

How is the learning process of digital soil mapping in a diverse group of land use planners?

Ricardo Simão Diniz Dalmolin^{(1)*} , Jean Michel Moura-Bueno⁽¹⁾ , Alessandro Samuel-Rosa⁽²⁾  and Carlos Alberto Flores⁽³⁾ 

⁽¹⁾ Universidade Federal de Santa Maria, Departamento de Ciência do Solo, Santa Maria, Rio Grande do Sul, Brasil.

⁽²⁾ Universidade Federal Tecnológica do Paraná, Curso de Agronomia, Santa Helena, Paraná, Brasil.

⁽³⁾ Empresa Brasileira de Pesquisa Agropecuária, Embrapa Clima Temperado, Pelotas, Rio Grande do Sul, Brasil.

ABSTRACT: The use of new technologies, the development of new software, and the advances in the machines ability to process data have brought a new perspective to soil science and especially to pedology, with the advent of digital soil mapping (DSM). To meet the demand for soil surveys in Brazil, it will be necessary to popularize the techniques used in DSM. To identify and map the soil to generate maps of land use capability, we proposed a theoretical and practical course focused on the training in DSM for professionals involved in the management of land resources. The methodology was divided into five modules: I. Introduction to pedology, soil-landscape relationship, soil survey and soil classification (theory); II. Identification of soils in the field and study of the soil-landscape relationship (practice); III. Digital soil mapping and geographic information system (theory) and obtaining environmental covariates (practice); IV. Statistical learners and quality measures of spatial predictions (theory) and spatial pseudo-sampling (practice); V. Database organization, calibration, and validation of predictive models (practice). Results such as the average level of confidence of the participants in the soil classification, as well as the number of pseudo-sampling classified by the participants, chosen statistical apprentice, environmental covariables used, and overall accuracy, were influenced by the participants level of knowledge regarding DSM soils and techniques. The structure, focus, and time of each module should be based on the participants needs. It is suggested that a survey should be carried out to consider the level of knowledge in relation to the topics addressed in DSM before the preparation and execution of the course. The contribution of individual experiences showed the importance of multidisciplinary in the teaching-learning process because it is a technique that involves soil knowledge, statistics, and mathematics applied to geoinformation science to understand soil variability in the landscape. The practical classes were fundamental, enabling an approximation of the content studied with the participants' reality and consolidation of the acquired knowledge. In general, the course was well evaluated by the participants regarding the contents covered and practical field training and laboratory geoprocessing, who reported that the practical classes were fundamental for the appropriation of knowledge in DSM. This course could be a model for the PronaSolos, which tend to have heterogeneous groups of participants, being necessary to plan specific protocols to tend the particularity.

* Corresponding author:
E-mail: dalmolin@ufsm.br

Received: March 08, 2019

Approved: November 4, 2019

How to cite: Dalmolin RSD, Moura-Bueno JM, Samuel-Rosa A, Flores CA. How is the learning process of digital soil mapping in a diverse group of land use planners? Rev Bras Cienc Solo. 2020;44:e0190037.
<https://doi.org/10.36783/18069657rbc20190037>

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.



Keywords: expert knowledge, pedometry, PronaSolos, soil education, soil survey.

INTRODUCTION

Soil class and properties for rational land use planning, prediction of future scenarios such as erosion, sedimentation, climate change, and also as a data source for modeling, are an urgent need (Amundson et al., 2015; Dalmolin and ten Caten, 2015; FAO/ITPS, 2015). In Brazil, most of the available maps are classified as small scale, smaller than 1:250,000, suitable for state or region level land use planning (Santos et al., 2013). In addition, most of these maps were generated by the traditional soil survey method, which is characterized by not meeting society's current demand for low-cost, short-term quantitative information (Hartemink and McBratney, 2008; Sanchez et al., 2009). This inability of soil maps obtained through the traditional method is related to the partial loss of information on soil variability in the landscape since the discrete model employed results in choroplectic maps (ten Caten et al., 2011), which impose abrupt boundaries between soil classes. This feature of traditional surveys results in difficulties in the practical application of map information (Sanchez et al., 2009).

To fulfill the need to map Brazilian soils and to properly approach land use planning, the National Program of Soil Survey and Interpretation (PronaSolos) (Polidoro et al., 2016) was recently established. The main objective of PronaSolos is to map the soils of the entire Brazilian territory in scales varying from 1:25,000 to 1:100,000 (Polidoro et al., 2016). However, conducting of soil surveys in the demand required by Brazil will only be possible with the use of new mapping techniques. The use of new technologies, the development of new softwares (Arrouays et al., 2017), and the advances in the machines ability to process data (Heung et al., 2016) have brought a new perspective to soil science and especially to pedology, with the advent of digital soil mapping (DSM) (McBratney et al., 2003; Lagacherie and McBratney, 2007). To meet the demand for DSM, provided by the actions of PronaSolos, it will be necessary to popularize the techniques used in this new methodology. In addition to the investments in soil mapping, the training of new pedologists with special attention to DSM should be considered (Arrouays et al., 2017). Thus, soil scientists working on predictive soil mapping need to incorporate these techniques and methodologies used in DSM, morphometry, and proximal remote sensing to meet the demand for spatial soil information (Hartemink, 2015).

Despite a large number of undergraduate courses with an emphasis on agrarian sciences in Brazil, few professionals are trained to work in the field of pedology, and even less are familiar with techniques required in DSM (Dalmolin and ten Caten, 2015). Some countries at the forefront of new developments in DSM, such as Australia and the Netherlands, have training courses aimed at this technique. The first DSM course in Australia was held in 2011 at the University of Sydney, meeting a request from the Australian Agricultural Land Assessment Program (Minasny and McBratney, 2016). The course was structured to develop user skills demonstrating how to use DSM techniques developed in the research to design soil maps for land use planning purposes. These authors report the positive experience of these training courses and are essential to initiate DSM activities across Australia. In the Netherlands, there are a series of training courses applied to DSM developed at the International Soil Reference and Information Center, whose main objective is to produce soil maps and information using local, regional, and global data sets. In Brazil, however, it has only been reported that EMBRAPA Solos has developed training courses in DSM (Baca et al., 2013; Vasques et al., 2013). The EMBRAPA Solos courses, aimed at training professionals from Brazil and other Latin American and Caribbean countries, showed that it is possible to carry out low cost theoretical and practical laboratory training using free software and data available in soil databases and environmental covariates derived, for example, Shuttle Radar Topography Mission (SRTM) and Landsat Mission.

MacMillan and Hengl (2019), discussing the future of predictive soil mapping (PSM) observed that it is necessary to adopt new methods and ideas associated with PSM within a new collaborative and open operational framework. Concerning collaborative and voluntary contributions from citizen scientists, Hengl et al. (2018) go further, asserting that there is a role in PSM for crowdsourcing to engaged citizen volunteers in collecting field observations and measurements to extend the soil and environmental relationships. According to Rossiter et al. (2015), soil observations require fieldwork. These authors state that soil maps are known by the user who relies on soil maps for decision making, especially those who are linked to agriculture or land planners. For increasing initiatives to a better understanding of soil, Rossiter et al. (2015) suggest multiple initiatives could reach projects in DSM, among them, training opportunities.

The DSM involves solid knowledge of pedology, statistics, and mathematics (Lagacherie and McBratney, 2007). Thus, in the proposition of this DSM course, we focus on the theoretical, practical field and laboratory software knowledge, and it is clear that the methodological proposition of this course is not only on pedagogical emphasis but on specific knowledge directed at pedology and DSM. Within this perspective, this work aims to: (i) present the first Brazilian experience of a theoretical and practical course, including field practice, focused on the DSM training for professionals with different levels of soil mapping knowledge; (ii) to evaluate whether the degree of experience in soil mapping of participants influence the products generated by DSM technique. Moreover, the report about the structure of the course, experience, and results may even serve as a basis for the planning of training courses provided in PronaSolos.

MATERIALS AND METHODS

General structure of the course

The theoretical-practical course of DSM was developed at the Agronomic Institute of Paraná (IAPAR), in Londrina, Paraná State, Brazil. Twenty-three professionals participated, including undergraduate teachers, researchers, rural extension workers, land-use planners, and policy-makers. The course has been presented through theory classes and practical sessions for five days, totaling 40 hours (approximately 8 hours by module). The course was structured in five modules, starting with basic pedological concepts, both in the classroom and in the field, followed by the basic concepts of DSM and geographic information systems (GIS), and its practical application for soil mapping (Figure 1). A detailed description is presented in the following sections.

Practical field and DSM application activities were developed in a catchment with 3,212 ha, here called BH, located near the IAPAR headquarters. The participants of the course were assigned the task of generating a detailed map of soil taxonomic classes [in the second categorical level of the Brazilian Soil Classification System, SiBCS (Santos et al., 2018)] of the BH - along with uncertainty measures - using DSM techniques. In a traditional soil survey, a detailed map is one that is produced using an observation density of one observation per 0.8-4.0 ha, considering a minimum mappable area (MMA) of 0.4 ha, and published with a cartographic scale of 1:10,000 (Rossiter, 2000). In DSM, data are handled exclusively in the digital environment and maps are published in the form of raster images. Thus, the concepts of MMA and cartographic scale lose meaning, giving rise to the concepts of spatial resolution and pixel size. There is no direct equivalence between these concepts. However, an approximation can be made from the MMA and the fact that at least four pixels are required to identify a rectangular object in a digital image. If we take the MMA of 0.4 ha as a rectangular object and divide it into four elements, we find what would be the pixel size, i.e., $0.5 \times \sqrt{4000} \text{ m}^2 = 31.62 \text{ m}$. A number of different digital landscape mapping projects have a similar pixel size, i.e., 30 m, such as LANDSAT (NASA, 2009), SRTM V3.0 (NASA/JPL, 2013), and TOPODATA (Valeriano and Rossetti, 2012).

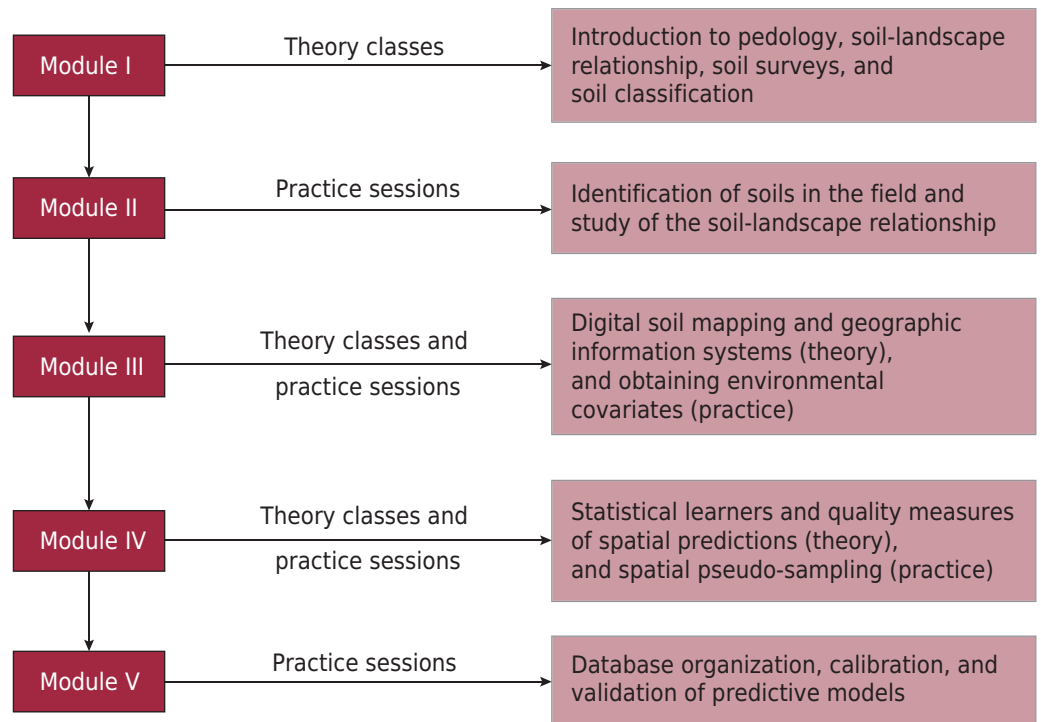


Figure 1. Flowchart of the theoretical-practical course of DSM.

Given the easy access to these environmental covariates data for DSM, the 30 m pixel size was adopted for the generation of a detailed soil map of the BH.

Description of course modules

Module I

The theory classes on pedology presented basic concepts about soil morphology, soil profile, pedogenic horizons, and diagnostic horizons, soil identification and classification, and the SiBCS structure (Santos et al., 2018). All theoretical concepts about soil surveys (IBGE, 2015), including the soils and their variability in the landscape and soil-landscape relationships were also addressed. Module I provided the necessary theoretical framework for the field practices and the knowledge needed for the relationships to be established in the DSM approach.

Module II

In the practice sessions (Figure 2), the theoretical concepts approached in the module I were explored with more emphasis on the field soil survey stages such as landscape information acquisition and identification of soil taxonomic classes. The taxonomic class recognition process took place by identifying the diagnostic horizons (Figure 2a) defined by SiBCS for each soil class. In this module, much emphasis has been placed on the soil-landscape relationship, the main soil formation factors (Figures 2b and 2c) in which relationships have been established between the elevation, slope, and curvature of the terrain with the soil classes in the landscape. From this, the instructors established the conceptual model of pedogenesis together with the participants (Figure 2d). In this model, in BH the *Latosolos Vermelhos* occur in the summit positions with declivity varying from 3 to 8 %. The *Nitossolos Vermelhos* are found on the surfaces of deflection of undulated relief, with slope varying from 8 to 20 %. On the slopes with inflection surfaces, strong undulated and sometimes mountainous relief, the *Neossolos Regolíticos* and *Neossolos Litólicos* are found. The *Cambissolos Flúvicos* are located in the lower open areas of the BH, close to the streams. These five classes of soils in the WRB system (IUSS Working Group



Figure 2. Field practice sessions with identification and description of soil profiles (a), a study of the soil-landscape relationship (b and c), and construction of the conceptual model of pedogenesis in the study area (d)

WRB, 2015) correspond respectively to Rhodic Ferralsol, Rhodic Nitisol, Umbric Leptosol, Leptic Regosol, and Fluvic Cambisol. More details about BH soil can be found in the report of the semi-detailed soil survey of the municipality of Londrina (Bognola et al., 2011).

Module III

The main concepts of the DSM were approached with emphasis on the model $S = f(S, C, O, R, P, A, N)$, in which S: soil, C: climate, O: organism, R: relief, P: parent material, A: age, N: location, for quantification of the correlation between soil and environmental conditions, and production of graphical representations of soil in a digital environment (McBratney et al., 2003). The statistical concepts of “variable response” or “dependent variable”, and “covariable” or “independent variable” were presented. The nature of soil data was defined in three types: “continuous”, such as carbon and clay content, “ordinal”, as the drainage class and the stoniness class, and “categorical”, as the taxonomic class. The GIS concepts such as coordinate reference systems (CRS), data representation models (vector and matrix), CRS transformation and structure, and nomenclature of files and folders in GIS were discussed. For the realization of the DSM work and exercises, the free software QGIS version 2.18 (QGIS Development Team, 2017) and R version 3.4.0 (R Core Team, 2017) were used because they are developed in collective collaboration, are flexible and without cost, being among the most used by the DSM scientific community (Samuel-Rosa et al., 2015; Vaysse and Lagacherie, 2015; Heung et al., 2016; Arrouays et al., 2017; Chagas et al., 2017). The need to obtain information on soils and environmental covariates to supply the predictive models and perform soil mapping was also addressed in this module.

In the practice sessions, it was demonstrated how covariates are made available and can be obtained, for example, from the Topodata project (<http://www.dsr.inpe.br/topodata/>), WorldClim (<http://worldclim.org/version2>), and United States Geological Survey - USGS (<https://earthexplorer.usgs.gov>). It was highlighted that pre-existing soil profile data, for

example, legacy data from the Embrapa Brazilian Soil Information System (https://www.bdsolos.cnptia.embrapa.br/consulta_publica.html), IBGE (<http://www.downloads.ibge.gov.br/tematicos/solos>), and World Soil Information System - WoSIS (<http://www.isric.org/explore/wosis>), can be used in conjunction with the new field samplings. To facilitate participants' understanding of the sources and use of environmental covariates, as well as the correlation between soil and environmental characteristics used in examples of global DSM projects, the results of SoilGrids1km were presented (Hengl et al., 2014).

For this task, the vector files of the area (catchment boundary, contour lines, hydrography, among others) and the environmental covariates in matrix files obtained from Topodata (elevation, slope, horizontal curvature, and vertical curvature) were provided to the participants.

Module IV

In this module, participants were grouped into four groups, named 1, 2, 3, and 4. Group 1 was composed of technicians with little knowledge in soil science and whose academic formation may not have addressed subjects such as soil survey and classification. Group 3 was composed of technicians who worked for years with soil survey and are researchers and professors in this area, while groups 2 and 4 presented previous knowledge of the subject and work on a daily basis with soils. This grouping was carried out to evaluate the influence of the previous level of knowledge on soil mapping in the pseudo-sampling stage and effect on quality of the final product (predicted map) generated by the DSM technique. The DSM courses conducted so far did not use this method.

The goal of Module IV was to obtain the data for calibration of the models for the DSM. As BH has 3,212 ha, at least 803 observations were required (3,212 ha/4 ha). As the duration of the course was only 40 hours, there was no time to visit all the 803 necessary points in the field. The alternative was to use pseudo-observations of the soil, sampled computationally. Pseudo-observations are based on the use of the theoretical model of pedogenesis - created in Module II - to deduce the taxonomic class of the soil in an unvisited site of BH (Figure 3) based on probabilities. To avoid participants' tendency in choosing the sites of soil pseudo-observations, their location was defined using a completely random mechanism in QGIS (QGIS geoalgorithm > vector selection tool > random points within fixed polygons). To maximize the spatial coverage of BH and ensure that an area of 30 × 30 m \cong 0.4 ha (MMA) had only a single pseudo-observation inside it, it was established that the minimum distance between two neighboring pseudo-observations should be $\sqrt{[(30 \text{ m} \times 2)^2 + (30 \text{ m} \times 2)^2]} = 84.85 \text{ m}$.

Since the pseudo-observations of the soil are obtained deductively, on a computer, it was discussed with the participants about the considerable uncertainty about these observations due to the lack of empirical data collected in the field and analyzed in the laboratory. To represent this uncertainty, the concept of *degree of confidence* (DC) on the taxonomic class of a soil profile was presented. First, in discussion with all participants, it was agreed that even considering data with a complete description of a soil profile, confidence in the accuracy of the taxonomic class should be 98 %. This was assumed because the data contains variations from sampling and from the laboratory, which can lead to misclassification. Then, assuming that in BH there are only five taxonomic classes, and not considering their spatial distribution, it was agreed that by using a completely random classification, the taxonomic class of a soil profile would be correct in at least 20 % ($1/5 = 0.20 \times 100 = 20 \%$) of the time. Once the upper and lower degrees of confidence in the soil classification were established, the four groups were asked to indicate their DC in the soil taxonomic class in the following situations: observation by auger, observation in the soil profile, and pseudo-sampling on the computer screen.

From the set of 803 random points established in QGIS, each participant performed the pseudo-sampling of as many points as they needed, based on their pedological knowledge

and auxiliary data, as contour lines, Google Earth 3D satellite imagery (Figure 3a) and terrain covariates (elevation and slope) of the area (Figure 3b), establishing the respective DC for each soil taxonomic classification.

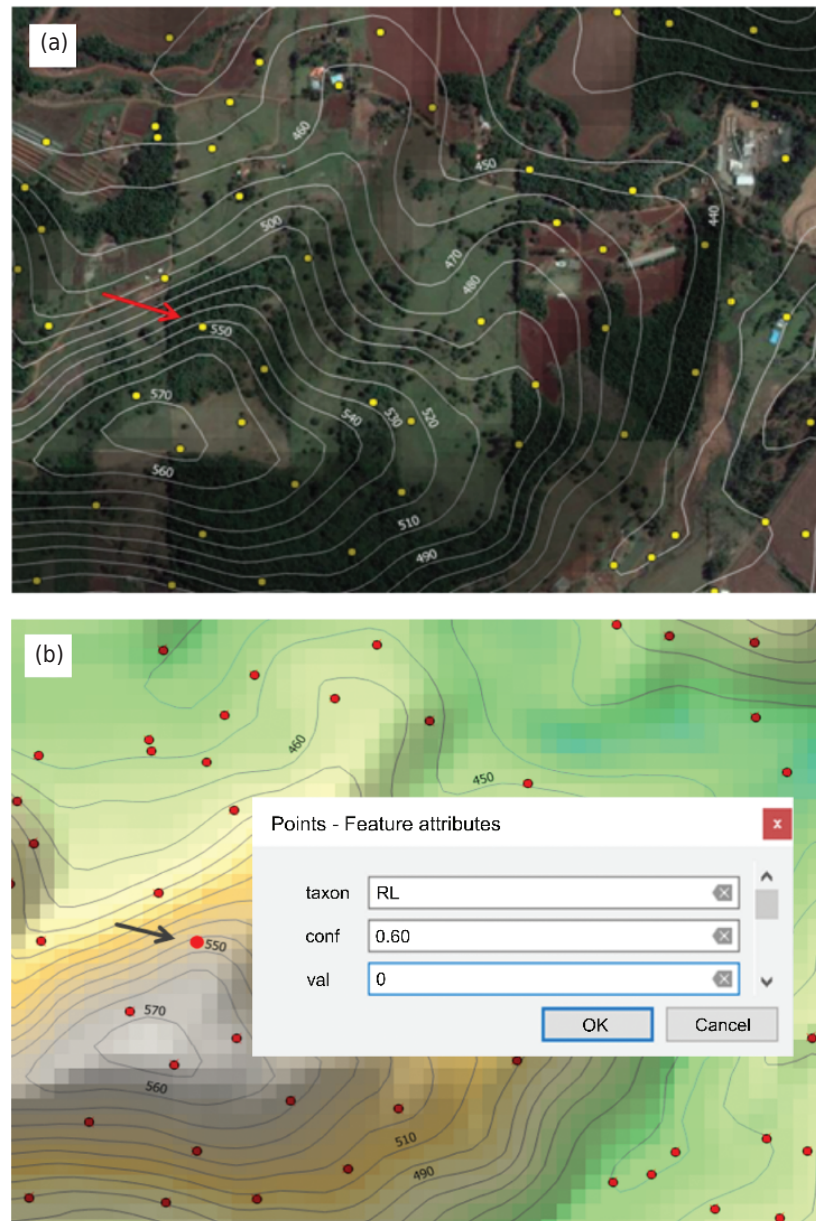


Figure 3. Pseudo-sampling on computer simulations based on the soil-landscape relationship. Google Earth image with contour lines (equidistant 10 meters) (a) and Digital elevation model, contours and taxonomy assignment, and confidence level for each point (b).

Table 1. Average confidence level (DC) of each group of participants in the classification of the soil in one of the five taxonomic classes identified in the BH

Soil information available	Group				
	Everyone	1	2	3	4
Complete profile	0.98	-	-	-	-
Incomplete profile	-	0.65	0.85	0.95	0.90
Augering/Ravine	-	0.70	0.75	0.90	0.80
Pseudo-sampling	-	0.50	0.70	0.75	0.60
Random classification ⁽¹⁾	0.20	-	-	-	-

⁽¹⁾ Probability of correct classification of a soil observation using a completely random classification, considering that five taxonomic classes were identified in the study area ($1/5 = 0.20$).

The DC in the soil classification from different levels of information from each group of participants is shown in table 1. Group 1 achieved the lowest DC's, and group 3, the largest. Groups 2 and 4 indicated intermediate values.

This module also addressed the basic concepts of the most used machine learning methods in the construction of soil prediction models (Table 2). Because the prediction models present an estimate of how uncertain they are about a predicted value, the concept of uncertainty and its representation in categorical variables (taxonomic classes), such as theoretical purity, Shannon's entropy, and confusion index (Kempen et al., 2009) were presented.

Module V

The guidelines for the organization of the data set containing the information of the soil class and environmental covariates for each point obtained in the pseudo-sampling step were presented. This set was the database for training prediction models. The prediction of soil classes was performed using a Python script specifically developed to access, from QGIS, the machine learning methods implemented in R, which is available at <https://github.com/samuel-rosa/qgis-r>.

In the prediction step, a cross-validation method was adopted for the predictive models (Filzmoser et al., 2009). After this procedure, the external validation of the digital soil map was carried out using 64 real soil observations, most of them located outside the study area, derived from the semi-detailed soil survey of the municipality of Londrina - PR (Bognola et al., 2011). External validation differs from cross-validation by the fact that the data used for validation is not used to feed the statistical apprentice. Thus, cross-validation is a measure obtained in the initial phase of work, using the available data. External validation is always a later phase, carried out using data obtained in the field after the soil map was elaborated. Data from the semi-detailed survey of soils of the municipality of Londrina are available at the Free Brazilian Repository for Open Soil Data (www.ufsm.br/febr) under the identification code ctb0022. The Free Brazilian Repository for Open Soil Data is a repository that stores accessible soil data for various applications (Samuel-Rosa et al., 2020).

The results of this stage were the digital soil map, uncertainty maps, and metadata table. The metadata table is composed of general data information used in the prediction and validation, the importance of predictor covariates, machine learning method used and values of cross-validation and external validation. The uncertainty maps are composed of measures of theoretical purity, Shannon entropy, and confusion index. Theoretical purity is the highest predicted value of probability at a point and varies between 0 and 1, where 1 means maximum theoretical purity, that is, the machine learning method has great confidence about the class of the predicted soil. On the other hand, 0 means minimal theoretical purity, in which the machine learning method has great uncertainty about the

Table 2. R statistical package, method of implementation, description, and reference of the machine learning methods used in the detailed DSM course

Package	Method	Description	Reference
rpart	rpart	Classification and regression tree	Therneau and Atkinson (2019)
MASS	lda	Linear discriminant analysis	Venables and Ripley (2002)
nnet	multinom	Penalized multinomial regression linear	Venables and Ripley (2002)
nnet	nnet	Artificial neural network	Venables and Ripley (2002)
randomForest	rf	Random forest	Liaw and Wiener (2002)
kernlab	svmRadial	Support vector machines with radial basis kernel function	Karatzoglou et al. (2018)

predicted class. The Shannon entropy is a measure of the “disorder” of prediction and also varies from 0 to 1, where 1 means maximum disorder; that is, the machine learning method has very little confidence about the predicted class, and 0 means maximum confidence about the predicted class. The index of confusion is a measure of the confusion the model makes between the two most likely classes. Like the two previous measures, the confusion index ranges from 0 to 1, where 1 means maximum confusion. The R code needed to reproduce these results was implemented in the Python script mentioned above, which is available at <https://github.com/samuel-rosa/qgis-r>.

Evaluation

Each participant generated their own maps and metadata table, which were presented and discussed together at the end of the course, considering pseudo-sampling strategy, the influence of the number of points, correlation of the predicted classes with the observations made in the field activity and the results of accuracy and uncertainty obtained in prediction. Results obtained by the participants will be discussed and presented through letters (A, B, C, D...) in order to keep them anonymous.

For the evaluation of the course, participants were given an evaluation questionnaire with open and closed questions. In the closed questions, scores ranging from 1 (minimum grade) to 5 (maximum grade) were given.

RESULTS

The results of the practical activity of DSM showed that participants who were familiar with the topics covered in the theoretical presentation in module I were the ones who had an effective participation, reporting personal experiences related to what was exposed and discussed, pointing out that the time dedicated to these subjects could be adapted according to the groups needs. It was demonstrated that the ease for soil identification, profile description, and establishing soil-landscape relationships developed in module II, depended on the academic background and previous experience from the participants.

Operational difficulties were observed in the practical activities in module II, even after the theoretical approach about the concepts of DSM and GIS. The main questions were related to the source, meaning, acquisition, and application of environmental covariates in DSM. It was clear that including more detailed information about statistics and also about the ways to obtain environmental covariates and their relationship with the distribution of soils in the landscape was a necessity.

In the topic of pseudo-sampling in module III, we noticed that the knowledge about the soil-landscape relationship of the study area or the tacit pedological knowledge developed by a few more experienced participants (pedologists), besides knowledge in GIS, facilitated the understanding and execution of this step. The results show that participants with more experience in soil mapping and GIS produced a higher number of pseudo-observations.

The digital soil map (Figure 4a) obtained by one of the participants of the course showed much similarity to the distribution of soils in the landscape observed in the construction stage of the pedogenesis model in the field practice. The confusion index map (Figure 4b) shows locations of higher uncertainty of the predictive model, where a higher number of samples is necessary to improve the quality of the soil map. Similarly, these places of greatest uncertainty were described by participants as places of greater difficulty to construct the pedogenesis model.

Regarding the model construction stage and soil class prediction, the greatest operational difficulties were observed, indicating the need for adjustments in to the workload for

this stage of the course, to consolidate learning and mastery of the software operation technique used in MDS. Such fact was reported in the evaluation carried out by the course participants.

In relation to the models used to adjust the predictive models, many questions arose regarding the theoretical basis and statistical assumption of each model and in which scenario to use each one. These doubts arose when doing in group analysis of the results of table 3. It was observed a better performance for linear models compared to the random forest (Table 3). The results of cross-validation and external validation from participant H showed that random forest is subject to overfitting and, therefore, poor at generalization. The questions derived from the discussion of the results in table 3 showed the clear need for greater workload aimed at teaching and learning of theoretical concepts of statistical learners.

In general, regardless of the calibration data, the machine learning methods identified the vertical curvature and elevation as the most important covariates, with slope, in all cases, being the second most important. Vertical curvature and elevation switched

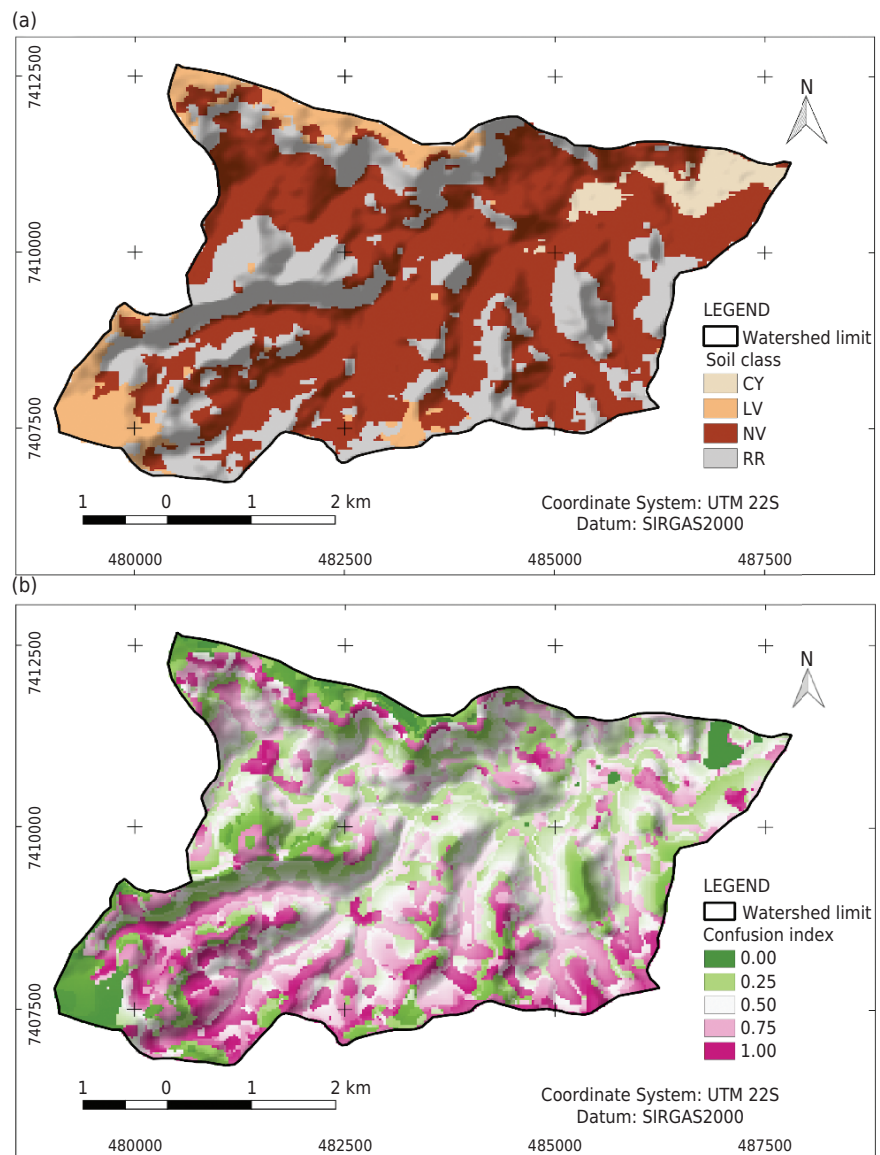


Figure 4. Digital soil map (a) and confusion index (b) produced by one of the participants of the course. LV: *Latosolos Vermelhos* (Rhodic Ferralsol); NV: *Nitossolos Vermelhos* (Rhodic Nitisol); RR: *Neossolos Regolíticos* (Umbric Leptosol); CY: *Cambissolos Flúvicos* (Fluvic Cambisol) (Santos et al., 2018).

Table 3. Participants, number of pseudo-samples performed, machine learning method used for prediction, covariates used in order of importance for the model, classes predicted by the calibrated model, and global purity in cross-validation and external validation. The top three models are highlighted in bold

Participant	Pseudo-samples	Statistical learner	Covariates	Predicted classes ⁽¹⁾	CV ⁽²⁾	EV ⁽³⁾
A	122	Penalized Multinomial Regression	Vertical curvature, Slope, Elevation	LV, NV, RL, RR	0.52	0.50
B	714	Penalized Multinomial Regression	Vertical curvature, Slope, Elevation	CY, LV, NV, RL	0.55	0.42
C	240	Linear Discriminant Analysis	Elevation, Slope, Horizontal curvature	CY, LV, NV, RR	0.80	0.45
D	175	Penalized Multinomial Regression	Vertical curvature, Slope, Elevation	CY, LV, NV, RL, RR	0.70	0.58
E	429	Penalized Multinomial Regression	Vertical curvature, Slope, Horizontal curvature	CY, LV, NV, RL, RR	0.48	0.56
F	813	Random Forest	Elevation, Slope, Vertical curvature	CY, LV, NV, RL, RR	0.74	0.53
G	342	Random Forest	Elevation, Slope, Vertical curvature	CY, LV, NV, RL, RR	0.55	0.44
H	236	Random Forest	Elevation, Slope, Vertical curvature	CY, LV, NV, RL, RR	0.71	0.48

⁽¹⁾ LV: *Latossolo Vermelho* (Rhodic Ferralsol); NV: *Nitossolo Vermelho* (Rhodic Nitisol); RL: *Neossolo Litólico* (Leptic Regosol); RR: *Neossolo Regolítico* (Umbric Leptosol); CY: *Cambissolo Flúvico* (Fluvis Cambisol). ⁽²⁾ CV: Cross-Validation. ⁽³⁾ EV: external validation.

positions in logistic regression and random forest. Regarding the covariates used as predictors, several questions also arose about how many, which, and when to use a given covariate. There were still manifestations about sources of covariates and how to obtain them.

Participant A had the fewest number of pseudo-samples, and the results showed that the class CY was not predicted, probably due to the small representativity of this class in the calibration data. The same pattern can be attributed to participant C, which results had no class RL predicted, besides the amplitude between the accuracy of cross-validation and external validation. In this case, the lack of knowledge about the lower predictive potential of the Linear Discriminant Analysis machine learning by participant C associated with the low number of pseudo-samples resulted in lower accuracy (Table 3).

Regardless of the high number of pseudo-samples, participant B did not distribute the observations properly, since class RR was not predicted. The importance of the quality of pseudo-sampling is demonstrated by participant D, who performed only 175 observations but obtained an accuracy of 0.58 in external validation using the Penalized Multinomial Regression learner. Participant E used 429 pseudo-samples and obtained an accuracy of 0.56 using the same learner as participant D. On the other hand, Random Forest only approaches the best models when the number of observations is very high, which is demonstrated by the results from participant F. This participant is part of group 1 (poor knowledge of soil science), so the participant, knowing their limitations in pedological knowledge, chose to perform excessive pseudo-sampling in order to achieve good accuracy in the prediction. This reflected in the model's greater ability to better identify the transitions between soil classes when more information about the soil is provided, enabling the creation of a higher number of rules with the covariates.

Concerning the steps of pseudo-sampling and spatial prediction, it was observed that the performance of the machine learning methods is related to the quality and quantity of observations for calibration of the models. It should be emphasized that quality

pseudo-samplings require knowledge of the soil-landscape relationship. In fact, the models used in DSM need to take into account the pedological knowledge for their construction and to be in agreement with reasonable hypotheses about the soil-landscape relationship (Rossiter, 2018).

The results of the course evaluation by the participants (Figure 5) show a positive scenario regarding items A, B, C, D, and E, with all answers being very good or excellent (marks 4 and 5). The item F, in which the participant was asked if he/she felt capable of applying the acquired knowledge, was the one that presented the highest percentage (45 %) of negative answers, mainly due to the insecurity of having the first contact with the subject, but this percentage decreased in item H, in which the ability to replicate the information received in future DSM training courses was questioned.

DISCUSSION

It was clear that it is necessary to know beforehand about the training and experiences of the course participants to direct the necessary training time in the basic soil classification modules (modules I and II). Participants reported the need for more time allocated for field practice classes for a better understanding of the soil-landscape relationship. This excessive reliance on field training, classification and soil identification for better understanding of the soil-landscape relationship and field mapping has been reported in a study by Hudson (1992) and later by Scull et al. (2003), who report that the acquisition of tacit pedological knowledge is a slow and expensive process, as it requires a lot of field training.

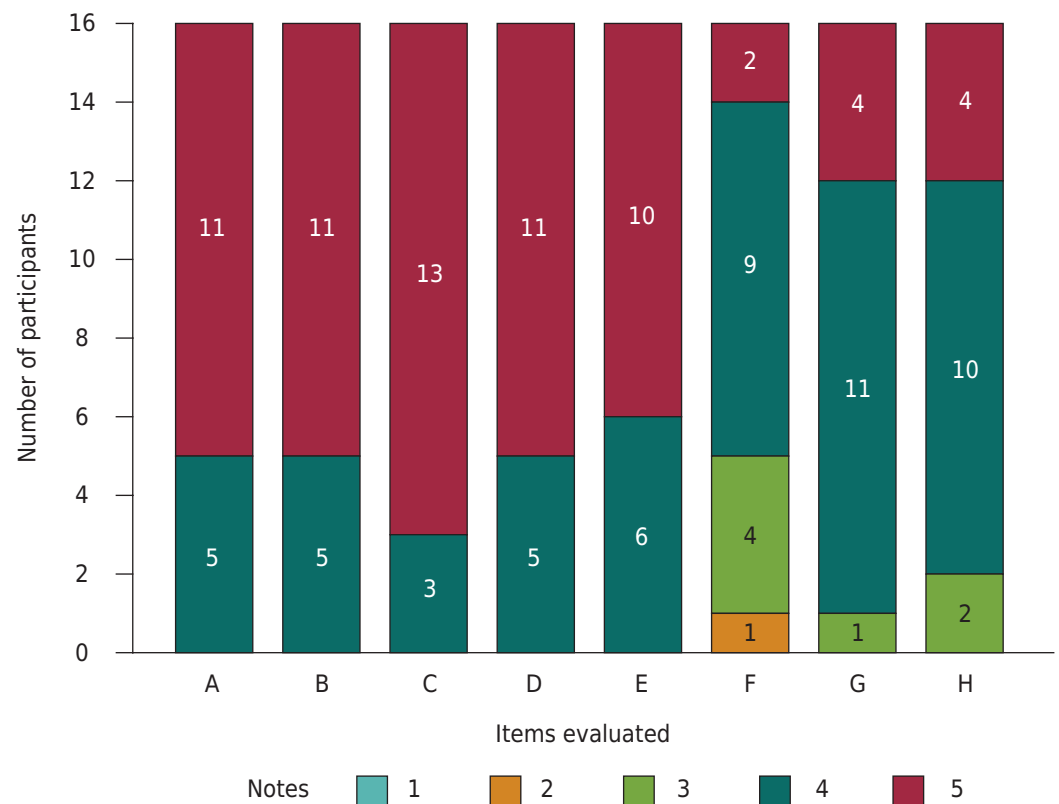


Figure 5. Evaluation carried out by the course participants, considering 5 the highest mark and 1 the lowest mark. Items evaluated: (A) the course met the expectations, and the objectives were achieved; (B) logical sequence of the course; (C) field class; (D) adequacy of practical classes; (E) the theoretical classes were illustrative, relevant and adjusted to the proposed subject; (F) Do you feel able to apply the acquired knowledge? (G) the knowledge acquired in the course are applicable in your work routine? (H) Do you feel able to replicate this knowledge in DSM courses that might be organized?

In addressing the foundations of DSM in module III, we observed little or no knowledge of the participants regarding the subject, resulting in innumerable basic questions about the subject. It is interesting to note that some participants were familiar with the use of GIS and geoprocessing techniques obtained in courses (face-to-face and distance learning) from educational institutions and on-line (e.g., in the Education Portal <https://www.portaleducacao.com.br>). The search for knowledge, according to what is demanded by these professionals, reinforces Lobry de Bruyn et al. (2017) statement that a multidimensional approach to soil education is needed that balances traditional models with new models to create a learning environment that facilitates changes and consequently learning. On the other hand, some basic concepts for building and managing a spatial database such as file formats for spatial data, coordinate reference systems, and directory naming standards, files and data tables, were not well known. This demonstrates that, regardless of the participants' experience, there is a need to level knowledge about GIS and spatial data.

We observed that the practice classes were fundamental in the appropriation of new knowledge, allowing an approximation of the content studied with the reality of the participants, corroborating studies by Minasny and McBratney (2016) and Arrouays et al. (2017), and this was reported as very positive in the evaluation of the present course. In spite of the advances in DSM in Brazil, with a recognized prominence in the world scenario, not only in the number of articles, but also with good citation indices (Cancian et al., 2018), the difficulties in understanding the DSM bases clearly demonstrate that the teaching of DSM in Brazil is restricted to a few Post-Graduate Programs in Soil Science or related areas (Dalmolin and ten Caten, 2015; Dalmolin et al., 2017). It should be emphasized that DSM is a recent subject compared to the other areas of soil science, demonstrating that the theoretical approach in DSM should have a greater workload in the course syllabus proposed here (Baca et al., 2013; Minasny and McBratney, 2016).

We observed that the knowledge provided by teaching DSM needs to be meaningful for the people involved. This is usually easily achieved as the learner is more familiar with the terms and concepts of the subject matter (Fazenda, 2014). The effectiveness of teaching DSM is directly related to the previous knowledge about the training of the participants, knowledge in pedology, understanding of the soil-landscape relationship, level of GIS training, and knowledge in statistics, as well as the distribution of the workload between modules. According to Hartemink et al. (2014), finding the balance between different professionals with deep and creative knowledge is a challenging task for soil science educators. Training technicians with different levels of knowledge in subjects related to DSM such as spatial modeling, multivariate statistics, organization, and use of soil databases, GIS, programming languages, etc., was a major challenge identified by Baca et al. (2013).

Knowing that teaching is not transferring knowledge, but creating possibilities for its own construction, the previous experience of the individual, used in the teaching-learning process, has become a differential in producing more significant results during the course. Minasny and McBratney (2016) reported that training in DSM creates learning possibilities, knowledge replication, and DSM techniques are moving from research to operational. This favored the understanding of the object of study in its environment, providing richness of details to the construction of knowledge, promoting broad reflections and interrelations, which diminished the problem of knowledge fragmentation (Morin, 2015).

Considering that the process of landscape interpretation and mapping is not always an individual task, the advantage of working with distinct backgrounds was evident, since it allowed the interconnection of contents from several areas of knowledge. This type of observation is common in all levels of learning, from the elementary school to higher education (Fazenda, 2014).

The baselines used to distribute class time per module should, as far as possible, consider the opinion of the participants and contemplate their needs. Considering that teaching requires critical reflection on the practice, the evaluations made by the participants can reveal flaws in the teaching strategies to which they are submitted and provide corrections for future practices. The main points reported were the need for more workload for field practice and study of the soil-landscape relationship, especially for participants in groups 1 and 2. It was also reported by all participants the need for more attention from instructors to the theoretical concepts of machine learning and obtaining predictor covariates, as well as practice in software employed in DSM. In a distance learning course about DSM, Baca et al. (2013) extended the total time in two weeks at the request of the trainees so that there was more time to access the content and perform the exercises with the support of the instructors.

In the evaluation of the course, although the participant's suggestions did not encompass the whole scenario involved in a DSM course, mainly because they did not consider the limitations of human and material resources, they were important and may indicate the readjustment of the workload. Still need a greater emphasis on the practical activities that consolidate the learning in DSM and the commitment of the participants in the continued study of the theoretical bases that encompass the DSM technique. The evaluation process allows to know the view of the participants, revealing their perceptions and serving as input for the reflection of the educators on the execution of the pedagogical practice (Fazenda, 2014).

Our observation was that the practice classes developed in an environment familiar to the participants motivated and facilitated the understanding of DSM. The teaching-learning process proved to be effective when using images from areas for training and validation of maps in landscapes that were common to the participants. The acquisition of pedological knowledge is a slow process, but field training in familiar landscapes can accelerate the learning process (Hudson, 1992). It was observed that the construction of the conceptual model of pedogenesis of the study area was a process guided by the previous knowledge of each of the participants, which manifested the various relationships that could be established for the production of the final map. In this line, Arrouays et al. (2017) emphasize that DSM should be conducted at a regional or local level to be consistent with its use and application, to ensure end-user involvement and efficient collection of soil data.

The low familiarization of the participants with DSM bases is related to the publication of results on this topic being restricted to scientific articles, emphasizing the need to disseminate this information through other means, for example, informative bulletins that aim the practical use of the knowledge (Minasny and McBratney, 2016). According to these authors, the distribution of computer code and manuals with protocols for DSM facilitates the practical application of this technique.

PronaSolos predicts an investment of approximately 1,3 billion dollars in 30 years to map Brazilian soils, and DSM methodologies will be used (Polidoro et al., 2016). According to Arrouays et al. (2017), there is a need for intensive training in pedology and DSM. This corroborates the reported results and experience show the need for planning the protocols involved in the five modules presented in this DSM courses, as well as prior knowledge of the participants skills.

The course described here is within this perspective and could be one of the guiding principles for future training for PronaSolos. As confirmed by Baca et al. (2013) and Vasques et al. (2013), and according to this work, it is possible to effectively capacitate professionals using free software and data available in soil databases and environmental covariates repositories. It is noteworthy that participants who reported not being able to replicate the course have little knowledge in soils, machine learning, and geoprocessing. Therefore, future courses should make a previous diagnosis,

aiming to group the participants in homogeneous groups in relation to knowledge in soils and DSM and thus plan the most appropriate DSM teaching protocol for each group demand.

CONCLUSIONS

The structure, focus, and time of each module should be based on the participants' needs. It is suggested that a survey should be carried out to consider the level of knowledge in relation to the topics addressed in DSM before the preparation and execution of the course, aiming at assisting in the planning of the techniques and in the level of deepening of the concepts.

The development of the course in an environment familiar to the users facilitated the teaching-learning process, since using common data helps the visualization and solution of problems.

The contribution given in the discussions according to the participants' experiences highlighted the importance of multidisciplinary in the teaching-learning process in DSM, because it is a technique that involves soil knowledge, statistics, and mathematics applied to geoinformation science to understand soil variability in the landscape.

The course was well evaluated by the participants, who reported that the practical classes were fundamental to approach the studied content to their reality.

This course could be a model to meet the needs of PronaSolos, which tend to have heterogeneous groups of participants, being necessary to plan specific protocols to tend the specific demand of each one.

ACKNOWLEDGEMENTS

We thank Dr Antonio Carlos de Azevedo for comments that greatly improved the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: Ricardo Simão Diniz Dalmolin (Lead), Jean Michel Moura-Bueno (Lead), Alessandro Samuel-Rosa (Lead), and Carlos Alberto Flores (Supporting).

Methodology: Ricardo Simão Diniz Dalmolin (Supporting), Jean Michel Moura-Bueno (Lead), Alessandro Samuel-Rosa (Lead), and Carlos Alberto Flores (Supporting).

Software: Jean Michel Moura-Bueno (Lead) and Alessandro Samuel-Rosa (Lead).

Validation: Ricardo Simão Diniz Dalmolin (Supporting), Jean Michel Moura-Bueno (Lead), and Alessandro Samuel-Rosa (Lead).

Formal analysis: Ricardo Simão Diniz Dalmolin (Supporting), Jean Michel Moura-Bueno (Lead), and Alessandro Samuel-Rosa (Lead).

Investigation: Ricardo Simão Diniz Dalmolin (Supporting), Jean Michel Moura-Bueno (Lead), Alessandro Samuel-Rosa (Lead), and Carlos Alberto Flores (Supporting).

Resources: Ricardo Simão Diniz Dalmolin (Supporting), Jean Michel Moura-Bueno (Lead), and Alessandro Samuel-Rosa (Lead).

Data curation: Ricardo Simão Diniz Dalmolin (Supporting), Jean Michel Moura-Bueno (Lead), and Alessandro Samuel-Rosa (Lead).

Writing - original draft: Ricardo Simão Diniz Dalmolin (Lead), Jean Michel Moura-Bueno (Lead), Alessandro Samuel-Rosa (Lead), and Carlos Alberto Flores (Supporting).

Writing - review and editing: Ricardo Simão Diniz Dalmolin (Lead), Jean Michel Moura-Bueno (Lead), Alessandro Samuel-Rosa (Supporting), and Carlos Alberto Flores (Supporting).

Visualization: Ricardo Simão Diniz Dalmolin (Lead), Jean Michel Moura-Bueno (Lead), Alessandro Samuel-Rosa (Supporting), and Carlos Alberto Flores (Supporting).

Supervision: Ricardo Simão Diniz Dalmolin (Lead), Jean Michel Moura-Bueno (Lead), and Alessandro Samuel-Rosa (Supporting).

Project Administration: Ricardo Simão Diniz Dalmolin (Lead).

Funding Acquisition: Ricardo Simão Diniz Dalmolin (Lead).

REFERENCES

- Amundson R, Berhe AA, Hopmans JW, Olson C, Sztein AE, Sparks DL. Soil and human security in the 21st century. *Science*. 2015;348:1261071. <https://doi.org/10.1126/science.1261071>
- Arrouays D, Lagacherie P, Hartemink AE. Digital soil mapping across the globe. *Geoderma Reg*. 2017;9:1-4. <https://doi.org/10.1016/j.geodrs.2017.03.002>
- Baca JFM, Vasques GM, Dart RO, Brefin MLMS, Olmedo GF. Capacitação em mapeamento digital de solos. Parte 1 - Cursos presenciais e à distância para técnicos da América Latina e Caribe. In: XXXIV Congresso Brasileiro de Ciência do Solo; 23 jul - 2 ago 2013; Florianópolis. Florianópolis: Sociedade Brasileira de Ciência do Solo; 2013.
- Bognola IA, Curcio GR, Gomes JBV, Caviglione JH, Uhlmann A, Cardoso A, Carvalho AP. Levantamento semidetalhado de solos do Município de Londrina. Londrina: IAPAR; 2011.
- Cancian LC, Dalmolin RSD, ten Caten A. Bibliometric analysis for pattern exploration in worldwide digital soil mapping publications. *An Acad Bras Cienc*. 2018;90:3911-23. <https://doi.org/10.1590/0001-3765201820180423>
- Chagas CS, Pinheiro HSK, Carvalho Junior W, Anjos LHC, Pereira NR, Bhering SB. Data mining methods applied to map soil units on tropical hillslopes in Rio de Janeiro, Brazil. *Geoderma Reg*. 2017;9:47-55. <https://doi.org/10.1016/j.geodrs.2017.03.004>
- Dalmolin RSD, ten Caten A. Mapeamento Digital: nova abordagem em levantamento de solos. *Investig Agrar*. 2015;17:77-86. <https://doi.org/10.18004/investig.agrar.2015.diciembre.77-86>
- Dalmolin RSD, ten Caten A, Dotto AC. Pedometria: uma breve contextualização nacional e mundial. *Boletim Informativo da Sociedade Brasileira de Ciência do Solo*. 2017;43:18-21. Available from: https://www.sbcs.org.br/wp-content/uploads/2018/01/boletimsbcs32017ebook_03_01_2018_10_45_30_id_36404.pdf
- Food and Agriculture Organization of the United Nations - FAO/ Intergovernmental Technical Panel on Soils - ITPS. Status of the world's soil resources (SWSR) - Main report. Rome: FAO / ITPS; 2015 [cited 2019 Jul 21]. Available from: <http://www.fao.org/3/a-i5199e.pdf>
- Fazenda ICA. Interdisciplinaridade: Um projeto em parceria. 7. ed. São Paulo: Loyola Jesuítas; 2014.
- Filzmoser P, Liebmann B, Varmuza K. Repeated double cross validation. *J Chemometrics*. 2009;23:160-71. <https://doi.org/10.1002/cem.1225>
- Hartemink AE. The use of soil classification in journal papers between 1975 and 2014. *Geoderma Reg*. 2015;5:127-39. <https://doi.org/10.1016/j.geodrs.2015.05.002>
- Hartemink AE, Balks MR, Chen Z-S, Drohan P, Field DJ, Krasilnikov P, Lowe DJ, Rabenhorst M, van Rees K, Schad P, Schipper LA, Sonneveld M, Walter C. The joy of teaching soil science. *Geoderma*. 2014;217-218:1-9. <https://doi.org/10.1016/j.geoderma.2013.10.016>

- Hartemink AE, McBratney A. A soil science renaissance. *Geoderma*. 2008;148:123-9. <https://doi.org/10.1016/j.geoderma.2008.10.006>
- Hengl T, Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG, Gonzalez MR. SoilGrids1km - Global soil information based on automated mapping. *PLoS ONE*. 2014;9:e105992. <https://doi.org/10.1371/journal.pone.0105992>
- Hengl T, Wheeler I, MacMillan RA. A brief introduction to Open Data, Open Source Software and Collective Intelligence for environmental data creators and users. *PeerJ Preprints*. 2018;6:e27127v2. <https://doi.org/10.7287/peerj.preprints.27127v2>
- Heung B, Ho HC, Zhang J, Knudby A, Bulmer CE, Schmidt MG. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*. 2016;265:62-77. <https://doi.org/10.1016/j.geoderma.2015.11.014>
- Hudson BD. The soil survey as paradigm based science. *Soil Sci Soc Am J*. 1992;56:836-41. <https://doi.org/10.2136/sssaj1992.03615995005600030027x>
- Instituto Brasileiro de Geografia e Estatística - IBGE. Manual técnico de pedologia. 3. ed. Rio de Janeiro: IBGE; 2015 [cited 2019 Out 21]. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv95017.pdf>
- IUSS Working Group WRB. World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. Rome: Food and Agriculture Organization of the United Nations; 2015. (World Soil Resources Reports, 106).
- Karatzoglou A, Smola A, Hornik K. kernlab: kernel-based machine learning lab. R package version 0.9-29; 2018. Available from: <https://CRAN.R-project.org/package=kernlab>.
- Kempen B, Brus DJ, Heuvelink GBM, Stoorvogel JJ. Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. *Geoderma*. 2009;151:311-26. <https://doi.org/10.1016/j.geoderma.2009.04.023>
- Lagacherie P, McBratney AB. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: Lagacherie P, McBratney AB, Voltz M, editors. *Digital soil mapping: an introductory perspective*. Amsterdam: Elsevier; 2007. p. 3-22.
- Liaw A, Wiener M. Classification and Regression by randomForest. *R News*. 2002;2:18-22.
- Lobry de Bruyn L, Jenkins A, Samson-Liebig S. Lessons learnt: sharing soil knowledge to improve land management and sustainable soil use. *Soil Sci Soc Am J*. 2017;81:427-38. <https://doi.org/10.2136/sssaj2016.12.0403>
- MacMillan RA, Hengl T. The future of predictive soil mapping. In: Hengl T, MacMillan RA, editors. *Predictive soil mapping with R*. Wageningen: OpenGeoHub foundation; 2019. p. 329-70.
- McBratney AB, Mendonça-Santos ML, Minasny B. On digital soil mapping. *Geoderma*. 2003;117:3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Minasny B, McBratney AB. Digital soil mapping: a brief history and some lessons. *Geoderma*. 2016;264:301-11. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Morin E. *A cabeça bem feita: Repensar a reforma, reformar o pensamento*. 22. ed. Rio de Janeiro: Bertrand Brasil; 2015.
- National Aeronautics and Space Administration - NASA. Landsat 7 science data users handbook. USA: NASA; 2009. Available from: https://landsat.gsfc.nasa.gov/wp-content/uploads/2016/08/Landsat7_Handbook.pdf.
- National Aeronautics and Space Administration - NASA/Jet Propulsion Laboratory - JPL. NASA Shuttle Radar Topography Mission Global 1 Arc Second [Data Set]. NASA EOSDIS Land Processes DAAC; 2013 [cited 2019 Jan 21]. Available from: <https://doi.org/10.5067/MEaSURES/SRTM/SRTMGL1N.003>
- Polidoro JC, Mendonça-Santos ML, Lumberras JF, Coelho MR, Carvalho Filho A, Motta PEF, Carvalho Junior W, Araújo Filho JC, Curcio GR, Correia JR, Martins ES, Spera ST, Oliveira SRM,

- Bolfe EL, Manzatto CV, Tosto SG, Venturieri A, Sá IB, Oliveira VA, Shinzato E, Anjos LHC, Valladares GS, Ribeiro JL, Medeiros PSC, Moreira FMS, Silva LSL, Sequinatto L, Aglio MLD, Dart RO. PronaSolos - Programa nacional de solos do Brasil (PronaSolos) - Dados eletrônicos. Rio de Janeiro: Embrapa Solos; 2016. (Documentos 183).
- Quantum Geographic Information System - QGIS. QGIS Development Team. Version 2.18. Las Palmas: Open Source Geospatial Foundation; 2017. Available from: https://www.qgis.org/pt_BR/site/forusers/download.html
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2017. Available from: <http://www.R-project.org/>.
- Rossiter DG. Past, present & future of information technology in pedometrics. *Geoderma*. 2018;324:131-7. <https://doi.org/10.1016/j.geoderma.2018.03.009>
- Rossiter DG. Methodology for soil resource inventories. 2nd ed. Enschede: International Institute for Aerospace Survey & Earth Sciences; 2000.
- Rossiter DG, Liu J, Carlisle S, Zhu A-X. Can citizen science assist digital soil mapping? *Geoderma*. 2015;259-260:71-80. <https://doi.org/10.1016/j.geoderma.2015.05.006>
- Samuel-Rosa A, Dalmolin RSD, Moura-Bueno JM, Teixeira WG, Alba JMF. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. *Sci Agric*. 2020;77:e20170430. <https://doi.org/10.1590/1678-992x-2017-0430>
- Samuel-Rosa A, Heuvelink GBM, Vasques GM, Anjos LHC. Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma*. 2015;243-244:214-27. <https://doi.org/10.1016/j.geoderma.2014.12.017>
- Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonça-Santos ML, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vågen T-G, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang G-L. Digital soil map of the world. *Science*. 2009;325:680-1. <https://doi.org/10.1126/science.1175084>
- Santos HG, Aglio MLD, Dart RO, Breffin MLMS, Souza JS, Mendonça LR. Distribuição espacial dos níveis de levantamento de solos no Brasil. In: XXXIV Congresso Brasileiro de Ciência do Solo; 23 jul - 2 ago 2013; Florianópolis. Florianópolis: Sociedade Brasileira de Ciência do Solo; 2013.
- Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbreiras JF, Coelho MR, Almeida JA, Araújo Filho JC, Oliveira JB, Cunha TJF. Sistema brasileiro de classificação de solos. 5. ed. rev. ampl. Brasília, DF: Embrapa; 2018.
- Scurr P, Franklin J, Chadwick OA, McArthur D. Predictive soil mapping: a review. *Prog Phys Geog*. 2003;27:171-97. <https://doi.org/10.1191/0309133303pp366ra>
- ten Caten A, Dalmolin RSD, Pedron FA, Mendonça-Santos ML. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos. *Rev Bras Cienc Solo*. 2011;35:53-62. <https://doi.org/10.1590/S0100-06832011000100005>
- Therneau T, Atkinson B. rpart: recursive partitioning and regression trees. R package version 4.1-15; 2019. Available from: <https://CRAN.R-project.org/package=rpart>.
- Valeriano MM, Rossetti DF. Topodata: Brazilian full coverage refinement of SRTM data. *Appl Geogr*. 2012;32:300-9. <https://doi.org/10.1016/j.apgeog.2011.05.004>
- Vasques GM, Dart RO, Baca JFM, Olmedo GF, Breffin MLMS. Capacitação em mapeamento digital de solos. Parte 2 - Estudo de caso: carbono do solo em Campos dos Goytacazes, RJ. In: XXXIV Congresso Brasileiro de Ciência do Solo; 23 jul - 2 ago 2013; Florianópolis. Florianópolis: Sociedade Brasileira de Ciência do Solo; 2013.
- Vaysse K, Lagacherie P. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Reg*. 2015;4:20-30. <https://doi.org/10.1016/j.geodrs.2014.11.003>
- Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer; 2002.