functions but returns the simplest model possible given a training set: a single mean for numeric outcomes. The percentage of the training set samples with the most prevalent class is returned when class probabilities are requested. The null model can be considered the simplest model that can be adjusted and that serves as a reference. Models that present similar or worse performances compared to the null model should be discarded. The best models had lower RMSE and MAE results than those found for NULL_MAE and NULL_RMSE. This shows that the final model is better than using the mean values, which also demonstrates a better quality in creating the models.

Given above, the null model considers using the mean value quantified by the collected samples (EQ. 4 and EQ. 5). This methodology is widely used, as well as spatialization processes in kriging when the variable in which spatialization is desired has spatial dependence (pure nugget effect). The equations are as follows:

$$NULL\_RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(\overline{Qtrain_i} - Qobs_i)^2\right]^{\frac{1}{2}} \quad \text{(Eq.4)}$$

$$NULL\_MAE = \frac{1}{n} \times \sum|\overline{Qtrain_i} - Qobs_i| \quad \text{(Eq.5)}$$

Where:

Qtrain = the mean of the training samples

Qobsi = the validation sample

N =number of samples (loop).

Here NULL_RMSE and NULL_MAE values lower than those observed in the prediction of the algorithm in the validation phase show that the use of means of the samples of the desired propriety agrees with the model created by the algorithms of the machine learning. The NULL_RMSE and NULL_MAE were calculated using the *nullMode* function of the caret package (Kuhn et al., 2020).

The final result of the performance of the algorithms of each attribute was obtained using the 75 loops, with the training results being the average of the performance and the results of the test samples calculated from the 75 external loops results using Equations 1, 2 ,and 3. The importance of the algorithms was calculated by the caret package (Kuhn and Johnson, 2013), each model presents its creation methodology. The final importance for each algorithm and attribute was determined from the importance created in the loop, being the average of the importance of the 75 repetitions.

3 Results

3.1. Geophysical sensor combinations, model performance, uncertainty, and covariate importance

The worst performance in modeling soil attributes occurred excluding the use of geophysical sensors (non-use of the geophysical sensor), where only parent material and terrain attributes were used (**Table 2**). In this case, the algorithms selected particular groups of terrain attributes for the modelling of each soil attributes (**Table 1**).

Table 2. Model performance for non-use geophysical sensors, for all soil attributes, based on R^2, RMSE, MAE, and NULL_RMSE.

Non-use of geophysical sensors	***R²***				
	Random Forest	***Cubist***	***SVM***	***LM***	-
Clay	0.38	0.386	0.259	0.285	-
Sand	0.284	0.292	0.278	0.225	-

Fe2O3	0.159	0.12	0.279	0.217	-
TiO2	0.12	0.125	0.226	0.16	-
SiO2	0.12	0.174	0.128	0.247	-
CEC	0.149	0.053	0.195	0.002	-
BS	0.131	0.028	0.113	0.003	-
OM	0	0.001	0.004	0.051	-
Non-use of geophysical sensors	***RMSE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_RMSE***
Clay	136.778	140.103	154.406	156.646	140.885
Sand	185.398	192.867	190.151	215.355	176.521
Fe2O3	61.686	66.432	59.453	66.357	53.341
TiO2	12.229	12.424	11.621	13.118	10.239
SiO2	41.701	41.323	42.595	38.976	35.45
CEC	41.3	50.065	41.141	997.529	36.139
BS	20.206	22.853	20.396	1189.64	17.142
OM	8.469	8.126	8.045	7.702	6.158
Non-use of geophysical sensors	***MAE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL MAE***
Clay	110.485	108.284	122.397	119.139	119.751
Sand	149.205	148.8	147.07	169.218	153.803
Fe2O3	40.742	44.028	36.812	43.673	41.578
TiO2	8.206	8.294	7.051	8.749	8.074
SiO2	31.757	31.715	31.432	29.458	29.534
CEC	28.931	33.168	27.072	149.114	27.187
BS	16.3	18.271	17.012	158.638	14.425
OM	6.357	4.813	5.992	5.719	4.813

Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: mmolc dm^{-3}. Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: mmolc dm^{-3}. Support Vector Machines (SVM); Linear Models (LM).

The Cubist algorithm (non-use of the geophysical sensor) showed the best performance in predicting soil texture, clay (R^2 of 0.386) and sand (R^2 of 0.292) contents, with the highest R^2 and the lowest RMSE and MAE values, concomitantly (**Table 2**). The importance of covariates to sand content prediction showed that minimal curvature, was the most important variable,

contributing 100% to the decrease mean accuracy. On the other hand, for clay content, the most important variable was parent material. In addition, for clay and sand, the tangential curvature and DEM showed an importance higher than 50% (**Fig. 4**).

When the geophysical sensor was not used, the SVM algorithm presented a moderate performance for Fe_2O_3 (R^2 0.279) and TiO_2 (R^2 0.226), whereas for SiO_2, the LM presented the best result, also with a moderate performance (R^2 0.247) (**Table 2**). The selected models simultaneously presented the highest R^2 and lowest RMSE and MAE values. The most important covariates for Fe_2O_3 and TiO_2 prediction by the SVM model were parent material (100%) and DEM (more than 50%). For SiO_2 prediction by the LM model, the most important covariates were DEM (100%) and standardized height (90%), whereas parent material contributed with 40% (**Fig. 4**).

For cation exchange capacity (CEC), the model with the best performance, after 75 runs was SVM, (R^2 of 0.223) (**Table 2**) when the geophysical sensor was not used. The most important covariates for CEC prediction to mean accuracy were DEM (100%), topographic wetness index (80%), and parent material (75%) (**Fig. 4**).

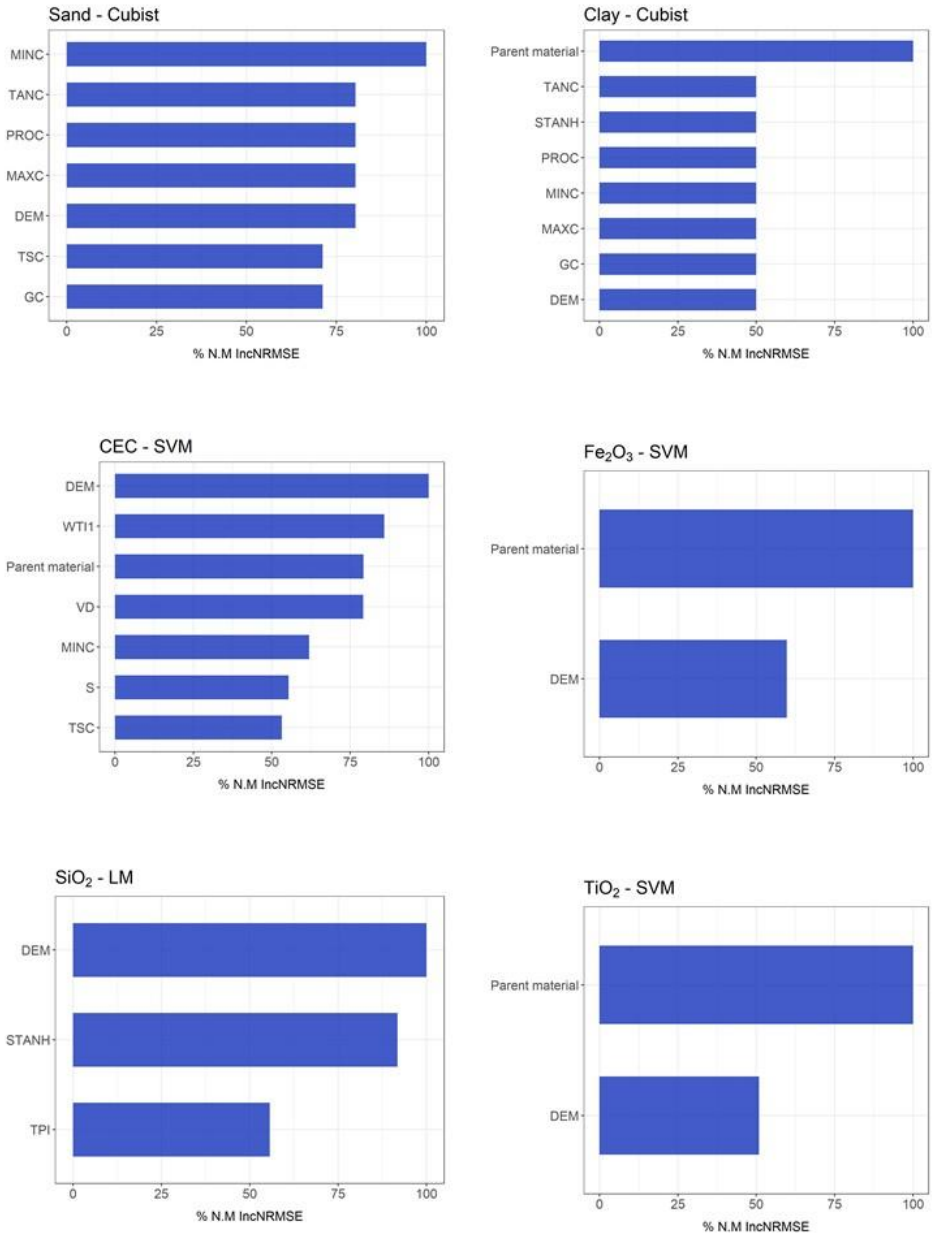


Figure 4. Variable importance for *non-use of geophysical sensors* (only variables that contributed more than 50% are presented here). For further details, see supplementary material.

All models showed a low performance in the prediction of base saturation (BS) and organic matter (OM), with R^2 values between 0.001 and 0.1 (**Tables 2, 3, 4, 5, and 6**).

The different combinations of geophysical sensors that contributed to the moderate modeling performance for soil attributes were as follows: Susceptibilimeter + Conductivimeter (S + C), Gamma-ray spectrometer + Conductivimeter (G + C), combined use of the three geophysical sensors (G + S + C) (**Tables 3, 4, and 6,** respectively). The R^2 values presented some variations between the R^2 of the best combination of geophysical sensors and the lowest R^2 values when the geophysical sensors were not used in the predictive models (**Tables 3, 4, and 6**). Among all the values of R^2 evaluated for this session, we considered all the highest values; among the highest values, we considered the lowest values as the worst results.

Table 3. Model performance for the combined use of susceptibilimeter and the conductivimeter, for all soil attributes, based on R^2, RMSE, MAE and NULL_RMSE.

Susceptibilimeter + Conductivimeter	***R²***				
	Random Forest	***Cubist***	***SVM***	***LM***	**-**
Clay	0.444	0.433	0.484	0.394	-
Sand	0.334	0.365	0.322	0.312	-
Fe2O3	0.314	0.407	0.153	0.383	-
TiO2	0.316	0.338	0.263	0.262	-
SiO2	0.141	0.25	0.169	0.101	-
CEC	0.139	0.178	0.223	0.124	-
BS	0.138	0.079	0.065	0.002	-
OM	0.032	0.077	0.039	0.056	-
Susceptibilimeter + Conductivimeter	***RMSE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_RMSE***
Clay	129.619	136.834	127.598	139.463	140.885
Sand	178.22	178.253	181.811	190.515	176.521
Fe2O3	55.378	52.416	64.573	54.36	53.341
TiO2	10.531	10.583	11.052	11.622	10.239
SiO2	41.116	39.138	42.22	46.013	35.45
CEC	41.878	41.91	40.134	48.52	36.139
BS	19.821	21.543	22.307	1219.091	17.142
OM	8.079	7.494	7.924	8.007	6.158
Susceptibilimeter + Conductivimeter	***MAE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_MAE***
Clay	102.841	105.12	92.812	106.083	119.751
Sand	145.441	139.737	146.016	153.815	153.803
Fe2O3	34.357	32.246	40.303	36.79	41.578
TiO2	6.457	6.593	6.65	8.199	8.074
SiO2	30.54	28.954	31.153	33.218	29.534
CEC	29.354	28.912	26.689	33.024	27.187
BS	15.824	17.372	18.953	161.284	14.425
OM	5.949	5.713	6.108	6.04	4.813

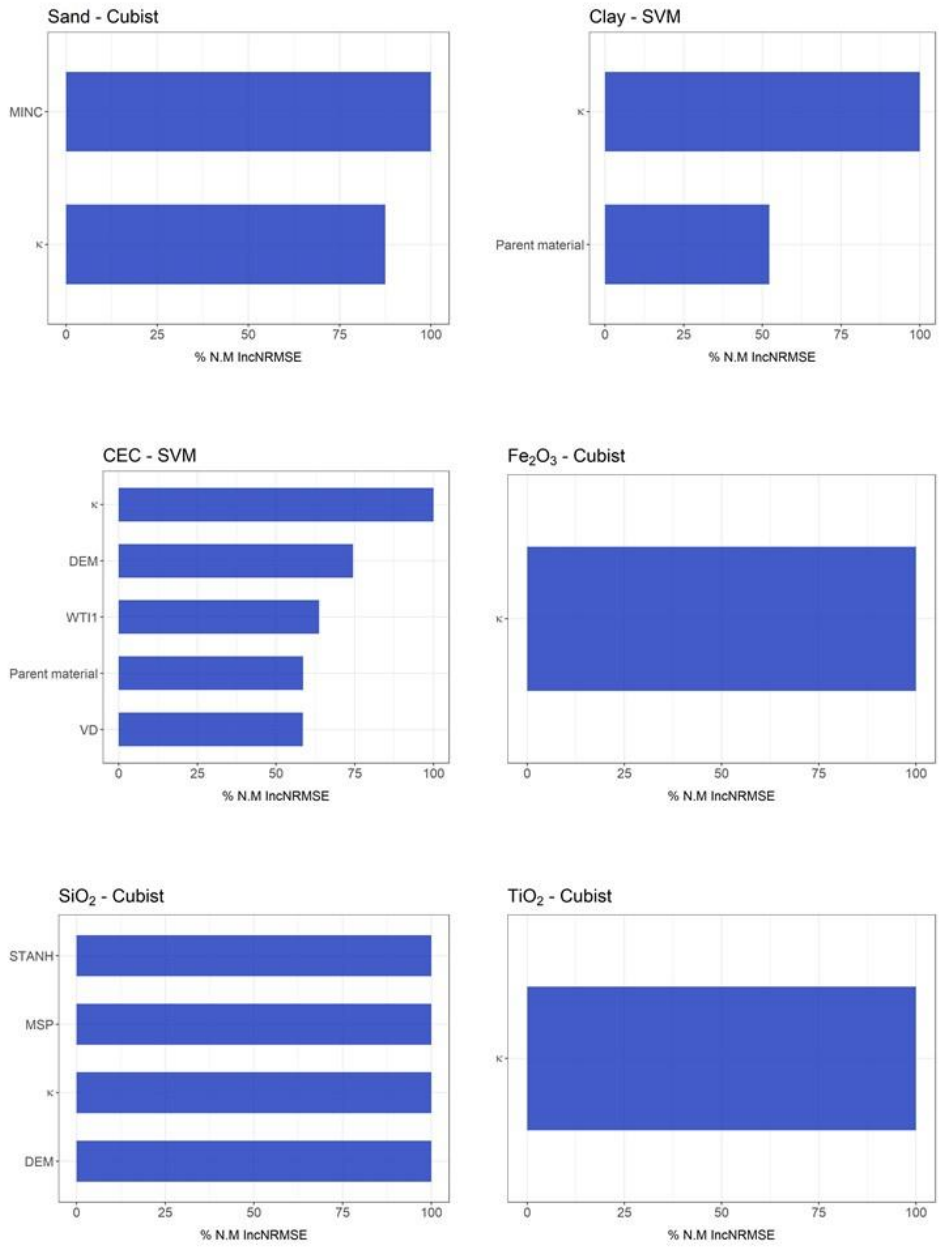
Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: $mmolc\ dm^{-3}$. Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: $mmolc\ dm^{-3}$. Support Vector Machines (SVM); Linear Models (LM).

Table 4. Model performance for the combined use of gamma-ray spectrometer and the conductivimeter, for all soil attributes based on R^2, RMSE, MAE and NULL_RMSE.

Gamma-ray spectrometer + Conductivimeter	***R²***				
	Random Forest	***Cubist***	***SVM***	***LM***	***-***
Clay	0.378	0.433	0.406	0.338	-
Sand	0.318	0.265	0.3	0.188	-
Fe2O3	0.22	0.282	0.158	0.249	-
TiO2	0.248	0.189	0.048	0.171	-
SiO2	0.16	0.163	0.17	0.178	-
CEC	0.14	0.077	0.241	0.002	-
BS	0.133	0.065	0.068	0.003	-
OM	0.001	0	0.059	0.047	-
Gamma-ray spectrometer + Conductivimeter	***RMSE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL RMSE***
Clay	137.097	134.231	134.035	146.116	140.885
Sand	179.808	197.657	182.644	225.909	176.521
Fe2O3	58.829	56.918	61.758	62.442	53.341
TiO2	11.011	12.026	13.076	13.035	10.239
SiO2	40.256	42.209	40.493	41.555	35.45
CEC	41.464	47.809	40.463	1499.11	36.139
BS	19.889	21.704	21.586	33.64	17.142
OM	8.567	8.356	7.72	7.738	6.158
Gamma-ray spectrometer + Conductivimeter	***MAE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL MAE***
Clay	108.636	105.954	106.779	117.816	119.751
Sand	145.511	160.722	148.469	181.07	153.803
Fe2O3	38.867	37.335	39.185	42.121	41.578
TiO2	7.265	8.241	8.197	9.198	8.074
SiO2	31.095	32.419	32.189	32.035	29.534
CEC	28.539	33.06	26.449	207.159	27.187
BS	15.812	17.471	17.325	24.294	14.425
OM	6.443	6.07	5.578	5.806	4.813

Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: $mmolc\ dm^{-3}$. Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: $mmolc\ dm^{-3}$. Support Vector Machines (SVM); Linear Models (LM).

For clay, the model with the best performance was the SVM algorithm (R^2 0.484) by S + C (**Table 3**), whereas that with the worst performance was the Cubist algorithm (R^2 0.38) by (G + S + C) (**Table 6**). For sand, the best model performance was obtained with the Cubist algorithm (R^2 0.365) by S + C (**Table 3**) and the worst also by Cubist (R^2 0.387) by (G + S + C). The most important covariates for clay prediction by the SVM model in S + C sensors combination were magnetic susceptibility (κ) (100%) and parent material (90%) (**Fig. 5**). For clay prediction by the Cubist model in G + S + C sensors combination, the most important covariate was parent material (100%) (**Fig. 6**). With respect to sand prediction, the most important covariates by the Cubist model in S + C were minimal curvature (100%) and magnetic susceptibility (κ) (80%) (**Fig. 5**) On the other hand, for G + S + C, the covariates that most contributed for sand prediction were DEM (100%), general curvature (80%) and minimal curvature (75%) (**Fig. 6**).



395 **Figure 5.** Variable importance for *Susceptibilimeter + Conductivimeter sensors* (only variables that contributed more than 50% are presented here (for further details see supplementary material).

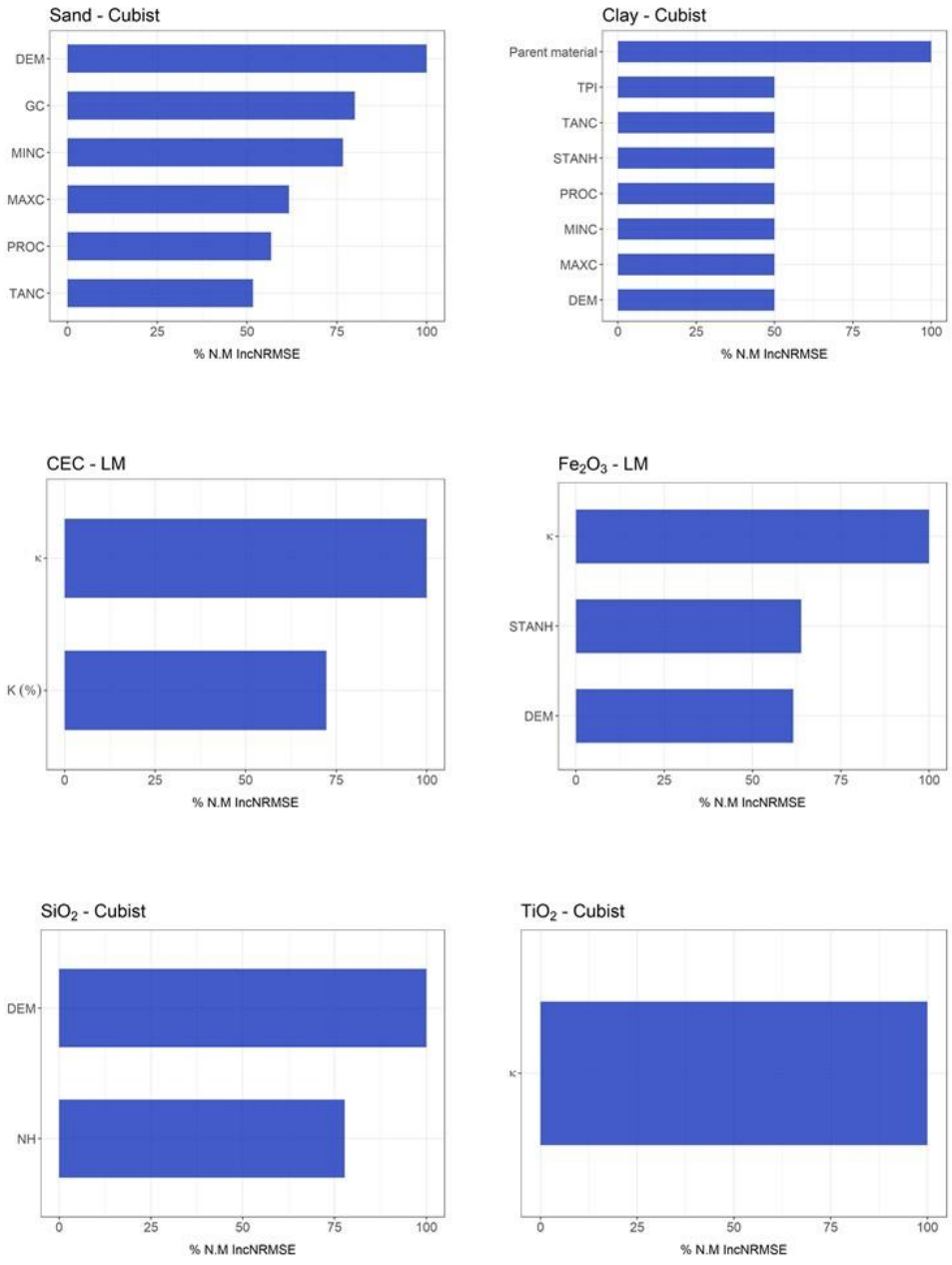
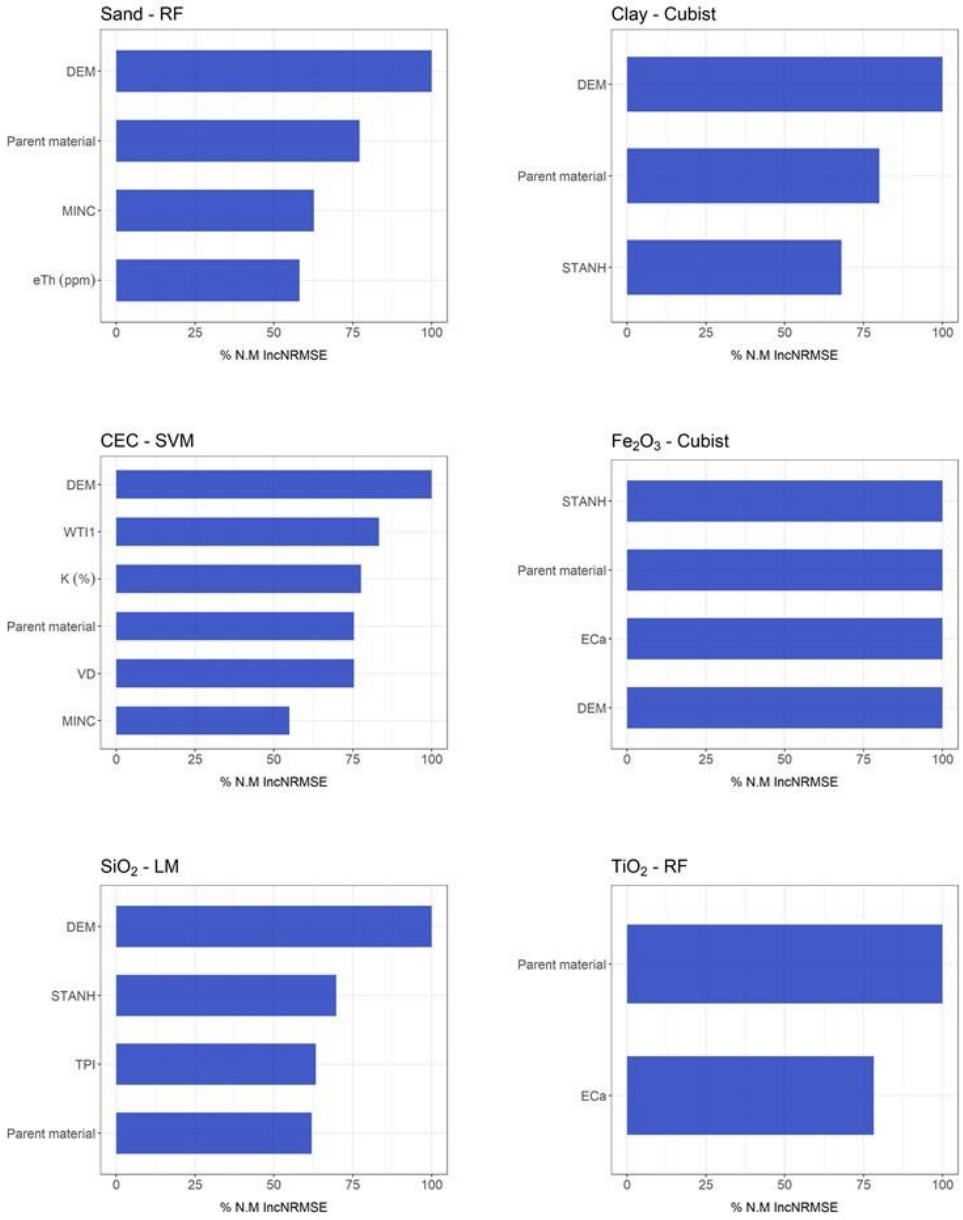


Figure 6. Variable importance for *Combined use of the three geophysical sensors* (only variables that contributed more than 50% are presented here (for further details see supplementary material).

For the elemental composition, the models employed greatly variable performance. For Fe_2O_3 the best model performance, was reached by the LM algorithm (R^2 0.441) by G + S + C (**Table 6**), while the worst performance was by the Cubist (R^2 0.282) by G + C (**Table 4**). With respect to TiO_2, the best model performance was by Cubist algorithm (R^2 0.358) by G + S + C (**Table 6**) and the worst was RF (R^2 0.248) by G + C (**Table 4**). For SiO_2, the best model performance was the Cubist algorithm (R^2 0.250) by S + C (**Table 3**) and the worst was the LM (R2 0.178) by G + C (**Table 4**). The importance of covariates in predicting Fe_2O_3 by LM in G + S + C, demonstrated that magnetic susceptibility (κ), standardized height and DEM were the most important variables, contributing 100%, 65%, 55%, respectively (**Fig. 6**). For Fe_2O_3 predicted by the Cubist algorithm by G + C, the most important covariates were standardized height, parent material, ECa and DEM (100%) (**Fig. 7**). For TiO_2 prediction by the Cubist algorithm by G +S + C the most important covariate was magnetic susceptibility (κ) (100%) (**Fig. 6**), while for the RF algorithm by G + C were parent material (100%) and ECa (75%) (**Fig. 7**). In relation to SiO_2 prediction by the Cubist by S + C, the most important covariates were standardized height, mid-slope position magnetic susceptibility (κ) and DEM (100%) (**Fig. 5**), while SiO2 predicted by the LM algorithm by G + C were DEM and standardized height (100% and 65%, respectively) to mean accuracy (**Fig. 7**).



415 **Figure 7.** Variable importance for *Gamma-ray spectrometer + Conductivimeter sensors* (only variables that contributed more than 50% are presented here (for further details see supplementary material).

In relation to CEC, the LM algorithm was the best model (R^2 0.317) by G + S + C (**Table 6**) and the worst was the SVM algorithm (R^2 0.223) by S + C (**Table 3**). The most important covariate for prediction of CEC by LM algorithm by G + S + C and by S + C was magnetic susceptibility (κ) (100%) (**Fig. 6 and 5**).

Overall, the best combination of geophysical sensors, which allowed the best model performance for different algorithms in the prediction of soil attributes, was Gamma-ray spectrometer + Susceptibilimeter (G + S) (**Table 5**).

Table 5. Model performance for combined use of gamma-ray spectrometer and susceptibilimeter, for all soil attributes, based on R^2, RMSE, MAE and NULL_RMSE.

Gamma-ray spectrometer + Susceptibilimeter	***R²***				
	Random Forest	***Cubist***	***SVM***	***LM***	***-***
Clay	0.465	0.441	0.494	0.366	-
Sand	0.422	0.152	0.367	0.233	-
Fe2O3	0.36	0.426	0.096	0.47	-
TiO2	0.308	0.282	0.284	0.328	-
SiO2	0.159	0.207	0.169	0.167	-
CEC	0.147	0.152	0.296	0.303	-
BS	0.169	0.082	0.112	0.002	-
OM	0.046	0.033	0.028	0.034	-
Gamma-ray spectrometer + Susceptibilimeter	***RMSE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_RMSE***
Clay	127.149	132.977	123.84	148.11	140.885
Sand	165.624	244.635	175.35	202.104	176.521
Fe2O3	53.418	52.737	67.759	48.513	53.341
TiO2	10.724	11.37	10.846	10.659	10.239
SiO2	40.898	40.244	42.207	42.993	35.45
CEC	41.902	44.296	38.723	37.645	36.139
BS	19.294	21.318	20.856	1024.32	17.142
OM	7.8	7.842	7.81	8.131	6.158
Gamma-ray spectrometer + Susceptibilimeter	***MAE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_MAE***
Clay	102.229	105.123	97.173	117.097	119.751
Sand	134.525	168.957	140.318	166.083	153.803
Fe2O3	33.284	32.411	42.282	33.124	41.578
TiO2	6.548	6.573	6.447	7.049	8.074
SiO2	30.394	29.691	30.396	32.951	29.534
CEC	28.977	30.945	25.376	25.815	27.187
BS	15.597	17.321	16.96	137.422	14.425
OM	5.805	5.836	5.966	6.262	4.813

Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: mmolc dm^{-3}. Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: mmolc dm^{-3}. Support Vector Machines (SVM); Linear Models (LM).

Table 6. Model performance for all combined use of geophysical sensors, for all soil attributes, based on R^2, RMSE, MAE and NULL_RMSE.

Combined use of the three geophysical sensors	***R²***				
	Random Forest	***Cubist***	***SVM***	***LM***	**-**
Clay	0.356	0.387	0.331	0.258	-
Sand	0.318	0.322	0.278	0.129	-
Fe2O3	0.281	0.406	0.309	0.441	-
TiO2	0.322	0.358	0.267	0.252	-
SiO2	0.162	0.212	0.21	0.125	-
CEC	0.171	0.266	0.246	0.317	-
BS	0.122	0.097	0.107	0.002	-
OM	0.003	0.073	0.002	0.047	-
Combined use of the three geophysical sensors	***RMSE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_RMSE***
Clay	139.61	139.41	144.532	160.894	140.885
Sand	180.339	188.745	189.768	256.078	176.521
Fe2O3	57.225	52.66	57.589	50.038	53.341
TiO2	10.472	10.547	11.053	11.499	10.239
SiO2	40.642	40.534	40.355	43.949	35.45
CEC	41.451	39.226	39.815	37.134	36.139
BS	19.951	21.749	21.178	1045.896	17.142
OM	8.234	7.569	8.134	7.752	6.158
Combined use of the three geophysical sensors	***MAE***				
	Random Forest	***Cubist***	***SVM***	***LM***	***NULL_MAE***
Clay	112.126	108.346	117.645	120.83	119.751
Sand	143.98	145.661	145.187	198.059	153.803
Fe2O3	35.597	32.751	35.387	34.724	41.578
TiO2	6.414	6.541	6.7	8.102	8.074
SiO2	30.215	30.197	30.001	33.649	29.534
CEC	29.014	27.169	26.201	25.273	27.187
BS	15.887	17.694	17.025	140.716	14.425
OM	6.223	5.854	5.945	5.798	4.813

Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: $mmolc\ dm^{-3}$. Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: $mmolc\ dm^{-3}$. Support Vector Machines (SVM); Linear Models (LM).

The best combination of sensors, resulting in the best model performance, was G + S. (**Table 5**). For soil texture, the SVM and RF algorithms showed the best performance for clay (R^2 0.494) and sand (R^2 0.422), respectively, by G + S, with the highest R^2 and lowest RMSE and MAE values (**Table 5**). The importance of covariates in predicting soil texture by the SVM (for clay) and the RF (for sand) demonstrated that magnetic susceptibility (κ) was the most important covariate (100%). In addition, parent material contributed 60% for clay prediction and DEM 60% for sand prediction (**Fig. 8**).

The LM algorithm presented the best performance for Fe_2O_3 (R^2 0.470) and TiO_2 (R^2 0.328), by G + S, whereas for SiO_2, the Cubist algorithm was most suitable (R^2 0.207), also by G + S (**Table 5**). The most important covariates for Fe_2O_3 and TiO_2 prediction via LM by G + S were magnetic susceptibility (κ) and standardized height (100 and 60%, respectively, for both) (**Fig. 8**). For SiO2 prediction via the Cubist algorithm by G + S, the most important covariates were mid-slope position and magnetic susceptibility (κ) (100% for both) (**Fig. 8**).

For CEC, the best model performance was obtained using the LM algorithm (R^2 0.303) by G + S (**Table 5**). In this case, the covariates that most contributed to model prediction were magnetic susceptibility (κ) (100%) and DEM (60%) (**Fig. 8**).

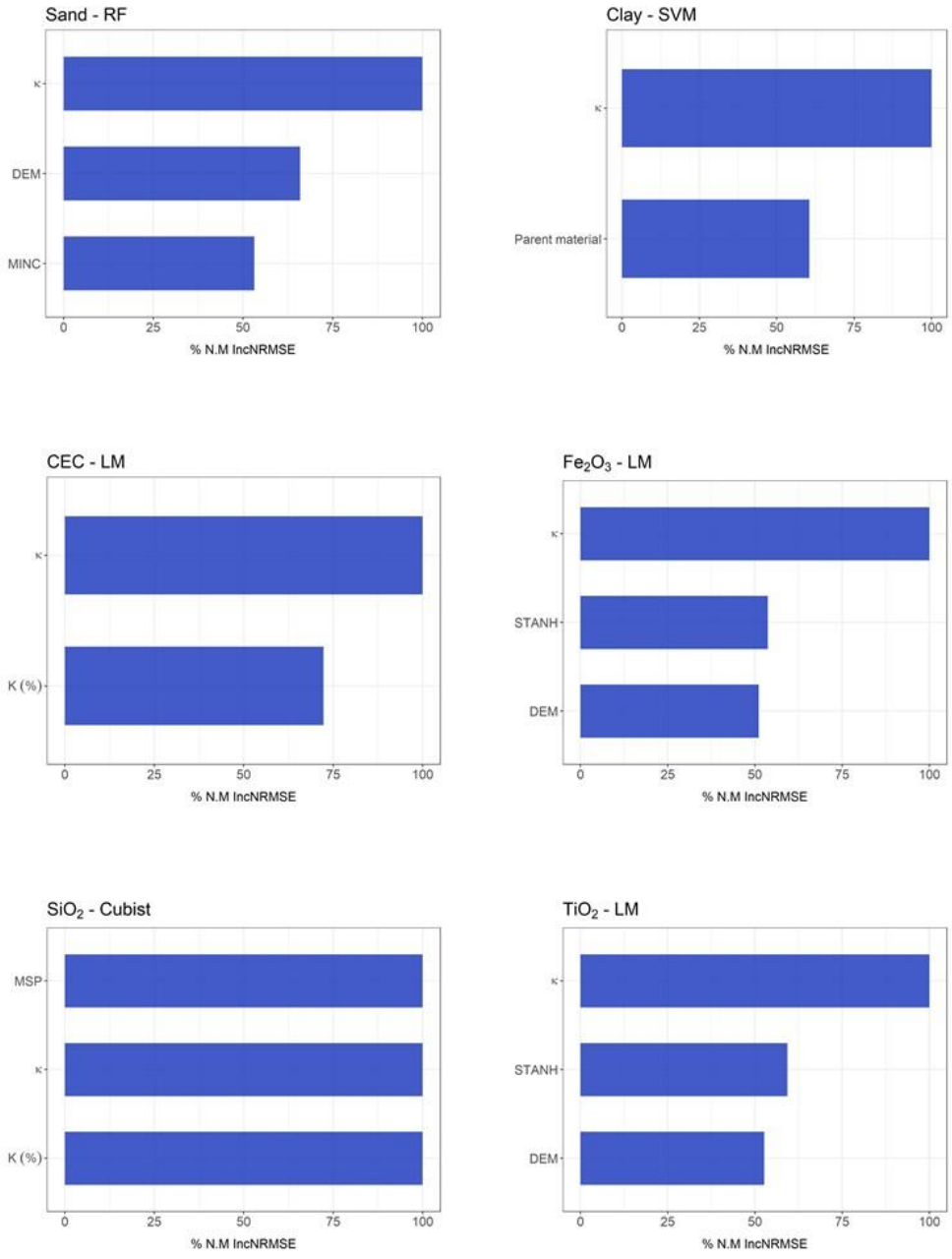


Figure 8. Variable importance for *Gamma-ray spectrometer + Susceptibilimeter sensors* (only variables that contributed more than 50% are presented here (for further details see supplementary material).

4 Discussion

4.1. Geophysical sensor combinations, models performance, and uncertainty

The methodological approach optimized the prediction of soil variables by applying different geophysical sensors combinations, parent material and terrain attributes for selecting covariates and models, as well as for assessing prediction uncertainty.

In general, without the use of geophysical sensors, the poorest results were obtained in terms of R^2, RMSE, and MAE for all prediction algorithms used for modeling soil attributes (**Table 2**). These results are consistent with Frihy et al. (1995), who also compared the combined use and the non-use of sensors regarding model geochemical attributes of soil by the Cubist algorithm and obtained the worst results without using the sensors. Most likely, this is a result of the highly complex interaction between soil forming factors and processes determining soil attributes (Jenny, 1994).

The moderate performance of the models can be attributed to the different combinations of the geophysical sensors pairwise, and the different data presented by the sensors contributed in different ways to the modelling process. In this regard, O'Rourke et al. (2016) also demonstrated a moderate performance of the models (R^2 ranging from 0.21 to 0.94) when using data from the VisNir, with R^2 ranging from 0.61 to 0.94 when using the pXRF sensor to model soil attributes. This might be related to the different sensors and, their relation with soil attributes. The VisNIR spectroscopy acts on targets with low energy levels, showing the ability to identify soil mineral species, strongly linked to soil attributes (Coblinski et al., 2021). In addition, pXRF spectroscopy allows the identification of total elementary contents by acting with high levels of ionizing energy, which is not identified by Vis-NIR, and is strongly correlated with minerals and soil attributes (Silvero et al., 2020). Therefore, the addition of pXRF with Vis-NIR data for obtaining information about soil constituents is highly efficient for modeling soil attributes.

The best combination of geophysical sensors was Gamma-ray spectrometer + Susceptibilimeter (G+S), with the highest values of R^2 and the lowest values of RMSE and MAE (**Table 5**). Most likely, the gamma-ray spectrometer and the susceptibilimeter are more closely associated with pedogenesis (argilluviation, ferralitization and others), pedogeomorphology, and soil attributes, as recently demonstrated by Mello et al. (2020); Mello et al. (2021), who modeled soil attributes such as texture, Fe_2O_3, TiO_2, SiO_2 and CEC in relation to thorium, uranium and potassium (K^{40}) levels as well as magnetic susceptibility.

In general, the Cubist algorithm was the best model for clay and sand content prediction (**Table 7**). Similar results have been found by Greve and Malone (2013); Ballabio et al. (2016); Nawar et al. (2016), and Silva (2019), who used the Cubist and Earth algorithm to predict soil texture using different data sources (3D imagery, Land Use and Cover Area frame Statistical survey, and reflectance spectroscopy). In all these models, the R^2 was not greater than 0.5, which can be explained by the small variation or limited distribution of the data set, causing poor modeling prediction. Zhang and Hartemink (2020) state that textural classes with fewer samples presented a more unstable prediction performance than those with more samples, which agrees with our results.

Table 7. Number of times that each model achieved the best performance for each soil attribute

Soil attributes	R^2			
	Random Forest	***Cubist***	***SVM***	***LM***
Clay		3	2	
Sand	2	3		
Fe_2O_3		2	1	2
TiO_2	1	2	1	2
SiO_2		3		1
CEC			3	2

Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: mmolc dm^{-3}. Clay and sand content in $g.kg^{-1}$; Fe_2O_3, TiO_2 and SiO_2 in $g.kg^{-1}$ CEC in $mmol_c\ dm^{-3}$; abbreviations: CEC: Cation Exchange Capacity; OM $g.dm^{-3}$; BS: mmolc dm^{-3}. Support Vector Machines (SVM); Linear Models (LM).

The better performance for elemental composition (Fe_2O_3, TiO_2 and SiO_2) was obtained using the Cubist algorithm, (**Table 7**), with an R^2 of0.2–0.47. This is contrasting with the results obtained by Henrique et al. (2018), who showed that the best model for predicting soil mineralogy Fe_2O_3 and TiO_2 (R^2 0.89 and 0.96, respectively) and RF only for Fe_2O_3 (R^2 0.95) by pXRF was the simple linear regression. In our study, the R^2 variation for the G + S combination was probably related to the low correlation with the parent material and, consequently, with soil mineralogy or to the limited number of samples and the high soil variability (Fiorio, 2013). However, it is important to highlight that *in situ*, various intrinsic environmental influences can interfere with modelling processes. For example, the relatively low R^2 values (approximately between 0.2 and 0.5) can be attributed to the difficulty in modeling soils and their attributes. This is related to the high complexity of soils, such as, the high spatial variability in surface and depth, the occurrence of geomorphic processes, weathering, and pedogenesis, and the different soil formation factors. For soil mineralogical attributes predicted by machine learning algorithms, the results can be classified as satisfactory from 0.2 to 0.5, as for the preliminary evaluation, since these values represent more informative results (Beckett, 1971; Dobos, 2003; Malone et al., 2009). According to Nanni and Demattê (2006), the R^2 may be explained by standardized laboratory conditions (such as temperature, humidity, substance concentrations, and other variables that interfere with the analysis results during their determination), with less environmental interference compared with direct field methods. For CEC, the best model performance was obtained for SVM (R^2 0.296) (**Table 5**). This result is corroborated by Liao et al. (2014), who compared the model performance of multiple stepwise regression, artificial neural network models, and SVM for CEC prediction and attributed their results to a nonlinear relationship between CEC and soil physicochemical properties. In

addition, in our previous study (Jafarzadeh et al., 2016), we demonstrated that, despite of the ability of SVM to predict CEC in acceptable limits, there is a poor performance in extrapolating the maximum and minimum values of CEC data. Despite this, uncertainties estimated for SVM predictions may not be associated with an incorrect classification, as pointed out by Cracknell and Reading (2013).

Even for the best combination of sensors (G + S) and the highest overall model performance, the R^2 values were not greater than 0.5 (**Table 5**). In models generated by field data, without sample preparation, R^2 values varying between 0.20 and 0.50 can be considered satisfactory and reliable (Dobos, 2003; Malone et al., 2009). In our study, the low R^2 values can be related to the limited number of collecting points or to the low field distribution, which does not represent the spatial variation of soil attributes; this is in agreement with Johnston et al. (1997) and Lesch et al. (1992), who evaluated soil salinity.

The best results for predictors of soil attributes through geophysical data have the lowest values when compared to the values of NULL_RMSE and NULL_MAE. This demonstrates that the use of machine learning models has less errors than the use of mean values for the entire area (**Table 5**), resulting in a better performance and accuracy.

The null model is a simple model (naive) that expresses the value of the mean of the Y (variable to be predicted or target variable). The RMSE and MAE values are calculated for the null model and further compared with MAE and RMSE values calculated by other models. If the RMSE and MAE values from other models present similar or worse performance than the null model, the model that compared it is not an informative model. In this case, it is better to choose a simple mean as a predictor rather than using a more complex model to explain a given phenomenon. The null model sets a minimum performance threshold to be reached by models (Kuhn et al., 2020); however, there are only few studies using NULL_RMSE and NULL_MAE as parameters for model evaluation and decision making.

4.1.2 Variables importance, model performance, and pedogeomorphology

In general, for all geophysical sensor combinations, the majority of terrain attributes used did significantly influence sand and clay content prediction (**Figs. 4, 5, 6, and 8**). However, in most cases, parent material and magnetic susceptibility strongly influenced clay content prediction, except for G + C (**Fig. 7**). Ließ et al. (2012) found that the best performance was obtained using the RF model, with elevation and overland flow distance strongly affecting the model performance. According to Bauer (2010), the greater sand/clay ratio upslope is explained by the selective transport of fine material downslope, whereas in the present study, the clay content increased because of the influence of parent material (diabase), as also demonstrated by Mello et al. (2020).

Magnetic susceptibility (κ), followed by DEM and parent material, were the key variables that contributed to sand and clay content prediction by RF and SVM, respectively, for G + S (**Fig. 8**). Siqueira et al. (2010) and Mello et al. (2020) found a positive correlation between soil magnetic susceptibility and clay content and a negative correlation between magnetic susceptibility and sand content. In fact, the mineralogical composition of the parent material strongly affects soil magnetic susceptibility (Ayoubi et al., 2018), mainly in tropical soils under the top of basalt spills (Da Costa et al., 1999), where our study was undertaken.

In general, for Fe_2O_3 and TiO_2, the most important variables were parent material, magnetic susceptibility, and DEM, which, in most cases, contributed 100% (**Figs. 4, 5, 6, 7, and 8**). In fact, the mineralogical composition of the parent material and the pedo-environmental conditions strongly influence the amount of Fe/Ti oxides in soils (Schwertmann and Taylor, 1989; Kämpf and Curi, 2000; Bigham et al., 2002) and accelerate redistribution by downslope erosion (Mello et al., 2020). Also, the mineralogical composition of the parent material (Mullins, 1977; Ayoubi et al., 2018) and the landform evolution (Blundell et al., 2009; Sarmast et al., 2017) control the magnetic susceptibility of soil. Since the sensors used record the surface response and topography effect, it is expected that the most important variables indicated by the models would be related to surface processes. For the best combination of sensors (G + S), magnetic susceptibility and standardized height were more important variables in the prediction of Fe_2O_3 (100%) and TiO_2 (55%) contents (**Fig. 8**), corroborating the expected surface processes and materials in the magnetic susceptibility of the soil (Shenggao, 2000; Damaceno et al., 2017) and the relief in the distribution of these materials (De Jong et al., 2000).

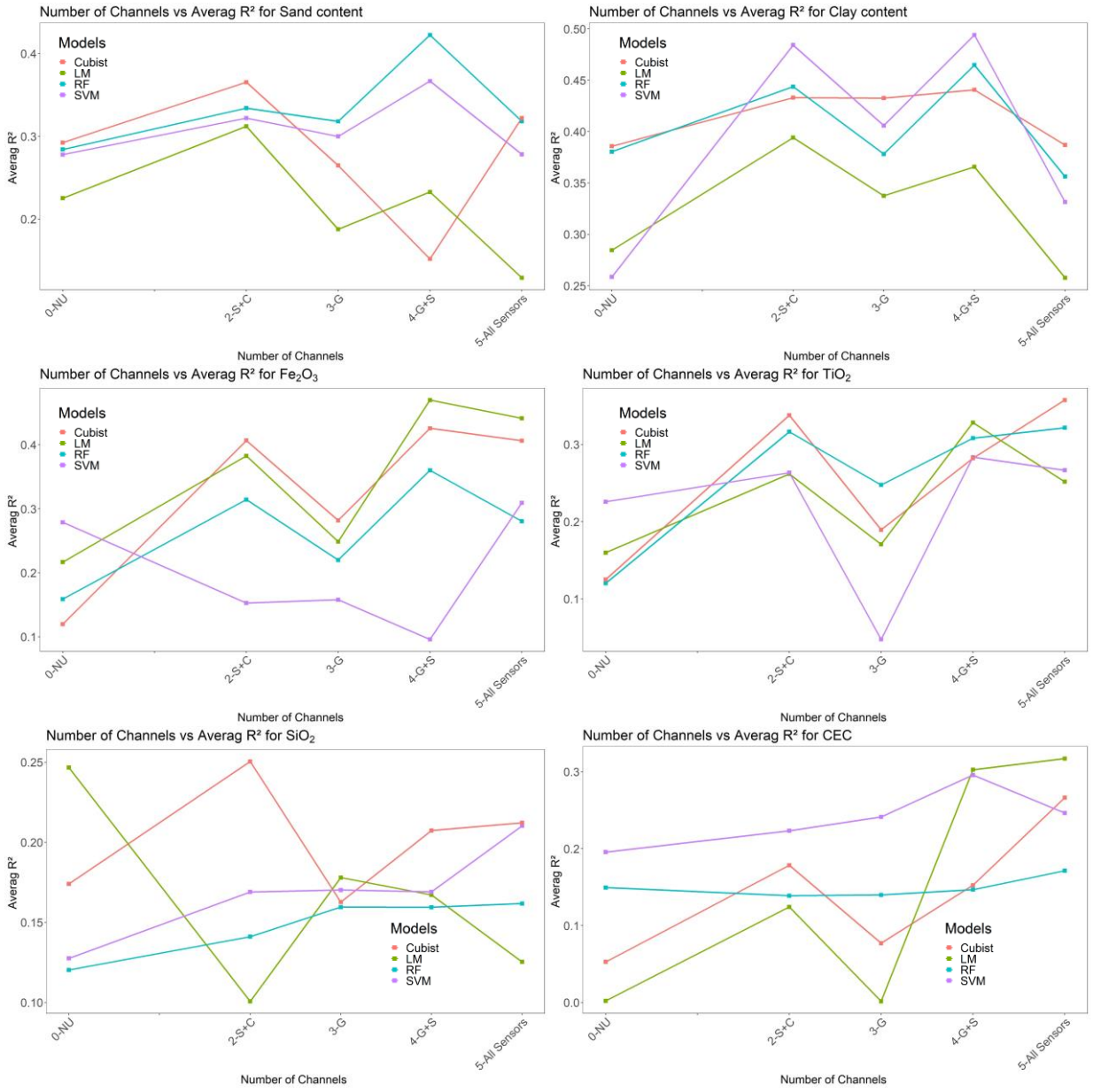
For SiO_2, the most important variable was DEM, which, in most of cases, contributed 100% (**Figs. 4, 5, 6, and 7**). The level of SiO_2 in soil is directly related to the nature of the parent material and the erosion processes at different topographic positions (Bockheim et al., 2014; Breemen and Buurman, 2003). This can explain the greater contribution of the DEM in the prediction models. For the best sensor combination (G + S), the variable that most contributed was mid-slope position, which also is related to topographic features.

For CEC, the variables DEM and magnetic susceptibility were the most important ones, contributing 100% in most of the cases (**Figs. 4, 5, 6, 7 and 8**). This can be explained by the high correlation between magnetic susceptibility, clay content, and CEC (Siqueira et al., 2010; de Souza Bahia et al., 2017; Mello et al., 2020). These variables vary with parent material and surface geomorphic processes, concentrating ferrimagnetic minerals (Frihy et al., 1995; Mello et al., 2020).

Considering that the gamma spectrometer sensor is composed of three channels (eU, eTh, and K^{40}), it can be called "three sensors". Thus, considering the combination of sensors used, it is possible to create a modeling performance graph using the number of sensors used through learning curves (**Fig. 9**). Such a learning curve shows a measure of the predictive performance of a given domain as a function of some measurements of varying amounts of learning effort (Perlich, 2010). In our case, the varying amounts were the number of sensors: non-use of geophysical sensors (0 sensors), S + C (two sensors), G + S (four sensors), and S + G + C (five sensors). In this analysis, the combination of G + C sensors will not be used because they present the same number of G + S sensors (four sensors). However, the combination G + C presented lower results than G + S.

For five soil properties (clay, sand, CEC, Fe_2O_3, and SiO_3), the best results did not occur with a greater number of sensors, showing that increasing the number of covariables can lead to a lower performance (**Fig. 9**). This fact is associated with the addition of a new sensor as a covariate, which may provide conflicting information for the set of the other sensors found, where the ECa may have presented conflicting values with the sensors generated by the gamma spectrometry channels, which generates a loss of performance when with sensors are combined. The application of the RFE importance selection method was able to amortize this, making it a reliable method to reduce this effect.

Figure 9. Learning curves calculated on the metric by which the parameters of the model were optimized and on the metric by which the model was evaluated and selected. The most common form of learning curves in the general field of machine learning shows predictive accuracy on the test examples as a function of the number of training examples (Perlich, 2010).



0-NU: Non-use of geophysical sensors; 2-S+C: (2 channels corresponding to susceptibilimeter + conductivimeter); 3-G: (3 channels corresponding to eU, eTh and K^{40} from gamma-ray); 4-G+S: (4 channels corresponding to eU, eTh and K^{40} from gamma-ray spectrometer + susceptibilimeter).

4.1.3 General evaluation

For this study, the independent RMQS data set was not large enough (75 sites). Therefore, validation using 74 sites provided erratic and inconsistent results, mainly when comparing different pedo-environmental indicators, even considering that this dataset, in theory, provides "unbiased" estimates of forecast performance (Loiseau et al., 2020). Similarly, Lagacherie et al. (2019) showed that the location and number of samples used for independent assessment can significantly impact the values of these indicators. This indicates that the greatest variations were observed for evaluation sets with less than 100 samples.

Modeling soil attributes using relief and geophysical data presented promising results for geosciences studies and soil scientists. The use of several algorithms from different "families", as well as the training and validation method, also made the study more robust and more reliable. In addition, machine learning models allow to define the importance of covariates, which is, sometimes, not possible when using ordinary spatialization methods, such as kriging and the inverse square of distance.

The "nested leave-one-out validation" method was useful with small sample sizes, being a potential tool to be used in geosciences studies. However, the academic community still knows little about the potential applicability of machine learning techniques.

5. Conclusions

It is possible to model soil attributes satisfactorily, with easily acquired input data (parent material + DEM) combined with data sets from different geophysical sensors. In addition, geophysical data from proximal sensors coupled with Cubist algorithms can provide accurate estimates for several soil attributes. This may reduce the need for new soil samples and wet chemistry methods.

The combination of geophysical sensors with the best model performance (higher R^2 and lower RMSE and MAE, concomitantly) for the prediction of soil attributes was Gamma-ray spectrometer + Susceptibilimeter (G + S). For this combination of sensors, the R^2 values were 0.494 (clay), 0.422 (sand), 0.470 (Fe_2O_3), 0.328 (TiO_2), 0.207 (SiO_2) and 0.303 (CEC) for the SVM, RF, LM, Cubist, and LM algorithms, respectively. The simultaneous use of three sensors did not optimize model performance. On the other hand, when the geophysical sensors were not used, soil attribute prediction by machine learning algorithms was less reliable.

In general, the algorithms showed varying performance levels. The Cubist algorithm was most suitable for clay, sand, Fe_2O_3, TiO_2, and SiO_2. For CEC, the best performance was obtained by SVM. The second-best algorithm performance observed using SVM for clay, RF for sand, and LM for Fe_2O_3, TiO_2, SiO_2, and CEC.

For soil attributes, we obtained R^2 values greater than 0.2, which are considered satisfactory for machine learning algorithms applied to field data without expensive laboratory analysis, especially when compared with data from fieldwork with the use of remote sensing covariates. All soil attributes were more reliably predicted considering an average value for the entire area. The use of the null model methodology provided a way of comparing the values generated by machine learning when it is not possible to use other methods. The use of four algorithms proved necessary since at least one of the soils attributes performed better in each of the tested algorithms.

The use of nested-LOOCV method was appropriate to be used in geoscience and soil science for modeling using a database with a small number of samples. In addition, the nested-LOOCV approach proved to be a robust method to evaluate the algorithm's performance, allowing concomitantly the optimisation and increasing the efficiency of training and testing of models.

The final model was more parsimonious, with an ideal number of covariates with a three-step selection. This reduced the effect of overfitting by the use of a large number of covariates. Also, the nested leave-one-out validation methodology proved to be appropriate for a small number of samples when compared to hold-out validation and cross-validation.

The covariables that most contributed to the prediction of soil attributes (clay, sand, Fe_2O_3, TiO_2, SiO_2, and CEC), in most of the algorithms used and sensor combinations, were DEM, magnetic susceptibility, parent material, and standardized height.

For each study area, a conceptual pedogeomorphological and geophysical model must be created due to the complex interaction among environmental variables, pedogenesis and soil attributes. These factors affect the geophysical variables which are detected and quantified by the sensors and will later serve as input data for the modeling processes.

The machine learning technique is a potential tool for modelling soil attributes with geophysical data when only field data with proximal sensors are available. The combined use of gamma-ray spectrometer and susceptibilimeter, allowed for an optimization of the models.

6. Authors contribution

Danilo César de Mello: conceived of the presented idea, carried out the experiment, developed the theoretical formalism, contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. He provided critical feedback and helped shape the research, analysis and manuscript.

Gustavo Vieira Veloso: designed the model and the computational framework and analysed the data, planned and carried out the simulations, performed the analytic calculations and performed the numerical simulations, modelling processing, evaluate algorithms performance, variables importance and statistical analyses.

Marcos Guedes de Lana: contributed to the interpretation of the results, took the lead in writing the manuscript. Devised the project, the main conceptual ideas and proof outline. He worked out almost all of the technical details. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Fellipe Alcantara de Oliveira Mello: contributed to the interpretation of the results, took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Raul Roberto Poppiel: contributed to the interpretation of the results, took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Diego Ribeiro Oquendo Cabrero: performed the analysis, drafted the manuscript and designed the figure. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Luis Augusto Di Loreto Di Raimo: performed the analysis, drafted the manuscript and designed the figure. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Carlos Ernesto Gonçalves Reynaud Schaefer: Critical revision of the article. All authors discussed the results and commented on the manuscript. He contributed to the interpretation of the results and verified the analytical methods.

Elpídio Inácio Fernandes Filho: Critical revision of the article. He designed the model and the computational framework and analysed the data. He contributed to the interpretation of the results and verified the analytical methods. All authors discussed the results and commented on the manuscript.

Emilson Pereira Leite: Critical revision of the article. He contributed to the interpretation of the results and verified the analytical methods. All authors discussed the results and commented on the manuscript.

José Alexandre Melo Demattê: Provided de financial support, leadership of the group, critical revision of the article. He contributed to the interpretation of the results and verified the analytical methods. Encouraged the co-authors to investigate a specific aspect and supervised the findings of this work.

7. Code and data availability

All analyzes and codes used in this research were developed in "R software" version 4.0.3, (R Development Core Team, 2020) (Kuhn et al., 2020).The codes and data used in this research can be found at https://zenodo.org/record/5733366#.YaTXa9DMKUk (DOI: 10.5281/zenodo.5733366). All packages used in "R software", as well as their respective versions are listed in the database and codes available in the *data_base.zip* in the indicated repository.

8. Acknowledgements

We would like to thank the National Council for Scientific and Technological Development (CNPq) for the first author scholarship (grant No. 134608/2015-1); the São Paulo Research Foundation (FAPESP) (grant No. 2014-22262-0) for providing essential resources to the Laboratory of Remote Sensing Applied to Soils from "Luiz de Queiroz" College of Agriculture (ESALQ/USP); the Geotechnologies in Soil Science group (GeoSS – website http://esalqgeocis.wixsite.com/english) and LabGeo – UFV - 'Post Graduation Program in Soil and Plant Nutrition – PGSNP' of the Soil Department of Federal University of Viçosa, Brazil; Institute of Geosciences at Campinas State University, for the support.

9. References

Agbu, P. A., Fehrenbacher, D. J. and Jansen, I. J.: Soil property relationships with SPOT satellite digital data in east central Illinois, Soil Sci. Soc. Am. J., 54(3), 807–812, 1990.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., De Moraes Gonçalves, J. L. and Sparovek, G.: Köppen's climate classification map for Brazil, Meteorol. Zeitschrift, 22(6), 711–728, doi:10.1127/0941-2948/2013/0507, 2013.

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E. and Sparks, D. L.: Soil and human security in the 21st century, Science (80-.)., 348(6235), 2015.

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A. B., McKenzie, N. J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I. O. A., Sanchez, P. A., Thompson, J. A. and Zhang, G.-L.: GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties, in Advances in Agronomy 125, pp. 93–134., 2014.

Ayoubi, S., Abazari, P. and Zeraatpisheh, M.: Soil great groups discrimination using magnetic susceptibility technique in a semi-arid region, central Iran, Arab. J. Geosci., 11(20), doi:10.1007/s12517-018-3941-4, 2018.

Bai, W., Kong, L. and Guo, A.: Effects of physical properties on electrical conductivity of compacted lateritic soil, J. Rock Mech. Geotech. Eng., 5(5), 406–411, doi:10.1016/j.jrmge.2013.07.003, 2013.

Ballabio, C., Panagos, P. and Monatanarella, L.: Mapping topsoil physical properties at European scale using the LUCAS database, Geoderma, 261, 110–123, doi:10.1016/j.geoderma.2015.07.006, 2016.

Barbuena, D., de Souza Filho, C. R., Leite, E. P., Miguel Jr, E., de Assis, R. R., Xavier, R. P., Ferreira, F. J. F. and Paes de Barros, A. J.: Airborne geophysical data analysis applied to geological interpretation in the Alta Floresta Gold Province, MT, Rev. Bras. Geofísica, 2013.

Batty, M. and Torrens, P. M.: Modelling complexity: the limits to prediction, Cybergeo Eur. J. Geogr., 2001.

Bauer, F. C.: Water flow paths in soils of an undisturbed and landslide affected mature montane rainforest in South Ecuador, Bayreuth, Alem., 2010.

Bazaglia Filho, O., Rizzo, R., Lepsch, I. F., Prado, H. do, Gomes, F. H., Mazza, J. A. and Demattê, J. A. M.: Comparison between detailed digital and conventional soil maps of an area with complex geology, Rev. Bras. Ciência do Solo, 37(5), 1136–1148, doi:10.1590/s0100-06832013000500003, 2013.

Beamish, D.: Gamma ray attenuation in the soils of Northern Ireland, with special reference to peat, J. Environ. Radioact., 115, 13–27, doi:10.1016/j.jenvrad.2012.05.031, 2013.

Beamish, D.: Relationships between gamma-ray attenuation and soils in SW England, Geoderma, 259–260, 174–186, doi:10.1016/j.geoderma.2015.05.018, 2015.

Beckett, P. H. T.: Soil variability: a review, Soils Fertil., 34(1), 1–15, 1971.

Bigham, J. M., Fitzpatrick, R. W. and Schulze, D. G.: Iron oxides, Soil Mineral. with Environ. Appl., 7, 323–366, 2002.

Blundell, A., Dearing, J. A., Boyle, J. F. and Hannam, J. A.: Controlling factors for the spatial variability of soil magnetic susceptibility across England and Wales, Earth-Science Rev., 95(3–4), 158–188, doi:10.1016/j.earscirev.2009.05.001, 2009.

Bockheim, J. G., Gennadiyev, A. N., Hartemink, A. E. and Brevik, E. C.: Soil-forming factors and Soil Taxonomy, Geoderma, 226–227(1), 231–237, doi:10.1016/j.geoderma.2014.02.016, 2014.

Breemen, Nico and Buurman, P.: Soil Formation, 2 nd., Laboratory of Soil Science and Geology, New YorkK, Boston, Dordrecht, London, Moscow., 2003.

Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A. and Edwards, T. C.: Geoderma Machine learning for predicting soil classes in three semi-arid landscapes, Geoderma, 239–240, 68–83, doi:10.1016/j.geoderma.2014.09.019, 2015.

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 do estado de São Paulo, Bol. técnico, 106, 94, 1986.

Camargo, L. A., Marques Júnior, J., Pereira, G. T. and Bahia, A. S. R. de S.: Clay mineralogy and magnetic susceptibility of Oxisols in geomorphic surfaces, Sci. Agric., 71(3), 244–256, doi:10.1590/S0103-90162014000300010, 2014.

Cardoso, R. and Dias, A. S.: Study of the electrical resistivity of compacted kaolin based on water potential, Eng. Geol., 226(January), 1–11, doi:10.1016/j.enggeo.2017.04.007, 2017.

César de Mello, D., Demattê, J. A. M., Silvero, N. E. Q., Di Raimo, L. A. D. L., Poppiel, R. R., Mello, F. A. O., Souza, A. B., Safanelli, J. L., Resende, M. E. B. and Rizzo, R.: Soil magnetic susceptibility and its relationship with naturally occurring processes and soil attributes in pedosphere, in a tropical environment, Geoderma, 372, doi:10.1016/j.geoderma.2020.114364, 2020.

Clevers, J. G. P. W., Van Der Heijden, G. W. A. M., Verzakov, S. and Schaepman, M. E.: Estimating grassland biomass using SVM band shaving of hyperspectral data, Photogramm. Eng. Remote Sensing, 73(10), 1141–1148, doi:10.14358/PERS.73.10.1141, 2007.

Coblinski, J. A., Inda, A. V., Demattê, J. A. M., Dotto, A. C., Gholizadeh, A. and Giasson, É.: Identification of minerals in subtropical soils with different textural classes by VIS–NIR–SWIR reflectance spectroscopy, CATENA, 203, 105334, 2021.

Correia, M. G., Leite, E. P. and de Souza Filho, C. R.: Comparação de métodos de estimativa de profundidades de fontes magnéticas utilizando dados aeromagnéticos da província mineral de Carajás, Pará, Brazilian J. Geophys., 28(3), 411–426, 2010.

Corwin, D. L., Lesch, S. M., Shouse, P. J., Soppe, R. and Ayars, J. E.: Identifying Soil Properties that Influence Cotton Yield Using Soil Sampling Directed by Apparent Soil Electrical Conductivity, , (1995), 352–364, 2003.

Da Costa, A. C. S., Bigham, J. M., Rhoton, F. E. and Traina, S. J.: Quantification and characterization of maghemite in soils derived from volcanic rocks in southern Brazil, Clays Clay Miner., 47(4), 466–473, doi:10.1346/CCMN.1999.0470408, 1999.

Cracknell, M. J. and Reading, A. M.: The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines, Geophysics, 78(3), WB113–WB126, doi:10.1190/GEO2012-0411.1, 2013.

Damaceno, J. G., de Castro, D. L., Valcácio, S. N. and Souza, Z. S.: Magnetic and gravity modeling of a Paleogene diabase plug in Northeast Brazil, J. Appl. Geophys., 136, 219–230, doi:10.1016/j.jappgeo.2016.11.006, 2017.

Darst, B. F., Malecki, K. C. and Engelman, C. D.: Using recursive feature elimination in random forest to account for correlated variables in high dimensional data, BMC Genet., 19(1), 65, 2018.

Demattê, J. A. M., Galdos, M. V, Guimarães, R. V, Genú, A. M., Nanni, M. R. and Zullo Jr, J.: Quantification of tropical soil attributes from ETM+/LANDSAT-7 data, Int. J. Remote Sens., 28(17), 3813–3829, 2007.

Demattê, J. A. M., Horák-Terra, I., Beirigo, R. M., Terra, F. da S., Marques, K. P. P., Fongaro, C. T., Silva, A. C. and Vidal-Torrado, P.: Genesis and properties of wetland soils by VIS-NIR-SWIR as a technique for environmental monitoring, J. Environ. Manage., 197, 50–62, doi:10.1016/j.jenvman.2017.03.014, 2017.

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

Dickson, B. L. and Scott, K. M.: Interpretation of aerial gamma-ray surveys - adding the geochemical factors, AGSO J. Aust. Geol. Geophys., 17(2), 187–200, 1997.

Dobos, E.: The application of remote sensing and terrain modeling to soil characterization, Innov. Soil-Plant Syst. Sustain. Agric. Pract., 328–348, 2003.

Domsch, H. and Giebel, A.: Estimation of soil textural features from soil electrical conductivity recorded using the EM38, Precis. Agric., 5(4), 389–409, doi:10.1023/B:PRAG.0000040807.18932.80, 2004.

Dragovic, S. and Onjia, A.: Classification of soil samples according to geographic origin using gamma-ray spectrometry and pattern recognition methods, Appl. Radiat. Isot., 65, 218–224, doi:10.1016/j.apradiso.2006.07.005, 2007.

EMBRAPA: Documentos 132 Manual de Métodos de, Embrapa, (ISSN 1517-2627), 230, 2011.

EMBRAPA: Manual de metodos de analises., 2017.

Farzamian, M., Monteiro Santos, F. A. and Khalil, M. A.: Application of EM38 and ERT methods in estimation of saturated hydraulic conductivity in unsaturated soil, J. Appl. Geophys., 112, 175–189, doi:10.1016/j.jappgeo.2014.11.016, 2015.

Ferreira, R. G., da Silva, D. D., Elesbon, A. A. A., Fernandes-Filho, E. I., Veloso, G. V., de Souza Fraga, M. and Ferreira, L. B.: Machine learning models for streamflow regionalization in a tropical watershed, J. Environ. Manage., 280, 111713, 2021.

Filho, B.: Universidade de São Paulo Escola Superior de Agricultura “ Luiz de Queiroz ” Comparação entre mapas de solos obtidos pelos métodos convencional e digital numa área complexa Osmar Bazaglia Filho Piracicaba, Master Diss. Soils Plant Nutr. 190 p, 190, 2012.

Fioriob, P. R.: Estimation of Soil Properties by Orbital and Laboratory Reflectance Means and its Relation with Soil Classification, Open Remote Sens. J., 2(1), 12–23, doi:10.2174/187541390100201012, 2013.

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

Frihy, O. E., Lotfy, M. F. and Komar, P. D.: Spatial variations in heavy minerals and patterns of sediment sorting along the Nile Delta, Egypt, Sediment. Geol., 97(1–2), 33–41, 1995.

Geonics, E. M.: EM38 Ground Conductivity Meter Operating Manual, Geonics Ltd. Ontario Mississauga, ON, Canada, 32, 2002.

Greve, M. B. and Malone, B. P.: High-Resolution 3-D Mapping of Soil Texture in Denmark, , doi:10.2136/sssaj2012.0275, 2013.

Grimley, D. A., Arruda, N. K. and Bramstedt, M. W.: Using magnetic susceptibility to facilitate more rapid, reproducible and precise delineation of hydric soils in the midwestern USA, Catena, 58(2), 183–213, doi:10.1016/j.catena.2004.03.001, 2004.

Harris, J. R. and Grunsky, E. C.: Computers & Geosciences Predictive lithological mapping of Canada ’ s North using Random Forest classi fi cation applied to geophysical and geochemical data, Comput. Geosci., 80, 9–25, doi:10.1016/j.cageo.2015.03.013, 2015.

Heil, K. and Schmidhalter, U.: Theory and Guidelines for the Application of the Geophysical Sensor EM38, , 38, 2019.

Hendrickx, ; Kachanoski, R. .: Miscible Solute Transport -Solute Content and Concentration - Indirect Measurement of Solute Concentration: Electromagnetic Induction, in Methods of Soil Analysis, vol. Chapter 6., 2002.

Henrique, S., Silva, G., Silva, E. A., Poggere, G. C., Linares, A., Junior, P., Gabriele, M., Gonçalves, M., Roberto, L., Guilherme, G. and Curi, N.: Soils and Plant Nutrition Modeling and prediction of sulfuric acid digestion analyses data from PXRF spectrometry in tropical soils, Sci. Agric., 2018.

Heuvelink, G. B. M. and Webster, R.: Modelling soil variation: past, present, and future, Geoderma, 100(3–4), 269–301, 2001.

Honeyborne, I., McHugh, T. D., Kuittinen, I., Cichonska, A., Evangelopoulos, D., Ronacher, K., van Helden, P. D., Gillespie, S. H., Fernandez-Reyes, D., Walzl, G., Rousu, J., Butcher, P. D. and Waddell, S. J.: Profiling persistent tubercule bacilli from patient sputa during therapy predicts early drug efficacy, BMC Med., 14(1), 1–13, doi:10.1186/s12916-016-0609-3, 2016.

Hothorn, T.: CRAN task view: Machine learning & statistical learning, 2021.

Hounkpatin, O. K. L., Op, F., Hipt, D., Yaovi, A., Welp, G. and Amelung, W.: Catena Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso), Catena, 166(April), 298–309, doi:10.1016/j.catena.2018.04.013, 2018.

IUSS Working Group WRB: World reference base for soil resources 2014. International soil classification system for naming

soils and creating legends for soil maps., 2015.

Jafarzadeh, A. A., Pal, M., Servati, M., FazeliFard, M. H. and Ghorbani, M. A.: Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction, Int. J. Environ. Sci. Technol., 13(1), 87–96, doi:10.1007/s13762-015-0856-4, 2016.

Javadi, S. H., Munnaf, M. A. and Mouazen, A. M.: Fusion of Vis-NIR and XRF spectra for estimation of key soil attributes, Geoderma, 385, 114851, 2021.

Jenny, H.: Factors of soil formation: A system of quantitative pedology, Dover publication, New York., 1994.

Jiménez, C., Benavides, J., Ospina-Salazar, D. I., Zúñiga, O., Ochoa, O. and Mosquera, C.: Relationship between physical properties and the magnetic susceptibility in two soils of Valle del Cauca Relación entre propiedades físicas y la susceptibilidad magnética en dos suelos del Valle del Cauca, Cauca. Rev. Cienc. Agri, 34(341), 33–45, doi:10.22267/rcia.173402.70, 2017.

Johnston, M. A., Savage, M. J., Moolman, J. H. and du Plessis, H. M.: Evaluation of Calibration Methods for Interpreting Soil Salinity from Electromagnetic Induction Measurements, Soil Sci. Soc. Am. J., 61(6), 1627–1633, doi:10.2136/sssaj1997.03615995006100060013x, 1997.

De Jong, E., Pennock, D. J. and Nestor, P. A.: Magnetic susceptibility of soils in different slope positions in Saskatchewan, Canada, Catena, 40(3), 291–305, doi:10.1016/S0341-8162(00)00080-1, 2000.

Jung, Y., Lee, J., Lee, M., Kang, N. and Lee, I.: Probabilistic analytical target cascading using kernel density estimation for accurate uncertainty propagation, Struct. Multidiscip. Optim., 1–19, 2020.

Kämpf, N. and Curi, N.: Óxidos de ferro: indicadores de ambientes pedogênicos e geoquímicos, Tópicos em ciência do solo, 1, 107–138, 2000.

Kuhn, M. and Johnson, K.: Applied predictive modeling, Springer., 2013.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B. and Team, R. C.: Package 'caret,' R J., 2020.

Lacoste, M., Lemercier, B. and Walter, C.: Regional mapping of soil parent material by machine learning based on point data, Geomorphology, 133(1–2), 90–99, doi:10.1016/j.geomorph.2011.06.026, 2011.

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

Leng, X., Qian, X., Yang, M., Wang, C., Li, H. and Wang, J.: Leaf magnetic properties as a method for predicting heavy metal concentrations in PM 2.5 using support vector machine : A case study in Nanjing, China, Environ. Pollut., 242, 922–930, doi:10.1016/j.envpol.2018.07.007, 2018.

Lesch, S. M., Rhoades, J. D., Lund, L. J. and Corwin, D. L.: Mapping soil salinity using calibrated electromagnetic measurements, Soil Sci. Soc. Am. J., 56(2), 540–548, 1992.

Levi, M. R. and Rasmussen, C.: Covariate selection with iterative principal component analysis for predicting physical soil properties, Geoderma, 219, 46–57, 2014.

Li, H., Wang, J., Wang, Q., Tian, C., Qian, X. and Leng, X.: Magnetic Properties as a Proxy for Predicting Fine-Particle-Bound Heavy Metals in a Support Vector Machine Approach, Environ. Sci. Technol., 51(12), 6927–6935, doi:10.1021/acs.est.7b00729, 2017.

Liao, K., Xu, S., Wu, J., Zhu, Q. and An, L.: Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China, J. Plant Nutr. Soil Sci., 177(5), 775–782, 2014.

Ließ, M., Glaser, B. and Huwe, B.: Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models, Geoderma, 170, 70–79, doi:https://doi.org/10.1016/j.geoderma.2011.10.010, 2012.

Lim, C.H., Jackson, M. L.: Dissolution for total elemental analysis, in Methods of Soil Analysis. Part 2: Chemical and Microbiological Properties., edited by Madison, pp. 1–12, American Society of Agronomy., 1986.

Loiseau, T., Richer-de-forges, A. C., Martelet, G., Bialkowski, A., Nehlig, P. and Arrouays, D.: Geoderma Regional Could airborne gamma-spectrometric data replace lithological maps as co-variates for digital soil mapping of topsoil particle-size distribution ? A case study in Western France, Geoderma Reg., 22, e00295, doi:10.1016/j.geodrs.2020.e00295, 2020.

Malone, B. P., McBratney, A. B., Minasny, B. and Laslett, G. M.: Mapping continuous depth functions of soil carbon storage and available water capacity, Geoderma, 154(1–2), 138–152, doi:10.1016/j.geoderma.2009.10.007, 2009.

McFadden, M. and Scott, W. R.: Broadband soil susceptibility measurements for EMI applications, J. Appl. Geophys., 90, 119–125, doi:10.1016/j.jappgeo.2013.01.009, 2013.

Mcneill, J. D.: Rapid, accurate mapping of soil salinity by electromagnetic ground conductivity meters, , (30), 2–3, 1992.

McNeill, J. D.: Geonics EM38 ground conductivity meter, Tech. Note TN-21. Geonics Ltd., Mississauga, Ontario, Canada, 1986.

Mello, D., Demattê, J. A. M., Silvero, N. E. Q., Di Raimo, L. A. D. L., Poppiel, R. R., Mello, F. A. O., Souza, A. B., Safanelli, J. L., Resende, M. E. B. and Rizzo, R.: Soil magnetic susceptibility and its relationship with naturally occurring processes and soil attributes in pedosphere, in a tropical environment, Geoderma, 372(March), 114364, doi:10.1016/j.geoderma.2020.114364, 2020.

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. and Ernesto Gonçalves Reynaud Schaefer, C.: Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes, Geoderma, 381, doi:10.1016/j.geoderma.2020.114736, 2021.

Minty, B. R. S.: A Review of Airborne Gamma-Ray Spectrometric Data-Processing Techniques, Aust. Gov. Publ. Serv., 1988.

Montanarella, L., Pennock, D. J., McKenzie, N. J., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Singh Aulakh, M. and Yagi, K.: World's soils are under threat, Soil, 2(2), 1263–1272, 2015.

Mullins, C. E.: Magnetic susceptibility of the soil and its significance in soil science–a review, J. soil Sci., 28(2), 223–246, 1977.

Nanni, M. R. and Demattê, J. A. M.: Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis, Soil Sci. Soc. Am. J., 70(2), 393–407, doi:10.2136/sssaj2003.0285, 2006.

Narjary, B., Meena, M. D., Kumar, S., Kamra, S. K., Sharma, D. K. and Triantafilis, J.: Digital mapping of soil salinity at various depths using an EM38, Soil Use Manag., 35(2), 232–244, doi:10.1111/sum.12468, 2019.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J. and Mouazen, A. M.: Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy, Soil Tillage Res., 155, 510–522, doi:10.1016/j.still.2015.07.021, 2016.

Neogi, S. and Dauwels, J.: Factored Latent-Dynamic Conditional Random Fields for Single and Multi-label Sequence Modeling, arXiv Prepr. arXiv1911.03667, 2019.

O'Rourke, S. M., Stockmann, U., Holden, N. M., McBratney, A. B. and Minasny, B.: An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties, Geoderma, 279, 31–44, doi:10.1016/j.geoderma.2016.05.005, 2016.

Pansu, M., Gautheyrou, J.: Handbook of Soil Analysis – Mineralogical, Organic and Inorganic Methods., Springer, Netherlands., 2006.

Perlich, C.: Learning Curves in Machine Learning., 2010.

Pozza, L. E. and Field, D. J.: The science of soil Security and food security, Soil Secur., 1, 100002, 2020.

Priori, S., Fantappiè, M., Bianconi, N., Ferrigno, G., Pellegrini, S. and Costantini, E. A. C.: Field-Scale Mapping of Soil Carbon Stock with Limited Sampling by Coupling Gamma-Ray and Vis-NIR Spectroscopy, Soil Sci. Soc. Am. J., 80(4), 954–964, doi:10.2136/sssaj2016.01.0018, 2016.

Reinhardt, N. and Herrmann, L.: Gamma-ray spectrometry as versatile tool in soil science: A critical review, J. Plant Nutr. Soil Sci., 182(1), 9–27, doi:10.1002/jpln.201700447, 2019.

Rhoades, J. D., Chanduvi, F. and Lesch, S. M.: Soil salinity assessment: Methods and interpretation of electrical conductivity measurements, Food & Agriculture Org., 1999.

Richards, L. A.: Diagnosis and improvement of saline and alkali soils, LWW., 1954.

Rochette, P., Jackson, M. and Aubourg, C.: Rock magnetism andn the interpretation of magnetic susceptibility, Rev. Geophys., 30(3), 209–226, 1992.

Rytky, S. J. O., Tiulpin, A., Frondelius, T., Finnilä, M. A. J., Karhula, S. S., Leino, J., Pritzker, K. P. H., Valkealahti, M., Lehenkari, P., Joukainen, A., Kröger, H., Nieminen, H. J. and Saarakkala, S.: Automating three-dimensional osteoarthritis histopathological grading of human osteochondral tissue using machine learning on contrast-enhanced micro-computed tomography, Osteoarthr. Cartil., 28(8), 1133–1144, doi:10.1016/j.joca.2020.05.002, 2020.

Sales, S. and C.: Terraplus KT-10 v2 User Manual, , User's Guide ver. 2.1, 2021.

Sarmast, M., Farpoor, M. H. and Esfandiarpour Boroujeni, I.: Magnetic susceptibility of soils along a lithotoposequence in southeast Iran, Catena, 156(March), 252–262, doi:10.1016/j.catena.2017.04.019, 2017.

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

Schuler, U., Erbe, P., Zarei, M., Rangubpit, W., Surinkum, A., Stahr, K. and Herrmann, L.: A gamma-ray spectrometry approach to field separation of illuviation-type WRB reference soil groups in northern Thailand, J. Plant Nutr. Soil Sci., 174(4),

536–544, doi:10.1002/jpln.200800323, 2011.

Schwertmann, U. and Taylor, R. M.: Iron oxides, Miner. soil Environ., 1, 379–438, 1989.

Shenggao, L.: Lithological factors affecting magnetic susceptibility of subtropical soils, Zhejiang Province, China, Catena, 40(4), 359–373, doi:10.1016/S0341-8162(00)00092-8, 2000.

Silva, E. B.: A Regional Legacy Soil Dataset for Prediction of Sand and Clay Content with Vis-Nir-Swir , in Southern Brazil, Rev. Bras. Cienc. do Solo, 1–20, 2019.

Silvero, N. E. Q., Di Raimo, L. A. D. L., Pereira, G. S., Magalhães, L. P. de, Terra, F. da S., Dassan, M. A. A., Salazar, D. F. U. and Demattê, J. A. M.: Effects of water, organic matter, and iron forms in mid-IR spectra of soils: Assessments from laboratory to satellite-simulated data, Geoderma, 375(December 2019), 114480, doi:10.1016/j.geoderma.2020.114480, 2020.

Siqueira, D. S., Marques, J., Matias, S. S. R., Barrón, V., Torrent, J., Baffa, O. and Oliveira, L. C.: Correlation of properties of Brazilian Haplustalfs with magnetic susceptibility measurements, Soil Use Manag., 26(4), 425–431, doi:10.1111/j.1475-2743.2010.00294.x, 2010.

Solutions, R.: Spectrum stabilization and calibration for the RSI RS-125 and RS-230 handheld spectrometers, 2009.

Sousa, I., Costa, L., Cavalcanti, I., Oliveira, C. De, Tavares, F. M., José, H., Polo, D. O., Sousa, I., Costa, L., Cavalcanti, I. and Oliveira, C. De: Uranium anomalies detection through Random Forest regression Uranium anomalies detection through Random Forest regression, , doi:10.1080/08123985.2020.1725387, 2020.

de Souza Bahia, A. S. R., Marques, J., La Scala, N., Pellegrino Cerri, C. E. and Camargo, L. A.: Prediction and mapping of soil attributes using diffuse reflectance spectroscopy and magnetic susceptibility, Soil Sci. Soc. Am. J., 81(6), 1450–1462, 2017.

Targulian, V.O.; Arnold, R.W.; Miller, B. A. . B.: Encyclopedia of Ecology, 2nd ed., The Netherlands; Elsevier: Amsterdam, Volume 4, 162–168., 2019.

Taylor, M. J., Smettem, K., Pracilio, G. and Verboom, W.: Relationships between soil properties and high-resolution radiometrics, central eastern Wheatbelt, Western Australia, Explor. Geophys., 33(2), 95–102, doi:10.1071/EG02095, 2002a.

Taylor, M. J., Smettem, K., Pracilio, G. and Verboom, W.: Relationships between soil properties and high-resolution radiometrics, central eastern Wheatbelt, Western Australia, Explor. Geophys., 33(2), 95–102 [online] Available from: https://doi.org/10.1071/EG02095, 2002b.

Teixeira, P. C., Donagemma, G. K., Fontana, A. and Teixeira, W. G.: Manual de métodos de análise de solo, Rio Janeiro, Embrapa. 573p, 2017.

Terra, F. S., Demattê, J. A. M. and Viscarra Rossel, R. A.: Proximal spectral sensing in pedological assessments: vis–NIR spectra for soil classification based on weathering and pedogenesis, Geoderma, 318(January), 123–136, doi:10.1016/j.geoderma.2017.10.053, 2018a.

Terra, F. S., Demattê, J. A. M. and Viscarra Rossel, R. A.: Proximal spectral sensing in pedological assessments: vis–NIR spectra for soil classification based on weathering and pedogenesis, Geoderma, 318(October 2017), 123–136, doi:10.1016/j.geoderma.2017.10.053, 2018b.

Triantafilis, J., Lesch, S. M., La Lau, K. and Buchanan, S. M.: Field level digital soil mapping of cation exchange capacity using electromagnetic induction and a hierarchical spatial regression model, Aust. J. Soil Res., 47(7), 651–663, doi:10.1071/SR08240, 2009.

Valaee, M., Ayoubi, S., Khormali, F., Lu, S. G. and Karimzadeh, H. R.: Using magnetic susceptibility to discriminate between soil moisture regimes in selected loess and loess-like soils in northern Iran, J. Appl. Geophys., 127, 23–30, doi:10.1016/j.jappgeo.2016.02.006, 2016.

Vašát, R., Kode, R., Klement, A. and Brodský, L.: Combining re fl ectance spectroscopy and the digital elevation model for soil oxidizable carbon estimation, , 303(May), 133–142, doi:10.1016/j.geoderma.2017.05.018, 2017.

Viana, J. H. M., Couceiro, P. R. C., Pereira, M. C., Fabris, J. D., Fernandes Filho, E. I., Schaefer, C., Rechenberg, H. R., Abrahão, W. A. P. and Mantovani, E. C.: Occurrence of magnetite in the sand fraction of an Oxisol in the Brazilian savanna ecosystem, developed from a magnetite-free lithology, Soil Res., 44(1), 71–83, 2006.

Viscarra Rossel, R. A., Webster, R. and Kidd, D.: Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random forest kriging, Earth Surf. Process. Landforms, 39(6), 735–748, doi:10.1002/esp.3476, 2014.

Wilford, J. and Minty, B.: Chapter 16 The Use of Airborne Gamma-ray Imagery for Mapping Soils and Understanding Landscape Processes, Dev. Soil Sci., 31(C), doi:10.1016/S0166-2481(06)31016-1, 2006.

Wilford, J. and Thomas, M.: Modelling soil-regolith thickness in complex weathered landscapes of the central Mt Lofty Ranges, South Australia, 2012.

Wilford, P. N., Bierwirth, J. R. and Craig, M. A.: Application of airborne gamma-ray spectrometry in soiVregolith mapping and Applied Geomorphology, , 17(2), 1997.

Wong, M. T. F. and Harper, R. J.: Use of on-ground gamma-ray spectrometry to measure plant-available potassium and other topsoil attributes, Aust. J. Soil Res., 37(2), 267–277, doi:10.1071/S98038, 1999.

Xu, D., Zhao, R., Li, S., Chen, S., Jiang, Q., Zhou, L. and Shi, Z.: Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China, Eur. J. Soil Sci., 70(1), 162–173, 2019.

Zare, E., Li, N., Khongnawang, T. and Farzamian, M.: Identifying Potential Leakage Zones in an Irrigation Supply Channel by Mapping Soil Properties Using Electromagnetic Induction , Inversion Modelling and a Support Vector Machine, 2020.

Zhang, Y. and Hartemink, A. E.: Data fusion of vis – NIR and PXRF spectra to predict soil physical and chemical properties, Eur. J. Soil Sci., (May 2019), 316–333, doi:10.1111/ejss.12875, 2020.