Digital elevation model quality on digital soil mapping prediction accuracy

Qualidade do modelo digital de elevação na acurácia do mapeamento digital de solos

Elias Mendes Costa[1*], Alessandro Samuel-Rosa[2], Lúcia Helena Cunha dos Anjos[1]

[1]Universidade Federal Rural do Rio de Janeiro/UFRRJ, Departamento de Solos, Seropédica, RJ, Brasil
[2]Universidade Tecnológica Federal do Paraná/UTFPR-SH, Santa Helena, PR, Brasil
*Corresponding author: eliasmccosta@gmail.com

ABSTRACT

Digital elevation models (DEM) used in digital soil mapping (DSM) are commonly selected based on measures and indicators (quality criteria) that are thought to reflect how well a given DEM represents the terrain surface. The hypothesis is that the more accurate a DEM, the more accurate will be the DSM predictions. The objective of this study was to assess different criteria to identify the DEM that delivers the most accurate DSM predictions. A set of 10 criteria were used to evaluate the quality of nine DEMs constructed with different data sources, processing routines and three resolutions (5, 20, and 30 m). Multinomial logistic regression models were calibrated using 157 soil observations and terrain attributes derived from each DEM. Soil class predictions were validated using leave-one-out cross-validation. Results showed that, for each resolution, the quality criteria are useful to identify the DEM that more accurately represents the terrain surface. However, for all three resolutions, the most accurate DEM did not produce the most accurate DSM predictions. With the 20-m resolution DEMs, DSM predictions were five percentage points less accurate when using the more accurate DEM. The 5-m resolution was the most accurate DEM overall and resulted in DSM predictions with 44% accuracy; this value was equal to that obtained with two coarser resolution, lower accuracy DEMs. Thus, identifying the truly best DEM for DSM requires assessment of the accuracy of DSM predictions using some form of external validation, because not necessarily the most accurate DEM will produce the best DSM predictions.

Index terms: Multinomial logistic regression; predictor variables; collinearity; Shannon entropy.

RESUMO

Modelos digital de elevação (MDE) usados no mapeamento digital de solo (MDS) são comumente selecionados com base em medidas e indicadores (critérios de qualidade) que refletem como um determinado MDE representa a superfície da paisagem. A hipótese é que, quanto mais acurado for um DEM, mais acuradas serão as predições do MDS. O objetivo deste estudo foi avaliar diferentes critérios para identificar o MDE que fornece as predições mais acuradas do MDS. Um conjunto de 10 critérios foi utilizado para avaliar a qualidade de nove MDEs construídos com diferentes fontes de dados, rotinas de processamento e três resoluções (5, 20 e 30 m). Modelos de regressão logística multinomial foram calibrados usando 157 observações de solo e atributos de terreno derivados de cada MDE. As predições de classe de solo foram validadas usando validação cruzada *leave-one-out*. Os resultados mostraram que, para cada resolução, os critérios de qualidade são úteis para identificar o MDE que representa mais precisamente a superfície da paisagem. No entanto, para todas as três resoluções, o MDE mais acurado não produziu as predições mais acuradas do MDS. Com os MDEs de resolução de 20 m, as predições do MDS foram cinco pp menos acuradas ao usar o MDE mais acurado. A resolução de 5 m foi o MDE mais acurado e resultou em predições com 44% de acurácia; este valor foi igual ao obtido com duas resoluções mais grossas, MDEs com menor acurácia. Assim, a identificação do melhor MDE para o MDS exige a avaliação da acurácia das predições do MDS usando alguma forma de validação externa, porque não necessariamente o MDE mais acurado produzirá as melhores predições do MDS.

Termos para indexação: Regressão logística multinomial; variáveis preditoras; colinearidade; entropia de Shannon.

INTRODUCTION

Digital soil mapping (DSM) uses statistical models to quantify the correlation of soil attributes with environmental conditions to make predictions at locations not sampled. In these models, soil attributes in a particular location, whether continuous or categorical, are taken as random dependent variables. Approximate information about soil-forming factors is included in these models as deterministic predictor variables, for example, digital elevation models (DEM) and satellite images and indexes and attributes derived from them, and categorical maps depicting the geology, geomorphology, and land use. The large number and variety of deterministic predictor variables usually enables DSM models to explain a large proportion of the soil spatial variation, e.g. more than 3/4 (Heung et al., 2016; Hengl et al., 2017).

Predictor variables are often chosen based on their availability, the required level of spatial detail of DSM predictions, and/or knowledge of their ability to explain soil spatial variation (Miller et al., 2015). For example, geological and climatic data, despite the strong conceptual correlation with soil attributes, are frequently generalized, thus rarely used if DSM predictions with fine spatial detail are required (ten Caten et al., 2012). Most DSM projects prioritize satellite images and DEMs, which are available at multiple levels of spatial detail and have excellent spatial coverage (ten Caten et al., 2012; Vasques; Grunwald; Myers, 2012). However, a major difficulty is the selection of a single - optimally the best - satellite image and DEM among the many available for a given DSM project. This is because these data carry unknown and varying errors that could reduce the correlation with soil attributes and thus affect the accuracy of DSM predictions. These errors arise from the complex interplay between the methods of data generation, analytical procedures and characteristics of each site (Fisher; Tate, 2006; Florinsky, 1998; Hirt; Filmer; Featherstone, 2010).

There are several strategies to select a single satellite image and DEM for a DSM project. In the case of DEMs, a popular strategy is to assess its quality (Baltensweiler et al., 2017; Cavazzi et al., 2013; Chagas et al., 2010; Moura-Bueno et al., 2016; Neumann; Roig; Souza, 2012; Penížek et al., 2016; Pinheiro et al., 2012). In general, the DEM´s quality is defined by combining various indicators and measures, presumably related to the capacity of a DEM and derived attributes to represent terrain surface features. One example is vertical error, a measure computed at a finite number of ground control points and confronted with the maximum DEM error allowed by legislation or specified by the DEM manufacturer (Baltensweiler et al., 2017; Chagas et al., 2010; Nelson; Reuter; Gessler, 2009). Some indicators are agreement among original and derived stream network and contour lines, and configuration of derived contribution basins (Chagas et al., 2010; Neumann; Roig; Souza, 2012; Pinheiro et al., 2012). As additional criteria evaluation of number and extent of spurious artifacts, and descriptive statistics of DEM and derived attributes (Cavazzi et al., 2013; Chagas et al., 2010; Penížek et al., 2016; Pinheiro et al., 2012; Thompson; Bell; Butler, 2001).

Despite their widespread use, there is little empirical evidence that a DEM selected using any of the quality criteria produces more accurate DSM predictions than the rejected DEMs. This is because the predictions of DSM models calibrated using alternative, competing DEMs are seldom validated. The objective of this study was to assess if the DEM quality criteria identify the single DEM that delivers the most accurate DSM predictions as assessed via external validation. In this sense, a case study was carried out in an area of moderately rugged terrain, with variation of land use/cover, in southeastern Brazil.

MATERIAL AND METHODS

Study area and soil data

The study area is located in the municipality of Pinheiral (São José do Pinheiro Farm), *Médio Paraíba do Sul* region of Rio de Janeiro State, Southeastern Brazil (Figure 1). The area has 1462 ha and occupies the lower section of the *Ribeirão Cachimbal* watershed, a tributary of the *Paraíba do Sul* River. The climate is of Cwa type, with a warm and rainy summer and a dry winter with mild temperatures (Alvares et al., 2014). Annual rainfall is around 995 mm with average annual temperature of 20.9 °C (Portilho et al., 2011). It is inserted in the Atlantic Forest Biome, and original vegetation was submontane seasonal semi-deciduous forest. Elevations vary from 350 to 558 m a.s.l., with strongly undulated relief (slopes up to 75%). The geology shows dominance of metamorphic rocks (porphyroclastic microcline gneiss, biotite-muscovite gneiss and biotite gneiss), plus basic dykes (gabbro and basalt) and Tertiary (fluvial) and Quaternary (colluvial-alluvial) sediments (Sanson; Ramos; Mello, 2006).

Soil dataset is comprised of 157 point-support observations, yielding a moderately high sampling density (one observation per 10 ha, Figure 1). This includes legacy data from previous studies, a soil survey (27 observations), and data recorded in field trips of a graduate course of UFRRJ in the years of 2004, 2012, and 2013 (35 observations). Other 95 observations were obtained just for this study. Their location was defined intentionally, in a tacit manner, to cover most of the geological and geomorphological variation (pedological demand), and to have about the same number of observations for each soil taxon (statistical demand). Out of these, 20 complete soil profiles were described, analyzed and classified according to Brazilian soil survey standards. The remaining 75 observations were made with an auger and the soil taxon inferred from the interpretation of soil attributes and landscape conditions using tacit knowledge. The data is available at the Free Brazilian Repository for Open Soil Data (www.ufsm.br/febr) with identification code ctb0002.

Thirteen soil classes were identified in the study area (Figure 2). Due to the relatively small sample size (157), the resulting number of observations in some soil classes was too low (< 9) to calibrate the DSM models. Thus, soil classes were grouped based on the similarity of their definitions, resulting in six categories that were used for modeling.
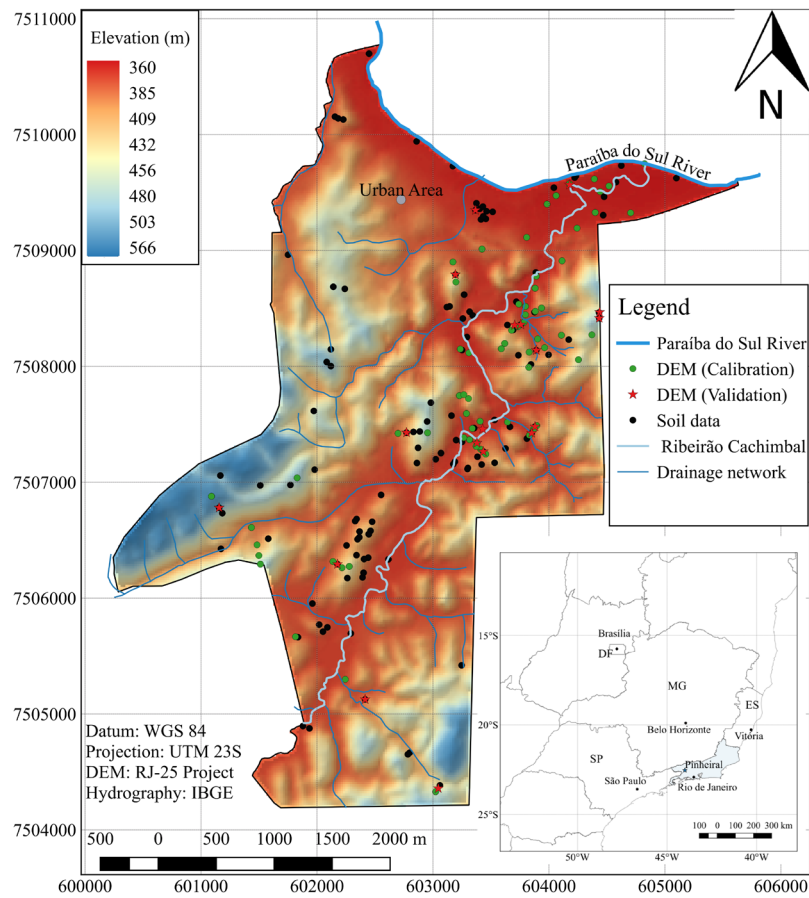
Figure 1: Location of study area in Southeastern Brazil, municipality of Pinheiral (inset map, blue star), and spatial distribution of point elevation (green – calibration; red – validation) and soil data (black). The drainage network and DEM background are suggestive of the current landscape setting.
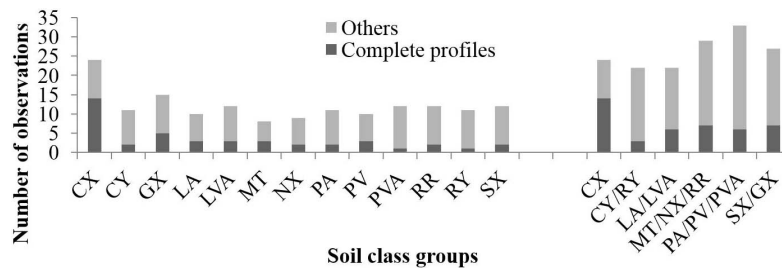

Figure 2: Distribution of soil classes and soil groups of the 157 soil profiles observed in the study area according to the Brazilian Soil Classification System (SiBCS). Classes *Latossolo Amarelo* (LA) and *Latossolo Vermelho-Amarelo* (LVA), and *Argissolo Amarelo* (PA), *Argissolo Vermelho* (PV) and *Argissolo Vermelho-Amarelo* (PVA), were grouped at the highest category level of the SiBCS. *Chernossolo Argilúvico* (MT), *Nitossolo Háplico* (NX) and *Neossolo Regolítico* (RR) were grouped based on commonality of some chemical, physical and/or morphological attributes, and the same parent material. *Cambissolo Flúvico* (CY) and *Neossolo Flúvico* (RY) all originated from fluvial sediments. *Planossolo Háplico* (SX) and *Gleissolo Háplico* (GX) were grouped based on their limited drainage, and clay illuviation and gleization processes, respectively. The *Cambissolo Háplico* (CX) class comprises soils with variable attributes, and greatest number of observations in the area.

Elevation data

To assess whether the quality criteria allows to selecting the DEM which results in most accurate DSM predictions, an elevation database was built using data collected in the field and other freely available on the Internet. These data were used to construct DEMs with three spatial resolutions (5, 20, and 30 m) using different processing routines to simulate scenarios of DEM quality (Table 1). The simulated scenarios, denoted by 'a', 'b', and 'c', correspond to conditions of high, medium and low expected capacity of a DEM to accurately depict the current terrain surface conditions. For each spatial resolution, this was achieved by using different quantities/levels of information to produce a DEM, e.g. DEM5a > DEM5b > DEM5c (Table 1). The data were handled using the Universal Transverse Mercator (UTM) projection, zone 23S, with horizontal *datum* WGS-84, while maintaining the original vertical *datum*.

The elevation data sources and processing are described below.

POINTS. Ground control points (86) measured in the field using a L1/L2 GNSS static receiver, whose location coincided with soil sampling locations obtained in this study. The elevation data were post-processed to achieve centimeter accuracy in orthometric elevation values using as vertical *datum* the tide gauge in Imbituba, Santa Catarina, Brazil. A set of 18-ground control points were reserved for later use as DEM validation data. They were selected as to represent the range of elevation values in the study area with a good spatial distribution given by the available data. The remaining 67-ground control points were used for DEM preparation.

IBGE. Cartographic database in vector format prepared by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) using plano-altimetric maps at a cartographic scale of 1:50000 (sheets Volta Redonda SF-23-Z-A-V-2 and Piraí-SF-23-Z-A-VI-1). It includes contour lines with an equidistant vertical interval of 20 m, and stream network. The contour lines and stream network, the latter assigned the correct flow direction, plus 67 ground control points, were used to generate DEMs at the three spatial resolutions evaluated. For this procedure, it was used the ANUDEM algorithm (Hutchinson, 1989) implemented in ArcGIS.

RJ-25. DEM of the RJ-25 Project with spatial resolution of 20 m (sheets 27432ne and 27441no). The DEM was obtained by analytical photogrammetric processing of air-borne photographs with a cartographic scale of 1:25000 and spatial resolution of 0.70 m. The DEM was interpolated to a spatial resolution of 5 m using the ANUDEM algorithm without the stream network data. From this DEM, the contour lines with an equidistant vertical interval of 10 m were extracted, and subsequently smoothed out. These contour lines, plus the 1:50000 IBGE stream network data and 67 ground control points (see above), were used to generate DEMs at the three evaluated spatial resolutions.

Table 1: Digital elevation models prepared with different spatial resolutions using various data sources and processing routines to simulate scenarios of digital elevation model data quality for digital soil mapping.

Resolution(m)	Scenario[1]	Code	Processing routine[2]
5	a	DEM5a	ANUDEM (CV 10 + HD + PTS)
	b	DEM5b	ANUDEM (CV 20 + HD + PTS)
	c	DEM5c	ANUDEM (RJ25)
20	a	DEM20a	ANUDEM (CV 10 + HD + PTS)
	b	DEM20b	ANUDEM (CV 20 + HD + PTS)
	c	DEM20c	RJ25
30	a	DEM30a	ANUDEM (CV 10 + HD + PTS)
	b	DEM30b	ANUDEM (CV 20 + HD + PTS)
	c	DEM30c	TOPODATA

[1] Simulated scenarios of DEM quality for DSM, henceforth denoted by 'a', 'b', and 'c', correspond to conditions of high, medium and low expected capacity of a DEM to accurately depict the current terrain surface conditions, respectively.
[2] CV20 = 20-m equidistant contour lines from the IBGE database. CV10 = 10-m equidistant contour lines derived from the RJ-25 Project DEM after interpolation to 5 m spatial resolution. HD = Stream network data from the IBGE database. PTS = elevation data from 67 ground control points. RJ25 = DEM data from the RJ-25 Project. TOPODATA = elevation data from the TOPODATA Project. ANUDEM = program developed by Hutchinson (1989) for creation of hydrologically correct DEMs.

TOPODATA. The DEM from the TOPODATA Project is a product of the post-processing of the original Shuttle Radar Topography Mission (SRTM) DEM by the *Instituto Nacional de Pesquisas Espaciais* (INPE). The post-processing was done to fill data gaps and to refine the SRTM DEM spatial resolution from 90 m to 30 m by using ordinary kriging (Valeriano; Rossetti, 2012). Sheet 22S45_ZN was used in this study.

A set of 10 quality criteria (Table 2) were used to assess the quality of three DEMs in each spatial resolutions (Figure 3) (Baltensweileret al., 2017; Cavazzi et al., 2013; Chagas et al., 2010; Moura-Bueno et al., 2016; Neumann; Roig; Souza, 2012; Penížek et al., 2016; Pinheiro et al., 2012; Thompson; Bell; Butler, 2001). Descriptive statistics are minimum, maximum, mean, and standard deviation of the elevation. Mean error (ME), root mean square error (RMSE) and median absolute error (MDAE) of elevation were computed at 18 external ground control points using,

$$ME = \frac{\sum_{i=1}^{n} d_i}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \quad (2)$$

$$MDAE = median\left(\left|d_i - median\left(d\right)\right|\right) \quad (3)$$

where d is the difference between the DEMs evaluated and the ground control points and n is the number of elevation points tested. The ideal condition would be ME equal to 0 and RMSE and MDAE as lower as possible. MDAE is particularly interesting because it is robust to outliers. Also, the normal distribution of the errors was evaluated as suggested by Höhle and Höhle (2009).

Table 2: Quality criteria used to select digital elevation models for digital soil mapping.

| Quality criteria	Description and interpretation	Reference
Descriptive statistics of elevation	Summary statistics of the elevation values across the study area. Shows the overall difference between the DEMs and the possible implications for the representation of the terrain. It is expected that DEMs that represent the surface with highest accuracy will have a greater amplitude of values. Some studies suggest that the greater the amplitude of elevation values, the lower DEM error statistics.	Chagas et al. (2010); Moura-Bueno et al. (2016); Neumann et al. (2012); Penížek et al. (2016); Pinheiro et al. (2012)
Error statistics	It is a comparison between the elevation values sampled from a DEM against ground control points (reference). Evaluates the vertical accuracy of the DEMs. It is expected that DEMs with lower values of root mean square error (RMSE) and mean error (ME) close to the zero more accurately represent the actual terrain surface.	Chagas et al. (2010); Baltensweiler et al. (2017); Moura-Bueno et al. (2016); Thompson, Bell and Butler (2001)
Spurious depressions	A spurious depression is a region of cells that drain inwards to a pit. The presence of a great number of large spurious depressions can significantly affect derivation of drainage network and characterization of terrain surface, consequently, the soil-landscape relationship in DSM modelling.	Chagas et al. (2010); Moura-Bueno et al. (2016); Pinheiro et al. (2012)
Slope classes	The distribution of slopes, according to predefined classes, is useful to identify the main differences between DEMs and possible implications for terrain representation. The evaluation of DEMs based on slope becomes more comprehensible when they are grouped into class intervals commonly used in soil characterization (Santos et al., 2015). It is expected that more accurate DEMs have a larger range of values and a larger frequency of steep slopes.	Pinheiro et al. (2012)
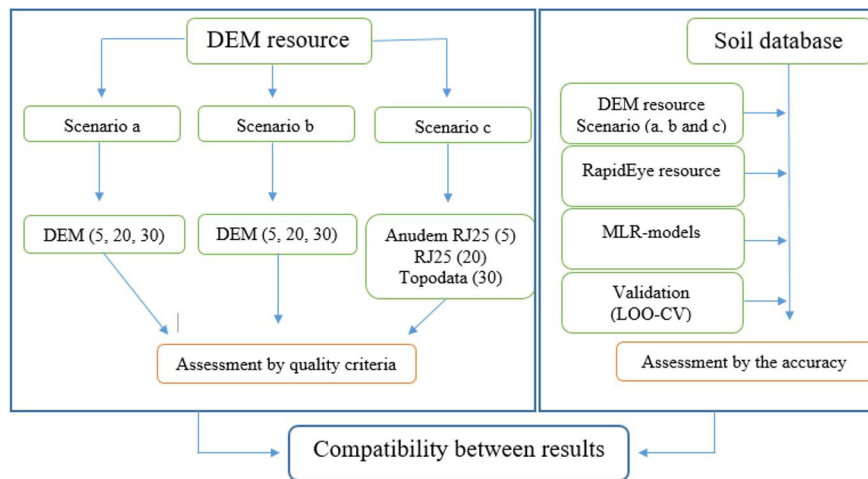
Figure 3: Flowchart of compatibility between the evaluation of DEMs that represent different scenarios and resolutions, and evaluation of predictions in the DSM using each DEM.

Spurious depressions were quantified in terms of number of depressions, number of cells, and relationship with total area. Slope classes were defined according to Santos et al. (2015), i.e. 0-3, 3-8, 8-20, 20-45, 45-75, and > 75%.

A commonly used quality criteria is the agreement between stream network and contour lines derived from the DEMs and some reference dataset, where the closer the agreement the higher the DEM quality (Chagas et al., 2010; Neumann; Roig; Souza, 2012; Pinheiro el al., 2012). This quality criterion was not used in this study because it would require making the strong assumption that a given dataset chosen as reference was an accurate representation of the terrain surface. Besides, the literature shows that its usage has consistently produced biased results. This is because the reference stream network and contour lines are generally used as data sources to produce DEMs - due to data availability limitations. Naturally, the stream network and contour lines derived from these DEMs have always presented a higher agreement with the reference dataset.

Predictor variables

A set of 13 terrain attributes was derived from each DEM, maintaining the DEM resolution, using SAGA GIS default settings. They are: catchment area (m²), aspect (degree), flow direction (dimensionless), vertical distance to stream network (m), stream network base level (m), relative slope position (dimensionless), profile curvature (m^{-1}), plan curvature (m^{-1}), slope (degree) elevation (m), topographic factor of the Universal Soil Loss Equation (dimensionless), convergence index (%), and topographic wetness index (dimensionless). The aspect, being a circular variable, was transformed to North exposition (dimensionless).

The five bands of a RapidEye satellite image (17 August 2011), provided by *Ministério do Meio Ambiente* (MMA), were also used as predictor variables. The image was corrected for atmospheric effects using the atmospheric radiative transfer model 6S (Antunes; Debiasi; Siqueira, 2014). The bands were used to generate the Normalised Difference Vegetation Index (NDVI) and the Soil Adjusted Vegetation Index (SAVI) predictor variables.

Evaluation of gains in DSM predictions

The capacity of each DEM in accurately making soil predictions was evaluated by calibrating multinomial logistic regression (MLR) models. The MLR is an extension of the binary logistic regression that handles more than two categories in the dependent variable. In this study, the dependent random variable C is composed of groups of soil classes and contains $k = 6$ categories (Figure 2). The correlation between predictor variables and probability of occurrence of each of the k classes of the dependent variable was quantified. The logistic function is represented by:

$$logit_j(s) = log\left(\pi_j(s) / \pi_k(s)\right) = \alpha_j + \beta'_j x(s), \ j = 1, ..., k-1, \quad (4)$$

Where $logit_j(s)$ is the natural logarithm of the ratio between the probability $\pi_j(s)$ that a given soil observation belongs to category j, subject to the values of the p predictor variables contained in the vector $x(s)$, and the probability $\pi_k(s)$ that that soil observation belongs to category k taken as reference (Agresti, 2002).

In this study, the reference category was the CX soil class group. The intercept of the logit model estimated for category j is given by α_j, while β'_j is a vector with the coefficients estimated for each of the p predictor variables.

Two groups of $m = 10$ MLR-models were calibrated using two different strategies. The first contained all predictor variables (complete version). The second had a subset of the predictor variables selected as to reduce collinearity (reduced version). The criteria used to select predictor variables was the absolute bivariate correlation, $|r|$, with the goal of keeping only predictor variables with $|r| < 0.90$.

For both complete and reduced versions, a MLR-model taken as baseline was calibrated using only the bands of the RapidEye sensor and derived indexes, $p = 7$; while the reduced version used bands 1 and 5 and the NDVI, $p = 3$. The other nine models, in the complete and reduced versions, were calibrated by adding terrain attributes derived from each of the nine DEMs to the baseline model, totaling $p = 20$ and $p = 13$ predictor variables, respectively. For all models, terrain attributes selected for the reduced strategy were: elevation, slope, north exposition, convergence index, catchment area, topographic wetness index, stream network base level, profile and plan curvature, and flow direction.

All MLR-models were evaluated using as quality measure the overall accuracy. This quantity measures the proportion of correct classifications made by a MLR-model. The calibration accuracy was calculated by cross-tabulating observed and fitted values - confusion matrix - of each calibrated MLR-model, while the validation accuracy was calculated by cross-tabulating observed and predicted values - error matrix - of a leave-one-out cross-validation. For the latter, one observation was temporarily removed from the dataset to be used as a validation observation. Each MLR-model was calibrated with remaining observations and used to predict the category of the validation observation. This procedure was repeated until all observations, $n = 157$, had been used once as validation observation. The calibration and validation overall accuracy was given by:

$$A = \frac{\sum_{i=1}^{k} E_{ij}}{n} \tag{5}$$

Where E is the confusion or error matrix of dimensions k x k, respectively.

The influence of DEM quality on the overall accuracy of digital soil maps was evaluated in a series of steps. The first, calibration and validation accuracies among

the MLR-models calibrated with all predictor variables. From the baseline MLR-model, it was determined the relative gain of using the terrain attributes as predictor variables. By comparing calibration against validation accuracy, it was evaluated the sensitivity of MLR-models to the calibration dataset, including soil observations and predictor variables with strong collinearity. The same procedure was adopted for analyzing the MLR-models calibrated with the reduced set of predictor variables. Comparison of the overall accuracy of MLR-models, calibrated using the complete and reduced set of predictor variables, was used to assess importance of reducing collinearity between predictor variables.

Finally, the set of MLR-models with superior leave-one-out cross-validation results were used to predict the probability of occurrence of each of the k classes of the random categorical variable C in a regular grid of points with 5-m separation distance covering the entire study area. The uncertainty of the spatial prediction of C at a given spatial location s was quantified using the Shannon entropy: where (c_i, s) is the estimated probability that the random variable C, at location s, takes the value c_i among the k possible values (Agresti, 2002). The usage of logarithm with base k scales the value of $H(s)$ between 0 and 1,

$$H(s) = -\sum_{i=1}^{k} \hat{\pi}(c_i, s) \log_k \hat{\pi}(c_i, s) \tag{6}$$

where 0 means no conditional uncertainty - one of the k categories has probability of occurrence equal to 1, and 1 the maximum conditional uncertainty - all k categories have equal probability of occurrence.

RESULTS AND DISCUSSION

DEM characteristics and quality for DSM

Overall, the evaluation of DEM characteristics and quality for DSM matched those from recent literature. For the spatial resolutions tested, the DEMs generated using IBGE database (DEM5b, DEM20b and DEM30b) showed values of minimum elevation smaller than other DEMs (Table 3). Meanwhile, larger minimum values were observed in the DEMs generated without ground control points (DEM5c, DEM20c and DEM30c). The maximum elevation value showed little variation between resolutions and, in general, DEMs using IBGE database (scenario 'b') showed the smallest maximum values. The lowest mean elevation values were observed in scenario 'b', around 402 m, while in scenarios 'a' and 'c' it was about 409 m. The same is true

for the standard deviation that was 37.3 m in scenario 'b' and 38.9 m in scenario 'a'. The greatest variation in standard deviation between resolutions for scenario 'c' is due to the different data source for this scenario. For 'b' and 'c' there is practically no difference between resolutions. Overall, DEM resolution influence on the descriptive statistics was little. In general, the ground control points separate for validation have descriptive statistics close to those for the DEMs.

As all the errors showed a normal, thus there is no bias in using criteria like RMSE and ME to evaluate errors (Höhle; Höhle, 2009). These are the two most commonly used measures of DEM accuracy (Fan; Atkinson, 2015). For all three spatial resolutions, the DEMs generated from

scenario 'a' had smaller RMSE and ME closer to zero. MDAE followed the same pattern except for resolution of 30 m. DEM20c also had a small ME, but it was associated with large RMSE and MDAE bigger the scenario 'a'. Of the two DEMs available on the web, TOPODATA (DEM30c) presented RMSE, MDAE and ME larger than the RJ25 (DEM20c). All DEMs overestimated the elevation values in 3.6to 6.6 m. As for the descriptive statistics, the resolution had little influence on the error measures.

Within each resolution, there was higher incidence of spurious depressions in scenario 'c', i.e. without use of stream network data and/or hydrologically consistent interpolation algorithm (Table 4).Without the use of

Table 3: Descriptive statistics (minimum, maximum and mean values, and standard deviation), mean error (ME), root mean square error (RMSE) and median absolute error (MDAE) of each digital elevation model evaluated (unit: meters). The three lowest values of ME, RMSE and MDAE are in boldface.

Code	Descriptive statistics				Validation		
	Minimum	Maximum	Mean	SD	ME	RMSE	MDAE
DEM5a	347.0	564.9	408.5	38.9	4.1	7.1	4.3
DEM5b	339.6	548.1	402.1	37.2	4.4	10.1	7.0
DEM5c	360.6	563.3	409.1	37.8	4.6	7.4	5.6
DEM20a	351.6	564.8	408.3	38.9	3.6	6.6	3.7
DEM20b	336.0	548.1	402.1	37.3	3.9	9.2	7.6
DEM20c	361.0	566.0	409.2	38.1	3.6	9.4	5.4
DEM30a	353.6	565.6	408.0	38.9	3.6	7.0	4.8
DEM30b	338.2	548.5	401.9	37.3	4.7	9.8	4.2
DEM30c	360.0	560.0	410.1	36.4	6.6	10.8	8.9
GCP	358.2	476.5	401.6	38.9	------	------	------

Note: SD = standard deviation; GCP= Ground control points (validation, 18 points).

Table 4: Number of spurious depression and cells involved and the relationship of depressions with the total area for each digital elevation model evaluated. The lowest values in each resolution are in boldface.

Code	Number of depressions	Number of cells	Relationship with total area (%)
DEM5a	101	202	0.03
DEM5b	57	114	0.02
DEM5c	978	1954	0.33
DEM20a	32	64	0.18
DEM20b	36	72	0.20
DEM20c	103	548	1.50
DEM30a	15	32	0.20
DEM30b	22	46	0.28
DEM30c	57	269	1.66

smoothing interpolator, DEMs derived from remote sensors carry with them, noise inherent to data collection method (Hengl; Gruber; Shestha, 2004).Consequently, this scenario resulted in larger numbers of cells involved and larger surface area of depressions in relation to total area. DEMs from scenario 'a', which used 10-m equidistant contour lines, showed a trend towards better (smaller) values. However, within each resolution, results for DEMs of scenarios 'a' and 'b' were only slightly different. The influence of the resolution was different from that observed in the descriptive statistic. Increasing the resolution produced a considerable decrease in the number of depressions and an increase of their size. This is due to the smoothing effect of having larger pixels.

The derived slope surfaces were considerably influenced by the spatial resolution in all three data scenarios (Figure 4). DEM5a, DEM5b, and DEM20cwere the only DEMs that yielded slopes >75%. Likewise, the resolution and scenarios influenced the relative distribution of slope classes. As the resolution increased, there was a decrease in the proportion of the steep slope class (45-75%), resulting in increasing the slopes between 20 and 45%. The same happened with scenarios, except for the resolution of 20 m, where scenario 'c' tends to produce a lower proportion of the steep-slope class. The plain class (0-3%) showed larger proportion in DEMs derived from scenario 'b' and smaller for scenario 'a' DEMs. The strongly sloping class (20-45%), except for DEM30c, showed the largest proportion, and it was largest in DEM5b.

When all cartographic data (hydrography and contour lines) and accurate available ground points are used in conjunction with hydrologically consistent algorithms, the ability of the resulting DEM in accurately represent morphology of the modeled terrainsurface (Tables 3, 4 and Figure 4) is better. This corroborates some literature results (Chagas et al., 2010; Moura-Bueno et al., 2016; Neumann; Roig; Souza, 2012; Pinheiro et al., 2012). Moreover, the finer the DEM spatial resolution the higher the accuracy of elevation values.

Based on the concept of DEM quality above, for each spatial resolution, DEMs of scenario 'a' can be considered as having superior quality, corroborating the initial hypothesis. Hence, DEMs of scenario 'c' would be of inferior quality. When evaluated across all three resolutions, DEM5a can be considered as the highest quality DEM (Table 5). As stated, this is an effect of the spatial resolution of a DEM on its capacity to represent the terrain surface. By definition, a DEM is more likely to represent the terrain surface when produced using
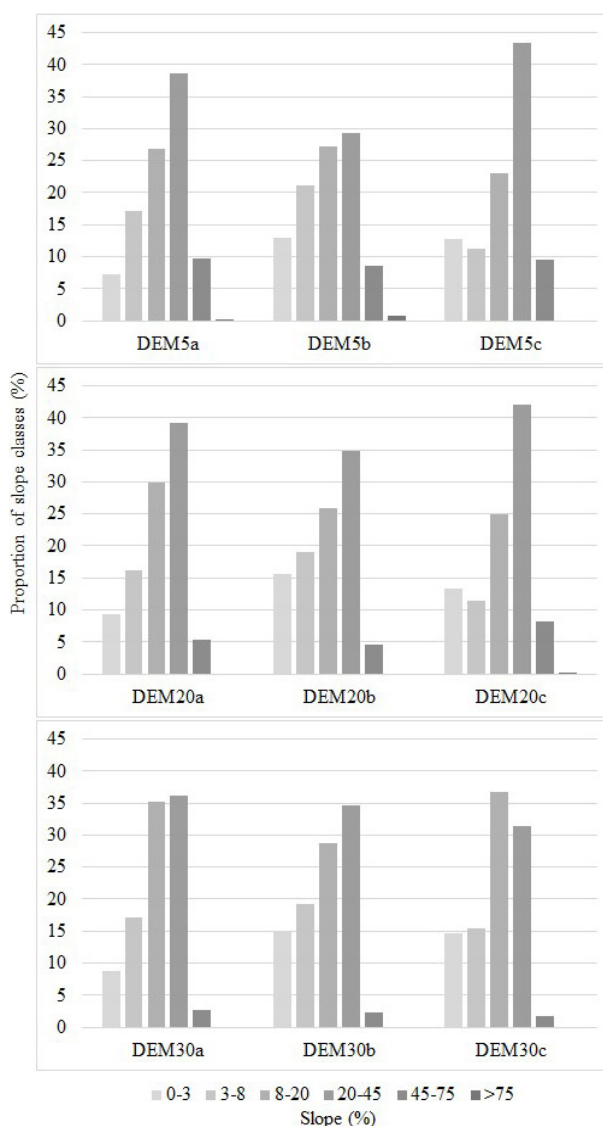

Figure 4: Proportion of slope classes in the study area (according to Santos et al., 2015).

Table 5: Summary results of the DEM quality assessment.

Criteria	5 m	20 m	30 m	Overall
Descriptive statistics of elevation	DEM5a	DEM20a	DEM30a	DEM5a
Error statistics	DEM5a	DEM20a	DEM30a	DEM20a
Spurious depressions	DEM5a,b	DEM20a	DEM30a	DEM5a,b
Slope classes	DEM5a,b	DEM20c	DEM30a	DEM5a,b

more sophisticated algorithms and more data with larger accuracy (Maynard; Johnson, 2014). Thus, DEM20a would be the second highest quality DEM, because it presented better error statistics than all other DEMs, and DEM30b and DEM30c would be the poorest DEMs.

DSM accuracy

The MLR-models calibrated using the superior DEM5a showed a trend towards larger overall calibration accuracy in both complete and reduced versions (Table 6). However, the difference to MLR-models calibrated with the poorest DEMs was small, and sometimes less than or equal to one percentage point (pp), such as DEM5a as compared to DEM30c in the complete version. There was a tendency for the greater the number of predictor variables the larger the similarity of accuracy values during calibration across MLR-models. However, higher performance of superior DEMs did not hold during validation of MLR-models in both complete and reduced versions: MLR-models calibrated using inferior DEMs tended to have higher validation accuracy. However, the difference in accuracy by using poorer or superior DEMs was small, usually less than 1 pp.

Overall, there was no direct relation between DEM quality - given the DEM quality concept above - and validation accuracy of calibrated MLR-models: the most accurate MLR-model, better result for DSM, was calibrated with the worse DEM for resolution of 5 m, DEM5c. In some cases, increasing DEM spatial resolution increased validation accuracy of the MLR-models, as for the DEMs of scenarios 'a' (complete version) and 'b' (complete and reduced versions) a contrary result of the DEM quality that increases with the finer resolution. Furthermore, the 5-m spatial resolution DEMs resulted in the MLR-models with smallest and largest validation accuracy, while the 30-m spatial resolution DEMs resulted in MLR-models of similar validation accuracy and very close to the most accurate MLR-model calibrated with DEM5c. This is in agreement with Cavazzi et al. (2013), Penížek et al. (2016), Samuel-Rosa et al. (2015), and Thompson, Bell and Butler (2001). According to these studies, the accuracy of DEMs is not directly related with the accuracy of digital soil maps. That is, depending on the terrain, level of detail of soil survey, strategy and density of sampling; the more generalized information can produce more accurate digital soil maps.

Reducing the number of predictor variables maintained or increased validation accuracy of MLR-models. Exception for baseline MLR-model that had a large reduction in number of predictor variables, from $p = 7$ to $p = 3$. For example, the complete MLR-model calibrated with the DEM5b was less accurate than baseline MLR-model; but in the reduced version, the former was ~2 pp more accurate than the latter, although the improvement was only ~1 pp. For the poorer DEMs, the accuracy of MLR-models increased by ~3-4 pp with selection of predictor variables; while among the superior DEMs the accuracy increased only for DEM5a (~6 pp).

Table 6: Overall accuracy of calibration and validation of multinomial logistic regression (MLR) models calibrated using two sets of predictor variables derived from satellite images (baseline) and digital elevation models (all others).

MLR-model	Complete (p = 20)		Reduced (p = 13)	
	Calibration	Validation	Calibration	Validation
Baseline (T)	44.6	37.6	41.4	34.4
T + DEM5a	72.0	38.2	73.2	44.0
T + DEM5b	64.3	35.0	64.3	36.9
T + DEM5c	72.0	42.0	71.3	47.1
T + DEM20a	73.9	38.9	61.8	38.2
T + DEM20b	71.3	38.9	69.4	38.9
T + DEM20c	67.3	39.1	64.7	43.0
T + DEM30a	70.1	42.7	66.2	42.0
T + DEM30b	70.7	40.8	66.2	44.0
T + DEM30c	72.9	38.7	68.4	43.2

Baseline model was adjusted using p = 7 and p = 3 predictor variables in their complete and reduced versions, respectively. Model(s) with best performance in each spatial resolution is (are) in boldface.

Because of this difference in prediction gains, the reduced MLR-models calibrated with DEM30b and DEM5a ended up with same validation accuracy, and the complete MLR-model calibrated with DEM30b had larger validation accuracy than that obtained with DEM5a. In both cases, the difference to the baseline MLR-model and the most accurate MLR-model was +10 pp and -3 pp, respectively.

Spatial predictions showed the gains obtained when DEMs are used to calibrate the MLR-models, regardless of their quality (Figure 5). DEMs generated using stream network associated with hydrologically consistent algorithm tended to underestimate the presence of the group of soil classes influenced by fluvial sediments, near the *Ribeirão Cachimbal*. All 'c' scenario DEMs predicted
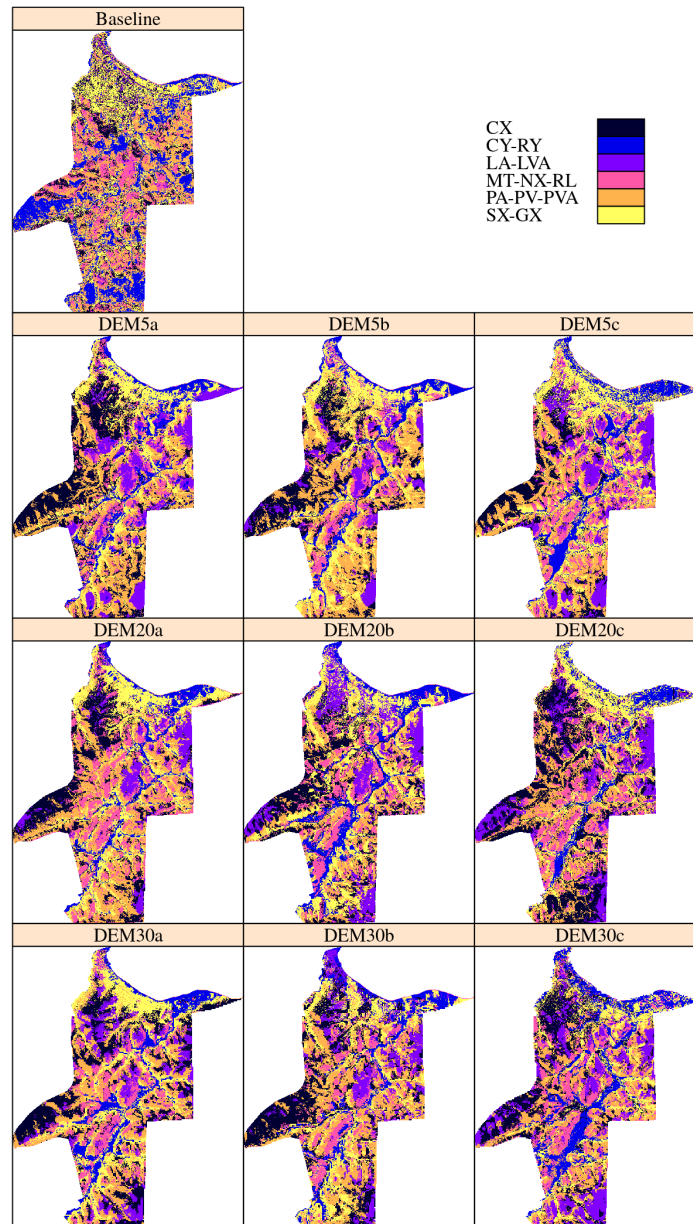

Figure 5: Spatial predictions of k = 6 groups of soil classes using multinomial logistic regression models calibrated using satellite images (baseline, p = 3) and digital elevation models (p = 13) with different levels quality – given the definition of quality used in this study.

large areas of these soils. For other groups, there seemed to be less concordance between predictions using different DEMs. In general, the nine MLR-models captured the same general pattern of spatial distribution of soil classes groups, with a dominance of CX, PA/PV/PVA and SX/GX.

Evaluation of the uncertainty of spatial predictions reinforced the importance of using DEM data to improve the MLR-models performance (Figure 6). Overall, predictions of the groups of soil classes in the plain relief surfaces, closer to the stream network and with

Figure 6: Shannon entropy (uncertainty) of predictions of multinomial logistic regression models calibrated using satellite images (baseline, p = 3) and digital elevation models (p = 13) with different levels quality (as defined in this study).

lower elevation, tended to show lower entropy. For other soil classes groups, the higher entropy reflect a greater difficulty in their spatial prediction. In general, there was little difference in the spatial pattern of uncertainty of different MLR-models.

Relation between DEM quality and soil mapping accuracy

The MLR-models most suitable for use in the study area were those calibrated using the less detailed scenario DEM5c and DEM30c. The DEM quality criteria did not lead to selection of the most suitable DEM for DSM; that is, the DEM capable of producing the most accurate digital soil map. In fact, a poorer DEM produced a more accurate map than a superior DEM.

Many criteria used to infer DEM quality for DSM were originally adopted from projects whose objective was the evaluation of DEMs for modeling the water flow on soil surface (Wise, 2000). Thus, several criteria have a dominant hydrological nature such as analyses of the presence of spurious depressions and delineation of the stream network. Other criteria appear to have limited physical basis, possibly having more of a heuristic nature, such as the descriptive statistics of elevation and slope. In addition, interpretation of some criteria at some points are subjective.

If the quality criteria were not directly linked to soil genesis and spatial distribution, it would be reasonable to suppose that a DEM that accurately represents present terrain surface is not necessarily the one that will produce most accurate digital soil map. One reason is that current soil features are the outcome more of preterit terrain surface than of present. The operational inability to represent preterit terrain surface could be due to the fact that different groups of soils - and soil attributes - often have larger correlation with predictor variables in one or more spatial resolutions (analysis scales) (Behrens et al., 2010). This empirical observation encouraged usage of multiscale soil prediction models, where DEMs with different spatial resolutions or terrain attributes derived using sampling windows of different sizes are used jointly (Behrens et al., 2014; Miller et al., 2015; Samuel-Rosa et al., 2015).

The success of multiscale soil spatial modelling has been attributed to differential action of soil formation processes (Behrens et al., 2010; Behrens et al., 2014). Although not tested in this study, the results suggest that density and spatial configuration of the calibration observations play a considerable role. For example, the accurate delineation of current stream network in some DEMs was detrimental to soil predictions. Such DEMs were constructed using the stream network data, commonly recommended for building hydrologically consistent DEMs for soil mapping e.g. for scenario 'a'. Here, hydrologically consistent DEMs have narrower river plains due to deepening of local drainage channels. The observations of the soil classes group with fluvial features are all located very close to the watercourses. The hydrologically consistent DEMs predicted that the area occupied by this group is smaller than it should be. Thus, depending on the density and location of their observation, it would have been more effective to have coarser information about areas influenced by a given watercourse - subjected to temporary or periodic flooding - than to know their exact geographic location. For Cavazzi et al. (2013) very detailed terrain attributes can generate too much "noise", inevitably harming the accuracy of the prediction. For similar sampling densities, the same idea seems to apply to satellite images, where it is more important to know general effect of a forest on soil spatial variation than that of individual forest trees (Samuel-Rosa et al., 2015). More than pedogenetic meaning, the results point to needing a balance of volume/density of soil data and spatial detail of predictor variables. Hydrologically consistent DEMs with finer spatial resolution may require larger sampling density for predictive gains to be significant; as shown by Baltensweiler et al. (2017), evaluating different DEM resolutions for predicting soil pH in an area with a very high sampling density (31 points per 1 ha).

To select a DEM it is necessary to consider working scale, objectives and resources (financial, human and infrastructure) available for the project. In many circumstances there will be at least two DEMs that meet the project objectives and are compatible with working scale and available resources; thus a second selection criterion should be used. This last and decisive criterion should be the accuracy of soil spatial predictions, such as it has been done here. The DEM that results in greater prediction accuracy, preferably measured using an external probabilistically selected sample, is the most appropriate for that project or, alternatively, the predictions from different models can simply be aggregated.

Considering the topographic variation, moderately rugged, more detailed DEMs produced discontinuous areas, reflecting the details of instant relief features and increasing the uncertainty about the predictions, while the coarser DEM information ignored these details reflecting better the soil cover development. In general, the CX occur in the highest part of the study area and/or flat summit. The SX-GX in the low elevations in flat areas the. The PA-PV-PVA in the shoulder, backslope and footslope. LA-LVA and MT-NX-RL have a strong relationship with the parent material, not considered here, so why these classes were not very well predicted.

CONCLUSIONS

This study showed that: The joint interpretation of DEM quality criteria are useful to identify DEMs with greater or lesser capacity to represent the terrain surface. DEMs with greater capacity to represent the terrain surface not necessarily are the DEMs that produce the most accurate spatial soil class predictions. For small sample sizes, selection of predictor variables can contribute more to improve the accuracy of digital soil maps than using a more accurate and finer spatial resolution DEM.

ACKNOWLEDGEMENTS

REFERENCES

AGRESTI, A. **Categorical data analysis**. New York John Wiley Sons. Gainesville, 2002. 710p.

ALVARES, C. A. et al. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, 22(6):711-728, 2014.

ANTUNES, M. A. H.; DEBIASI, P.; SIQUEIRA, J. C. DOS S. Avaliação espectral e geométrica das imagens Rapideye e seu potencial para o mapeamento e monitoramento agrícola e ambiental. **Revista Brasileira de Cartografia**, 66(1):105-113, 2014.

BALTENSWEILER, A. et al. Terrestrial laser scanning improves digital elevation models and topsoil pH modelling in regions with complex topography and dense vegetation. **Environmental Modelling and Software**, 95(1):13-21, 2017.

BEHRENS, T. et al. Multi-scale digital terrain analysis and feature selection for digital soil mapping. **Geoderma**, 155(3-4):175-185, 2010.

BEHRENS, T. et al. Hyper-scale digital soil mapping and soil formation analysis. **Geoderma**, 213:578-588, 2014.

CAVAZZI, S. et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? **Geoderma**, 195(196):111-121, 2013.

CHAGAS, C. S. et al. Avaliação de modelos digitais de elevação para aplicação em um mapeamento digital de solos. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 14(21):218-226, 2010.

FAN., L.; ATKINSON, P. M. Accuracy of digital elevation models derived from terrestrial laser scanning Data. **Geoscience and Remote Sensing Letters, IEEE**, 12(9):1923-1927, 2015.

FISHER, P. F.; TATE, N. J. Causes and consequences of error in digital elevation models. **Progress in Physical Geography**, 30(4):467-489, 2006.

FLORINSKY, I. V. Accuracy of local topographic variables derived from digital elevation models. **International Journal of Geographical Information Science**, 12(1):47-61, 1998.

HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PLOS ONE**, 12(2):1-40, 2017.

HENGL, T.; GRUBER, S.; SHESTHA, D. P. Reduction of errors in digital terrain parameters used in soil-landscape modelling. **International Journal of Applied Earth Observation and Geoinformation**, 5:97-112 , 2004.

HEUNG, B. et al. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. **Geoderma**, 265:62-77, 2016.

HIRT, C.; FILMER, M. S.; FEATHERSTONE, W. E. Comparison and validation of the recent freely available ASTER-GDEM ver1, SRTM ver4.1 and GEODATA DEM-9S ver3 digital elevation models over Australia. **Australian Journal of Earth Sciences**, 57(3):337-347, 2010.

HÖHLE, J.; HÖHLE, M. Accuracy assessment of digital elevation models by means of robust statistical methods. **ISPRS Journal of Photogrammetry and Remote Sensing**, 64(4): 398-406, 2009.

HUTCHINSON, M. F. A new procedure for gridding elevation and streamline data with automatic removal of spurious pits. **Journal of Hydrology**, 106:211-232, 1989.

MAYNARD, J. J.; JOHNSON, M. G. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs neighborhood extent. **Geoderma**, 230(231):29-40, 2014.

MILLER, B. A. et al. Impact of multi-scale predictor selection for modeling soil properties. **Geoderma**, 239(240):97-106, 2015.

MOURA-BUENO, J. M. et al. Assessment of digital elevation model for digital soil mapping in a watershed with gently undulating topography. **Revista Brasileira de Ciência do Solo**, 40:1-15, 2016.

NELSON, A.; REUTER, H. I.; GESSLER, P. DEM production methods and sources. **Developments in Soil Science**, 33:65-85, 2009.

NEUMANN, M. R. B.; ROIG, H. L.; SOUZA, A. L. F. DE. Digital elevation models obtained by contour lines and SRTM/Topodata, for digital soil mapping. **Journal of Soil Science and Environmental Management**, 3(5):104-109, 2012.

PENÍŽEK, V. et al. Influence of elevation data resolution on spatial prediction of colluvial soils in a Luvisol region. **PLOS ONE**, 11:1-18, 2016.

PINHEIRO, H. S. K. et al. Modelos de elevação para obtenção de atributos topográficos utilizados em mapeamento digital de solos. **Pesquisa Agropecuária Brasileira**, 47(1)1384-1394, 2012.

PORTILHO, A. P. et al. Balanço hídrico para Pinheiral, Rio de Janeiro. **Revista Agroambiental**, 4:21-28, 2011.

SAMUEL-ROSA, A. et al. Do more detailed environmental covariates deliver more accurate soil maps? **Geoderma**, 243(244):214-227, 2015.

SANSON, M. DE S. R.; RAMOS, R. R. C.; MELLO, C. L. Bacia sedimentares brasileiras: Bacia de Volta Redonda. **Fundação Paleontológica Phoenix**, 88:1-6, 2006.

SANTOS, R. D. et al. **Manual de descrição e coleta de solo no campo**. 7°ed. Revisada e ampliada. SBCS. Viçosa. 2015. 100p.

TEN CATEN, A. et al. Mapeamento digital de classes de solos: Características da abordagem brasileira. **Ciência Rural**, 42:1989-1997, 2012.

THOMPSON, J. A.; BELL, J. C.; BUTLER, C. A. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. **Geoderma**, 100(1-2):67-89, 2001.

VALERIANO, M. DE M.; ROSSETTI, D. DE F. Topodata: Brazilian full coverage refinement of SRTM data. **Applied Geography**, 32(2):300-309, 2012.

VASQUES, G. M.; GRUNWALD, S.; MYERS, D. B. Associations between soil carbon and ecological landscape variables at escalating spatial scales in Florida, USA. **Landscape Ecology**, 27(3):355-367, 2012.

WISE, S. Assessing the quality for hydrological applications of digital elevation models derived from contours. **Hydrological processes**, 14:1909-1929, 2000.