Digital soil mapping and its implications in the extrapolation of soil-landscape relationships in detailed scale

Mario Sergio Wolski[1], Ricardo Simão Diniz Dalmolin[2], Carlos Alberto Flores[3], Jean Michel Moura-Bueno[2], Alexandre ten Caten[4] and Douglas Rodrigo Kaiser[1]

[1]Universidade Federal da Fronteira Sul, Curso de Agronomia, Rua Jacob Reinaldo Haupenthal, nº 1.580, CEP 97900-000 Cerro Largo, RS, Brazil. E-mail: mariowolski@hotmail.com, douglasrodrigokaiser@gmail.com [2]Universidade Federal de Santa Maria, Centro de Ciências Rurais, Departamento de Solos, Avenida Roraima, nº 1.000, Cidade Universitária, Camobi, CEP 97105-900 Santa Maria, RS, Brazil. E-mail: dalmolin@ufsm.br, bueno.jean1@gmail.com [3]Embrapa Clima Temperado, Caixa Postal 403, CEP 96001-970 Pelotas, RS, Brazil. E-mail: carlos.flores@embrapa.br [4]Universidade Federal de Santa Catarina, Campus de Curitibanos, Rodovia Ulysses Gaboardi, Km 3, CEP 89520-000 Curitibanos, SC, Brazil. E-mail: ten.caten@ufsc.br

Abstract – The objective of this work was to test the extrapolation of soil-landscape relationships in a reference area (RA) to a topographic map (scale 1:50,000), using digital soil mapping (DSM), and to compare these results to those obtained in similar studies previously conducted in Brazil. A soil survey in a 10 km² RA, using conventional mapping techniques (scale 1:10,000), was made in order to map a 678 km² physiographically similar area (scale 1:50,000) using DSM. The decision tree technique was employed to build a predictive extrapolation model based on soil classes and eight terrain attributes in the RA. The validation of DSM by application of field observation points resulted in a 66.1% global accuracy and in 0.36 kappa index. The most representative soils in the area were correctly predicted, whereas the less representative and less frequent soils in the landscape (and consequently with reduced sampling) had their prediction compromised. The RA proportion, which equals 1.5% of the total area, is a limiting factor in the formulation of soil-landscape relationships to precisely represent the mapped area by DSM.

Index terms: landscape stratification, pedometrics, predictive soil mapping, reference area, terrain attributes.

Mapeamento digital do solo e suas implicações na extrapolação das relações solo-paisagem em escala de detalhe

Resumo – O objetivo deste trabalho foi testar a extrapolação das relações solo-paisagem de uma área de referência (AR), por meio de mapeamento digital de solos (MDS), para uma carta topográfica (1:50.000), e comparar os resultados aos obtidos em estudos similares anteriormente desenvolvidos no Brasil. O trabalho consistiu no levantamento de solos, com técnicas convencionais de mapeamento de uma AR de 10 km² (na escala 1:10.000), para mapear uma área fisiograficamente similar de 678 km² (na escala 1:50.000), tendo-se utilizado o MDS. A técnica de árvore de decisão (AD) foi utilizada para a construção do modelo preditivo de extrapolação, com base nas classes de solos e em oito atributos de terreno da AR. A validação do MDS, com pontos de observação de campo, resultou em 66,1% de exatidão global e 0,36 de índice kappa. Os solos mais representativos da área foram preditos corretamente, enquanto solos menos representativos e de menor ocorrência na paisagem e, consequentemente, com amostragem reduzida, tiveram sua predição comprometida. A proporção da AR, igual a 1,5% da área total, é um fator limitante à formulação das relações solo-paisagem para representar precisamente a área mapeada por MDS.

Termos para indexação: estratificação da paisagem, pedometria, mapeamento preditivo de solos, área de referência, atributos de terreno.

Introduction

In Brazil, among the studies using digital soil mapping (DSM) stand out those based on legacy data. According to Giasson et al. (2013), among the legacy data are maps and reports generated without proper field validation. As there are no estimates on the uncertainty involved in these soil inventories, any eventual errors may be propagated further. Additionally, most of the soil legacy data are presented in scales that are not compatible to support DSM at detailed scales (between 1:5,000 and 1:20,000). Therefore, these data do not supply the demand of soil information on agricultural planning

for watersheds and rural properties, according Dalmolin et al. (2004).

The DSM approach employs basic concepts of pedometrics by using soil spatial prediction functions (SSPF). Based on the assumptions derived from the soil-landscape relationships, these functions constitute a frame for empirical fitting among quantitative correlations between the soil and the environment (McBratney et al., 2003). This approach has been frequently used in order to predict soil classes and/ or properties in nonsampled locations at the regional, national, and global levels (ten Caten et al., 2012; Hengl et al., 2014; Dalmolin & ten Caten, 2015; Teske et al., 2015).

In an area having a small geographical extension into the same physiographic region, soil formation factors such as parent material, organisms, and climate may often show small spatial variations. In these locations, relief is the main factor involved in outlining soil types, and is responsible for most of the variation found in these landscapes. In such regions, small representative areas, previously mapped, may be designated as reference areas (RA), as suggested by Lagacherie et al. (1995). The RAs are used as a basis for extrapolation of the soil-landscape relationships, which enables mapping of the neighboring areas with similar characteristics.

The RA approach for DSM was applied at municipality scale in Rio Grande do Sul state, Brazil by ten Caten et al. (2011), who found that Lithic Leptosol showed the highest value (above 98%) of mapping accuracy (MA). According to the authors, this observation is associated with the fact that nine predictive covariates were used to generate the models derived from the relief that, in the evaluated area, shows a strong correlation with Lithic Leptosols. Bagatini et al. (2016) employed DSM on two watersheds in Rio Grande do Sul state, by extrapolating data from previously mapped soil found in the RA. In their model, a higher accuracy was observed in the model training areas, in comparison with the validation areas, and that the representativeness of soil classes affected the accuracy of the prediction models. Despite the intensification of DSM research in Brazil, studies based on RA are still scarce, and most of them employ legacy data for extrapolation of soil-landscape relationships to map neighboring regions. In general, legacy data does not show compatible scales for conducting more detailed surveys. In areas for which no soil maps are available, such as Missões region in Rio Grande do Sul state, it is necessary to develop a methodology to acquire these data, on a scale compatible with the user requirements. The use of RA, with a greater degree of detail than the area of interest, was the strategy used in the present study.

The objective of this work was to test the extrapolation of soil-landscape relationships in a reference area (RA) to a topographic map (scale 1:50,000), using digital soil mapping (DSM), and to compare these results to those obtained in similar studies previously conducted in Brazil.

Materials and Methods

The study area is depicted in the topographic chart of Carajazinho (SH.21-X-B-VI-1), scale 1:50,000, and is located in the physiographic region of Missões, in the state of Rio Grande do Sul, Brazil, occupying an approximate area of 678 km^2 (Figure 1 A). This area was chosen as it appropriately represents a region of the state which is a large producer of grains. However, there is not detailed soil information on this region. The climate of the region is a humid subtropical climate (Cfa, according to the classification of Köppen-Geiger), with rainfall uniformly distributed throughout the year, average annual precipitation of 1,771 mm, and average annual temperature of 18.5°C (Alvares et al., 2013). In general, the area has a flat to gently undulated terrain, with steep slopes near the drains, with occasional hills with slopes from 3 to 10%. The geological substrate consists of basalt of the Formação Serra Geral. After exploring the entire region of study, an area of 10 km^2 [1.5% of the total area (TA)] was chosen for the detailed soil survey and designated as the RA (Figure 1 B). A topographical survey was conducted in the RA, in order to generate a digital elevation model (DEM-RA) which was used for understanding the soil-landscape relationships, and for conventional soil mapping on a 1:10,000 scale. Twenty complete soil profiles were described and sampled using the method described by Santos et al. (2015), and 280 observation points were surveyed by auger sampling systems. Physical and chemical analyses of the soil profiles were performed according to Donagema et al. (2011). Soils were classified up to the series level, based on the Brazilian Soil Classification System (SiBCS) (Santos et al., 2013).

In this survey, we used the methodology described by the Brazilian Institute of Geography and Statistics (IBGE, 2015), and identified 23 mapping units (MUs) (Figure 1 C). In order to extrapolate the soil-landscape relationships obtained by DSM to the entire chart area, the MUs of the RA were grouped under the second categorical level of the SiBCS, generating a map with a simplified legend and five MUs (Figure 1 D), out of which 2,500 points comprising a density of 0.4 points

per hectare were selected by random sampling for prediction model training. On the basis of the DEM-RA representing the relief, eight covariates were extracted to predict soil distribution in the landscape, which included elevation (ELEV), topographic wetness index (TWI), profile curvature (PROF), plan curvature (PLA), aspect (ASP) and slope (SLOP), vertical distance from channel network (VDCN), and geoform classes. Extraction of these covariates
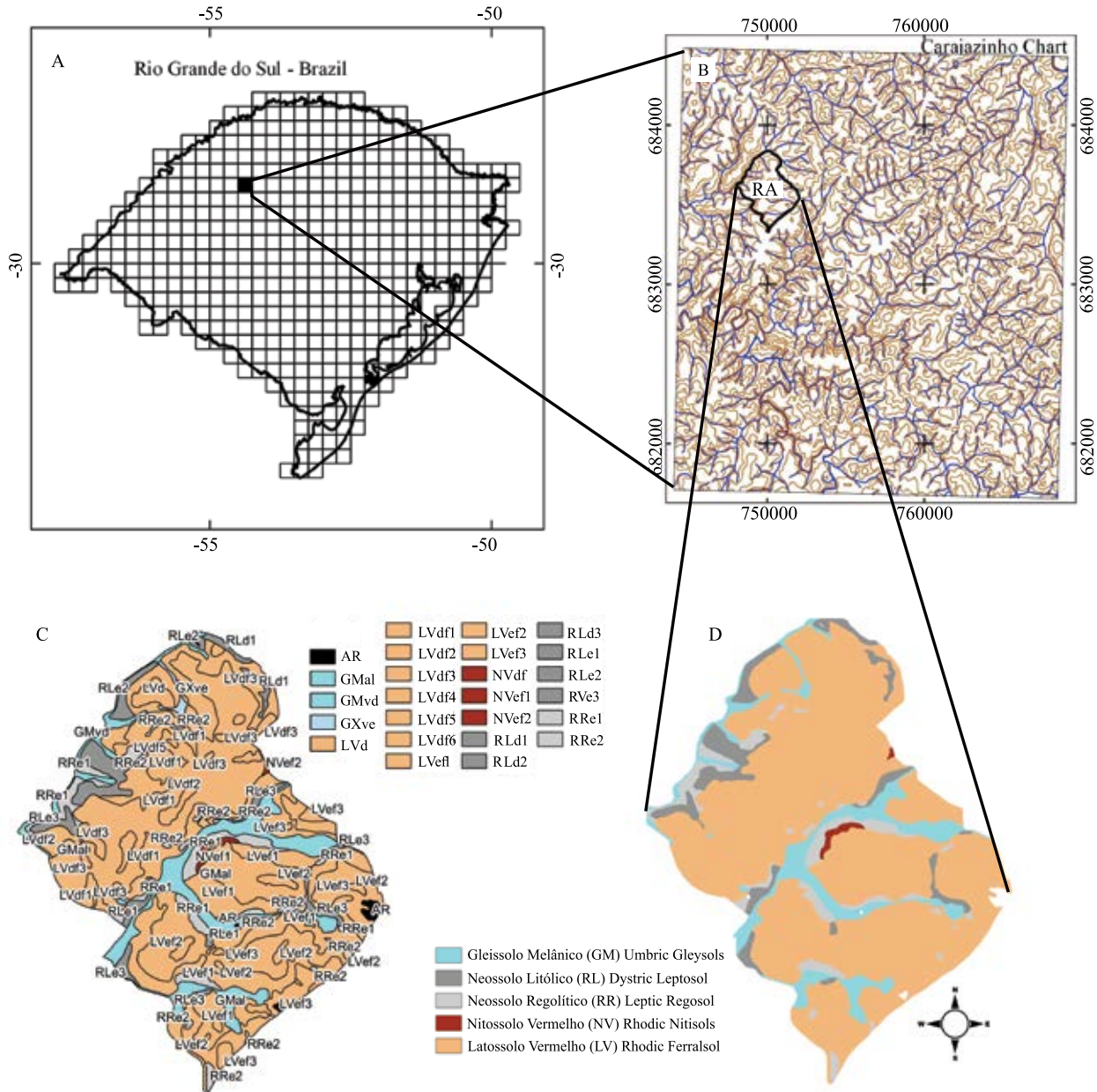

Figure 1. Chart of Carajazinho, RS, Brazil (A); reference area (RA) location in the Carajazinho chart (B); conventional soil map at 1:10,000 scale (C); and a map of the RA with a simplified legend (D). Datum: Sirgas2000, UTM projection, Zone 21S (Sirgas…, 2017).

was performed using the geographical information system software SAGA GIS, version 2.2.3 (SAGA User Group Association, University of Hamburg, Germany), and the geoform classes were obtained using the LandMappeR (MacMillan, 2003). For each of the 2,500 points in the RA, the predictive covariate values and the soil classes up to the second SiBCS categorical level were obtained. These data were exported to the software WEKA, version 3.6.3 (Hall et al., 2009), before proceeding to model training based on the decision tree (DT) methodology, which was executed using the algorithm J48, in order to establish the soil-landscape relationships in the RA.

The DEM for the entire area of study (DEM-chart) was generated at a resolution of 30 m by using contour interpolation, and vectorized and structured following the method described by Hasenack & Weber (2010). The same eight predictive covariates above mentioned for the RA were extracted from the DEM-chart, and an interpolation of the trained model using DT on the RA was performed using a SIG environment interface (SAGA GIS 2.2.3). Subsequently, a digital soil map was generated for the entire area (DSM-chart).

The DSM-chart was validated using 595 points, sampled in the field through observation of profiles and random drilling, along with soil identification up to the second SiBCS categorical level. An error matrix was generated (Congalton, 1991), and the following accuracy indicators were calculated: global accuracy (GA), producer accuracy (PA), user accuracy (UA), and kappa index (K). The GA specifies the proportion of accurate predictions in relation to the total number of pixels; PA expresses how much of the MU surface obtained using the conventional survey was mapped using DSM as well; UA indicates the probability of a DSM MU (represented by a pixel) coinciding with the MU of a conventional survey; K represents an association measure used to describe the concordance level in the MU prediction.

Moreover, in order to evaluate the extrapolation capacity of the soil-landscape relationships by the prediction model, a DSM data clipping for the RA boundaries was performed, and the DSM was validated using 149 field observation points randomly distributed in the RA. Further, an error matrix was generated to assess the evaluation accuracy using the indicators GA, PA, UA, and K.

The obtained results were compared with the findings of studies previously conducted in Brazil by using RA for DSM (ten Caten et al., 2011; Arruda et al., 2013, 2016; Villela, 2013; Silva et al., 2016). Scientific articles were searched using the Scielo, Scopus, and Web of Science platforms, based on a study by ten Caten et al. (2012). The main criterion for the data to be included in the table was that the RAs should represent up to 20% of the total surveyed area.

Results and Discussion

The DSM generated by extrapolating soil-landscape relationships of the RA to the Carajazinho chart (Figure 2) predicted five MUs: Gleissolo Melânico (GM) - Umbric Gleysols; Neossolo Regolítico (RR) - Leptic Regosol; Neossolo Litólico (RL) - Dystric Leptosol; Nitossolo Vermelho (NV) - Rhodic Nitisols; and Latossolo Vemelho (LV) - Rhodic Ferralsol. The highest representation is showed by the LV, which comprises 72% of the total area; this soil occurrence is predicted in areas with flat to gently undulated terrain and in geoforms having smooth top, rectilinear slopes, and platforms/plateaus.

The areas classified as GM had the second highest representation, comprising 17.37% of the total, and were predicted in regions near the drainage network. The RL and RR areas represented 11.31% of the total, predicted in landscapes with divergent slopes (RL) and divergent shoulders/convergent slopes (RR). The NV area had the lowest representation, accounting for only 0.11% of the total, which was predicted in the same locations as RR. The occurrence of RR, RL, and NV in similar locations often hindered the establishment of soil-landscape relationships. Arruda et al. (2016), also working with DSM, observed that similar behavior of environmental covariates complicated the distinction between soil classes.

Soils in the NV class were the less representatives both in the RA and in Carajazinho map, thus, they showed a lower number of samples in the model training, which resulted in a relatively small area predicted as NV. This fact agrees with Hengl et al. (2007), who showed that predictive mapping should be performed with a minimal representation from each soil class, in order to allow that the models capture the complexity of soil formation in the area to be mapped, ensuring that soils of all classes be represented.
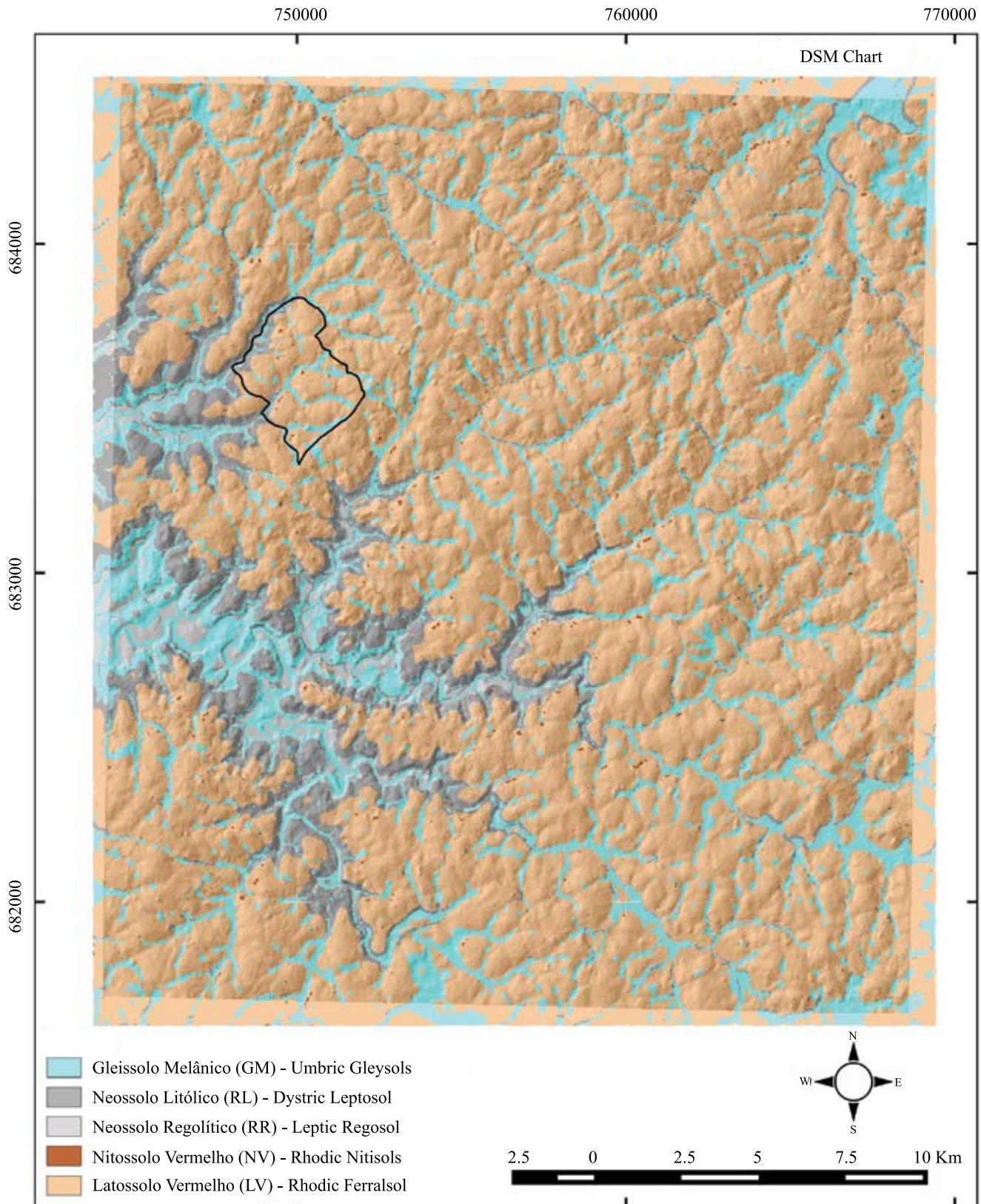
Figure 2. Digital soil map of the Carajazinho chart at 1:50,000 scale. Datum: Sirgas 2000, UTM projection, Zone 21S (Sirgas…, 2017).

In the error matrix for DSM (Table 1), UA values were as follows: LV showed the highest one, 77.5%, RL and GM had intermediate values, 40.0 and 37.4%, respectively; and RR, the lowest one, 25.0%. The percentages represent proportion of the area that was accurately mapped. The class NV recorded no value in the matrix, indicating that this class was not predicted by the model. Values for PA followed the same trend, as follows: the highest value for LV, 90.6%; intermediate values for MG (52.3%) and RL (34.8%); and, the lowest values for RR (4.1%) and NV (0.0%).

The soil classes RR and NV had low representation in the area of study, and thus, reduced sample number, which explains their low UA and PA values, highlighting the NV class. ten Caten et al. (2011), also observed this fact and showed that prediction for soils from less represented classes was compromised when performing DSM based on multiple logistic regression analyses. In their study, the topography was heterogeneous, varying from plain to mountainous. Since there is no defined protocol for DSM (Grunwald, 2009), we hypothesized that in an area with more homogeneous relief, the limitation described by ten Caten et al. (2011) could not occur in our area of study. Furthermore, soils from these classes occur contiguously in the entire region, predominantly in the geoforms with divergent shoulders and convergent slopes, which hinders the differentiation by our predictive model, which is entirely based on terrain attributes. Bagatini et al. (2016) extrapolated preexisting soil maps to physiografically similar areas using DSM in two distinct watersheds. Their models overestimated the MUs with higher representation and underestimated those less represented, both in the training and in the validation areas. Besides, the use of DT methodology in the DSM showed low efficiency for the extrapolation of preexisting soil maps. Their predictive model reproduced GA values of 63 and 68% in the training areas in both watersheds. On extrapolation, these GA values dropped to 50 and 54%.

In the generated DSM, GA was 66.1%, which indicates the proportion of accurately predicted observations (agreement between predicted soil classes and reference classes) in relation to the total number of observations. The K value was 0.36, which is lower than the average value of 0.47 observed in similar studies conducted in Brazil (ten Caten et al., 2012). However, Giasson et al. (2013) observed an average K of 0.37 in the Lajeado Grande watershed, which shows characteristics similar to the area of the present study. Likewise, Bagatini et al. (2015) obtained a GA similar to those in the present study; in their study, depending on the sampling point density, GA varied between 60 and 76% in the Rio Santo Cristo watershed, and between 54% (lowest density, 0.1 points ha^{-1}) and 74% (highest density, 4 points ha^{-1}) in the Rio Lajeado Grande watershed. This GA behavior was similar to that observed in the Carajazinho chart, which belongs to the same physiographic region evaluated by Giasson et al. (2013) and Bagatini et al. (2015).

High UA and PA values for LV and MG, especially those for the LV obtained in this study (AU of 77.5%, and PA of 90.6%), are associated with the high proportion of soils belonging to these classes in the study area, to a higher number of training samples, and to their positions in the landscape.

Table 1. Error matrix and evaluation parameters for DSM accuracy in the Carajazinho chart.

| | Classes | Validation points | | | | | ∑ rows | UA(%) |
		GM	RL	RR	NV	LV		
Predicted	GM	34	15	21	1	20	91	37.4
	RL	12	24	11	2	11	60	40.0
	RR	1	6	3	0	2	12	25.0
	NV	0	0	0	0	0	0	-
	LV	18	24	38	13	320	413	77.5
	∑ columns	65	69	73	16	353	576[*]	-
	PA (%)	52.3	34.8	4.1	0.0	90.6	-	-
							GA 66.1% K = 0.36	

GM, Gleissolo Melânico (Umbric Gleysols); RL, Neossolo Litólico (Dystric Leptosol); RR, Neossolo Regolítico (Leptic Regosol); NV, Nitossolo Vermelho (Rhodic Nitisols); LV, Latossolo Vermelho (Rhodic Ferralsol); UA, user accuracy; PA, producer accuracy; GA, global accuracy; K, kappa index. [*]Total number of field validation points was 595, out of which 19 points were classified as rocky outcrop.

The size of the RA (only 1.5% of the TA) also affects the quality of prediction. This factor could be verified by comparing the single reliability measurements with the prediction data. Due to its small representation in the RA, the NV limited the number of pixels predicted in MUs during the algorithm training, compromising the establishment of relationships between the environmental covariates.

The DSM results obtained for the Carajazinho chart allowed to evidence that the determination of spatial position and collection of data in the RA are the two most important steps in the DSM technique, since the RA should comprise all MUs depicting territorial expression, in order to be considered representative of the landscape to be mapped. However, there are no reports recommending a minimal RA size in relation to the TA. Nevertheless, Höfig et al. (2014) reported a reduced ability of the

predictive models to reproduce MUs with small representation in the evaluated areas.

In the comparison between the conventional RA map (Figure 3 A) and the snipped DSM-generated RA (Figure 3 B), a visual similarity in the spatialization of soil MUs was observed. The error matrix (Table 2) shows 72.92% GA and 0.54 K of, both higher than the values found for the entire Carajazinho chart area. UA and PA values for LV and MG were still higher, especially for LV (UA 80.4%, and PA 92.5%). UA and PA values for NV and RR also increased, probably due to the reduced distance of the extrapolation of soil-landscape relationships to the area of interest. Therefore, it is important to investigate the f environmental covariates of their ability to predict and classify soils located at different distances from the RA. The main studies with RA in Brazil present different methodologies and do not suggest a minimum
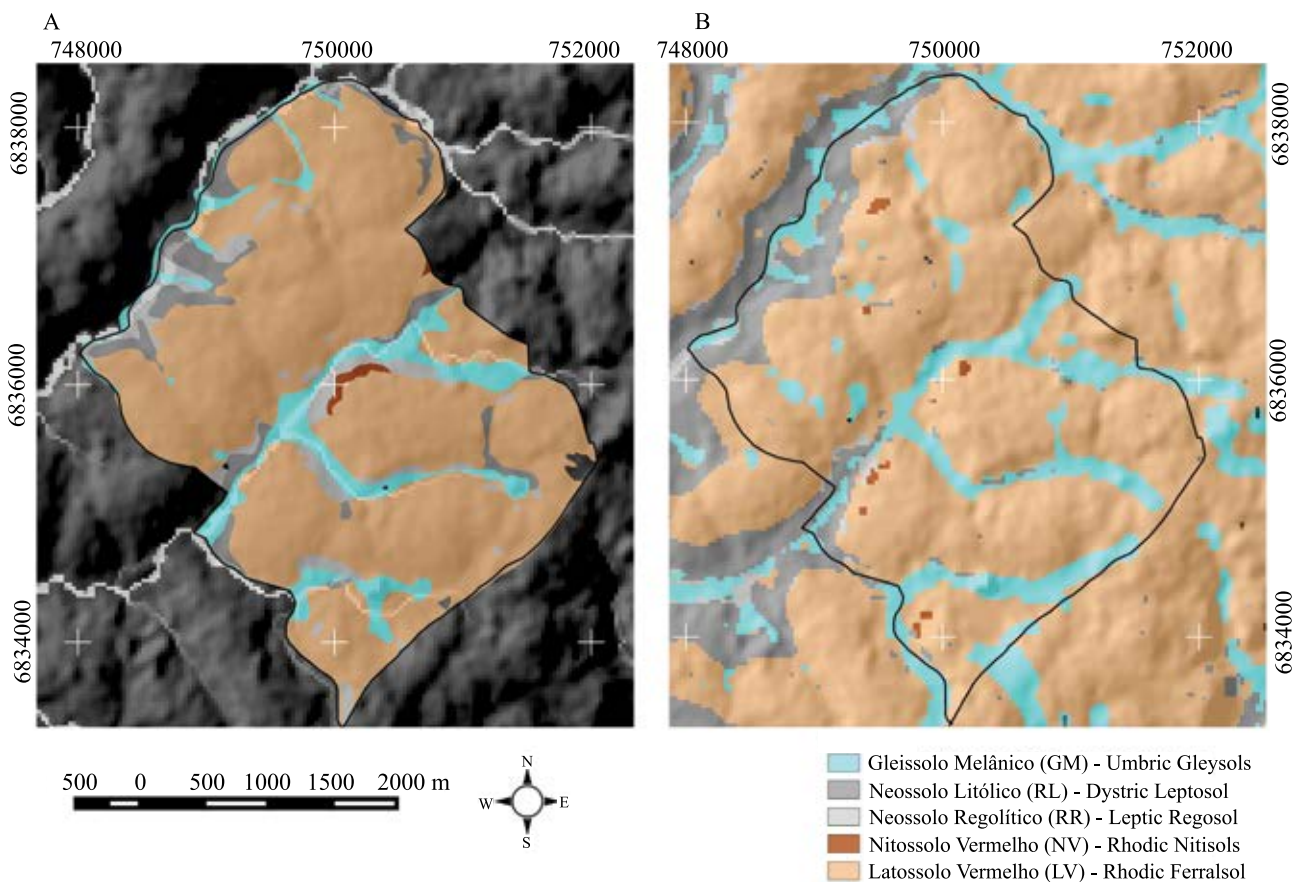

Gleissolo Melânico (GM) - Umbric Gleysols
Neossolo Litólico (RL) - Dystric Leptosol
Neossolo Regolítico (RR) - Leptic Regosol
Nitossolo Vermelho (NV) - Rhodic Nitisols
Latossolo Vermelho (LV) - Rhodic Ferralsol

Figure 3. Soil maps obtained by the conventional method (A), and the reference area (RA) section in the digital soil mapping (DSM) of the Carajazinho chart at 1:50,000 scale (B). Datum: Sirgas 2000, UTM projection, Zone 21S (Sirgas…, 2017).

size of RA to be used. There is no correlation between RA size, in relation to TA of extrapolation, and the DSM accuracy (Table 3).

The study 1 (Table 3) presents the work developed by Silva et al. (2016), who obtained high-accuracy values (70.97% GA, and 0.55 K) using an RA/TA of only 0.27% for extrapolation of different soil classes in an area with undulated terrain. The study 2 (Arruda et al., 2016) presents an RA of 500 ha located in the center of the area of study (scale 1:10,000), which was planned with the aim at building a DSM based on artificial neural networks (ANN), using environmental covariates that expressed soil-landscape relationships for an area of 110.72 km^2 in an undulated terrain, generating high values of GA (83.7%) and K (0.80).

Study 3 presents the results of this work, where a low-K value was obtained, in comparison with the other values shown in Table 3, due to the small RA size (1.47% of the extrapolation), contributing to the low accuracy of our DSM chart. In the study 4 (Arruda et al., 2013), five RAs (total 12 km²) were selected by grouping, in order to extrapolate soil-landscape

relationships to an area of 120 km², thus, obtaining 77.5% GA and 0.74 K. In the study 5 (Villela, 2013), a conventional detailed soil survey was performed in an area of 80 km², which was used as RA to extend the map using DSM techniques to an area of 730 km² with scale reduction and legend simplification. Four DSM charts were built, from which two were trained by models based on the conceptual model of the pedologist for the area of study, and two were trained by models based on the statistical analyses of information on the RA. The techniques proved efficient for MU prediction in the area of study, with GA varying between 74.62 and 88.81%, and K ranging from 0.68 to 0.85. In the study 6 (ten Caten et al., 2011) used an RA of 143 km² for extrapolation of soil-landscape relationships to an area of 874 km² with heterogeneous terrain attributes, and found 61.79% GA and a 0.46 K.

Therefore, the best results (high-GA and K values) were achieved by Villela (2013), who reported satisfactory quality indices in a plain terrain using an RA corresponding to 11% of the TA, which are values very close to the ones observed by Arruda et al.

Table 2. Error matrix and evaluation parameters for assessing the accuracy of the predicted map using digital soil mapping (DSM) of the reference area (RA).

	Classes	Validation points					∑ rows	UA (%)
		GM	RL	RR	NV	LV		
Predicted	GM	18	4	2	0	2	26	69.23
	RL	1	7	3	0	2	13	53.85
	RR	2	3	5	0	1	11	45.45
	NV	0	0	0	1	1	2	50.00
	LV	5	6	6	1	74	92	80.43
	∑ columns	26	20	16	2	80	144	-
	PA (%)	69.23	35.00	31.25	50.00	92.50	-	-
							GA 72.92%	K = 0.54

GM, Gleissolo Melânico (Umbric Gleysols); RL, Neossolo Litólico (Dystric Leptosol); RR, Neossolo Regolítico (Leptic Regosol); NV, Nitossolo Vermelho (Rhodic Nitisols); LV, Latossolo Vermelho (Rhodic Ferralsol); UA, user accuracy; PA, producer accuracy; GA, global accuracy; K, kappa index.

Table 3. Size correlations between the RA and the extrapolated area, as reported by the studies conducted in Brazil[1].

Study	Reference	Predominant relief	RA (%)	RA (km²)	TA (km²)	GA (%)	kappa index
1	Silva et al. (2016)	Undulated	0.27	4.85	1.771.9	70.97	0.55
2	Arruda et al. (2016)	Undulated	4.5	5	111	83.7	0.81
3	This study	Gently undulated	1.5	10	678	66.1	0.36
4	Arruda et al. (2013)	Undulated	10	12	120	77.5	0.74
5	Villela (2013)	Flat	11	80	730	74.6 to 88.8	0.68 to 0.85
6	ten Caten et al. (2011)	Heterogeneous	16.4	143	874	61.8	0.46

[1]RA, reference area; TA, total area; and GA, global accuracy.

(2013, 2016). It was not possible to infer the covariates responsible for the different results achieved in the mentioned studies (Table 3). Therefore, although the kappa values were considered to be at least satisfactory, it was impossible to conclusively establish an ideal ratio between RA and TA (Table 3).

This shows that further research is required to indicate the ideal size of an RA, considering the different relief configurations, the RA position in the area to be extrapolated, and strategies to obtain the cartographic base and soil map of the RA in regions where soil data are scarce.

The present work and the evaluation of previous studies using RA performed in Brazil reinforce the need for further research to establish protocols for the identification of the ideal RA size.

Conclusions

1. The use of a reference area (RA) in digital soil mapping (DSM) showed that it is possible to acquire soil information in unmapped areas.

2. Soil classes showing a low level of occurrence in the RA cannot be accurately predicted by extrapolation of soil-landscape relationships.

Acknowledgments

References

ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; GONÇALVES, J.L. de M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v.22, p.711-728, 2013. DOI: 10.1127/0941-2948/2013/0507.

ARRUDA, G.P. de; DEMATTÊ, J.A.M.; CHAGAS, C. da S. Mapeamento digital de solos por redes neurais artificiais com base na relação solo-paisagem. **Revista Brasileira de Ciência do Solo**, v.37, p.327-338, 2013. DOI: 10.1590/S0100-06832013000200004.

ARRUDA, G.P. de; DEMATTÊ, J.A.M.; CHAGAS, C. da S.; FIORIO, P.R.; SOUZA, A.B. e; FONGARO, C.T. Digital soil mapping using reference area and artificial neural networks. **Scientia Agricola**, v.73, p.266-273, 2016. DOI: 10.1590/0103-9016-2015-0131.

BAGATINI, T.; GIASSON, E.; TESKE, R. Expansão de mapas pedológicos para áreas fisiograficamente semelhantes por meio de mapeamento digital de solos. **Pesquisa Agropecuária**

Brasileira, v.51, p.1317-1325, 2016. DOI: 10.1590/S0100-204X2016000900031.

BAGATINI, T.; GIASSON, E.; TESKE, R. Seleção de densidade de amostragem com base em dados de áreas já mapeadas para treinamento de modelos de árvores de decisão no mapeamento digital de solos. **Revista Brasileira de Ciência do Solo**, v.39, p.960-967, 2015. DOI: 10.1590/01000683rbcs20140289.

CONGALTON, R.G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v.37, p.35-46, 1991. DOI: 10.1016/0034-4257(91)90048-B.

DALMOLIN, R.S.D.; KLAMT, E.; PEDRON, F. de A.; AZEVEDO, A.C. de. Relação entre as características e o uso das informações de levantamentos de solos de diferentes escalas. **Ciência Rural**, v.34, p.1479-1486, 2004. DOI: 10.1590/S0103-84782004000500023.

DALMOLIN, R.S.D.; ten CATEN, A. Mapeamento digital: nova abordagem em levantamento de solos. **Investigación Agraria**, v.17, p.77-86, 2015. DOI: 10.18004/investig.agrar.2015.diciembre.77-86.

DONAGEMA, G.K.; CAMPOS, D.V.B. de; CALDERANO, S.B.; TEIXEIRA, W.G.; VIANA, J.H.M. (Org.). **Manual de métodos de análise de solo**. 2.ed. rev. Rio de Janeiro: Embrapa Solos, 2011. 230p. (Embrapa Solos. Documentos, 132).

GIASSON, E.; HARTEMINK, A.E.; TORNQUIST, C.G.; TESKE, R.; BAGATINI, T. Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil. **Ciência Rural**, v.43, p.1967-1973, 2013. DOI: 10.1590/S0103-84782013001100008.

GRUNWALD, S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. **Geoderma**, v.152, p.195-207, 2009. DOI: 10.1016/j.geoderma.2009.06.003.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I.H. The WEKA data mining software: an update. **SIGKDD Explorations**, v.11, p.10-18, 2009. DOI: 10.1145/1656274.1656278.

HASENACK, H.; WEBER, E. (Org.). **Base cartográfica vetorial contínua do Rio Grande do Sul**: escala 1:50.000. Porto Alegre: UFRGS, Instituto de Biociências, Centro de Ecologia, 2010. (Série Geoprocessamento, 3).

HENGL, T.; TOOMANIAN, N.; REUTER, H.I.; MALAKOUTI, M.J. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. **Geoderma**, v.140, p.417-427, 2007. DOI: 10.1016/j.geoderma.2007.04.022.

HENGL, T.; JESUS, J.M. de; MACMILLAN, R.A.; BATJES, N.H.; HEUVELINK, G.B.M.; RIBEIRO, E.; SAMUEL-ROSA, A.; KEMPEN, B.; LEENAARS, J.G.B.; WALSH, M.G.; GONZALEZ, M.R. SoilGrids1km - Global soil information based on automated mapping. **PLoS ONE**, v.9, e114788, 2014. DOI: 10.1371/journal.pone.0105992.

HÖFIG, P.; GIASSON, E.; VENDRAME, P.R.S. Mapeamento digital de solos com base na extrapolação de mapas entre áreas fisiograficamente semelhantes. **Pesquisa Agropecuária**

Brasileira, v.49, p.958-966, 2014. DOI: 10.1590/S0100-204X2014001200006.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Manual técnico de pedologia**. 3.ed. Rio de Janeiro, 2015. 297p.

LAGACHERIE, P.; LEGROS, J.P.; BURFOUGH, P.A. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. **Geoderma**, v.65, p.283-301, 1995. DOI: 10.1016/0016-7061(94)00040-H.

MACMILLAN, R.A. **LandMapR Software Toolkit- C++ Version (2003)**: User's manual. Edmonton: LandMapper Environmental Solutions Inc., 2003. 110p.

MCBRATNEY, A.B.; MENDONÇA SANTOS, M.L.; MINASNY, B. On digital soil mapping. **Geoderma**, v.117, p.3-52, 2003. DOI: 10.1016/S0016-7061(03)00223-4.

SANTOS, H.G. dos; JACOMINE, P.K.T.; ANJOS, L.H.C. dos; OLIVEIRA, V.A. de; LUMBRERAS, J.F.; COELHO, M.R.; ALMEIDA, J.A. de; CUNHA, T.J.F.; OLIVEIRA, J.B. de. **Sistema brasileiro de classificação de solos**. 3.ed. rev. e ampl. Brasília: Embrapa, 2013. 353p.

SANTOS, R.D. dos; LEMOS, R.C. de; SANTOS, H.G. dos; KER, J.C.; ANJOS, L.H.C. dos; SHIMIZU, S.H. Manual de descrição e coleta de solo no campo. 7.ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2015. 100p.

SILVA, S.H.G.; MENEZES, M.D. de; OWENS, P.R.; CURI, N. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. **Geoderma**, v.267, p.65-77, 2016. DOI: 10.1016/j.geoderma.2015.12.025.

SIRGAS: Sistema de Referencia Geocéntrico para las Américas. Available at: <http://www.sirgas.org/index.php?id=64>. Accessed on: June 20 2017.

TEN CATEN, A.; DALMOLIN, R.S.D.; PEDRON, F.A.; MENDONÇA-SANTOS, M. de L. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos. **Revista Brasileira de Ciência do Solo**, v.35, p.53-62, 2011. DOI: 10.1590/S0100-06832011000100005.

TEN CATEN, A.; DALMOLIN, R.S.D.; MENDONÇA-SANTOS, M. de L.; GIASSON, E. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, v.42, p.1989-1997, 2012. DOI: 10.1590/S0103-84782012001100013.

TESKE, R.; GIASSON, E.; BAGATINI, T. Produção de um mapa pedológico associando técnicas comuns aos mapeamentos digitais de solos com delineamento manual de unidades de mapeamento. **Revista Brasileira de Ciência do Solo**, v.39, p.950-959, 2015. DOI: 10.1590/01000683rbcs20140285.

VILLELA, A.L.O. **Mapeamento digital de solos da Formação Solimões sob floresta tropical amazônica**. 2013. 101p. Tese (Doutorado) – Universidade Federal do Rural do Rio de Janeiro, Rio de Janeiro.