Decision trees for digital soil mapping on subtropical basaltic steeplands

Elvio Giasson[1]*; Eliana Casco Sarmento[1]; Eliseu Weber[2]; Carlos Alberto Flores[3]; Heinrich Hasenack[2]

[1]UFRGS – Depto. de Solos, C.P. 15100 – 91501- 970 – Porto Alegre, RS – Brasil.
[2]UFRGS – Centro de Ecologia, C.P. 15007 – 91501- 970 – Porto Alegre, RS – Brasil.
[3]Embrapa Clima Temperado, C.P. 403 – 96001- 970 – Pelotas, RS – Brasil.
*Corresponding author <giasson@ufrgs.br>

ABSTRACT: When soil surveys are not available for land use planning activities, digital soil mapping techniques can be of assistance. Soil surveyors can process spatial information faster, to assist in the execution of traditional soil survey or predict the occurrence of soil classes across landscapes. Decision tree techniques were evaluated as tools for predicting the ocurrence of soil classes in basaltic steeplands in South Brazil. Several combinations of types of decicion tree algorithms and number of elements on terminal nodes of trees were compared using soil maps with both original and simplified legends. In general, decision tree analysis was useful for predicting occurrence of soil mapping units. Decision trees with fewer elements on terminal nodes yield higher accuracies, and legend simplification (aggregation) reduced the precision of predictions. Algorithm J48 had better performance than BF Tree, RepTree, Random Tree, and Simple Chart.
Key words: GIS, soil survey, terrain analysis

Árvores de decisão para o mapeamento digital de solos em encostas basálticas subtropicais

RESUMO: Quando levantamentos de solos não estão disponíveis para atividades de planejamento de uso das terras, técnicas de mapeamento digital de solos podem ser úteis. Mapeadores de solos podem processar as informações espaciais rapidamente, auxiliando na execução de levantamentos de solos tradicionais ou prevendo a ocorrência de classes de solos na paisagem. Avaliaram-se técnicas de análise de decisão na predição da ocorrência de classes de solos em áreas de encostas basálticas no Sul do Brasil. Várias combinações de tipos de algoritmos de árvore de decisão e quantidade de elementos nos nós terminais das árvores de decisão foram testadas usando mapas de solos com a legenda original e com legenda simplificada. Em geral, o uso de árvores de decisão foi eficaz na predição de ocorrência de unidades de mapeamento de solos. Menor número de elementos no nó terminal das árvores de decisão produziu acurácias mais altas e a simplificação da legenda (agregação) reduziu a precisão das predições. O algoritmo J48 teve melhor desempenho que BF Tree, RepTree, Random Tree, e Simple Chart.
Palavras-chave: SIG, levantamento de solos, análise do terreno

Introduction

Soil surveys are almost a mandatory source of information for land use planning. However, finer scale soil maps are scarce in developing countries because conventional mapping techniques are time-consuming and costly. Moreover, conventional soil maps do not explicitly document a pedologist's mental model of the soil-landscape relationships that guide mapping work and which could support subsequent land use planning process. To overcome this limitation, the association of traditional soil survey techniques and new technologies of digital soil mapping (DSM) may improve the overall process of soil mapping, making the process more quantitative.

Environmental variables that control soil variation and distribution across landscapes can be quantified and mapped. Several approaches have been applied in DSM: logistic regressions (Figueiredo et al., 2008; Giasson et al., 2006; Giasson et al., 2008), discriminant analysis (Bell et al., 2000; Pavlik and Hole, 1997), fuzzy logic (Zhu, 1997), kriging (Voltz at al., 1997; Knotters at al., 1995), and decision trees (Lagacherie and Holmes, 1997; Bui and Moran, 2001). Classification and regression tree analysis represent a supervised approach to classification. Predictive soils mapping using decision tree analysis, which split up the datasets into blocks by a tree, can increase mapping efficiency and accuracy by extracting relationships between soil types and environmental variables, applying these relationships to predict soil types for unmapped areas and explicitly documenting the process.

For the establishment of relationships between these variables and soil spatial distribution, decision trees may be used for predicting the occurrence of soil map units based on terrain and hydrologic variables. In decision tree analysis, observations enter at the root node and a test is applied to best separate the observations

into classes, making groups purer. The observation then passes along to the next node and the process of testing the observations to split them into classes continues until the observation reaches a terminal node. Observations reaching a particular terminal node are classified the same way. Many terminal nodes may make the same classification. Several different paths may be followed for an observation to become part of a particular class.

The objective of this study was to test the use of decision tree algorithms for producing digital soil maps. Using as predictive variables soil distribution information and terrain parameters, several decision tree algorithms were used with variable number of objects in terminal nodes of trees. Predicted soil maps were evaluated compared to conventional soil survey map.

Material and Methods

The study area is located in Vale dos Vinhedos, Bento Gonçalves (Rio Grande do Sul State, Brazil) (Figure 1), and comprises an area of 6,735 km^2 between longitudes 51º34'31.86" W and 51º33'1.86" W and latitudes 29º10'31.78" S and 29º09'1.78" S. It is located in the physiographic region Planalto das Araucárias, a plateau formed by basaltic rocks of the Serra Geral Formation with relief characterized by steep slopes (IBGE, 1986). The regional climate is subtropical with mild summer (Köppen Cfb class), with mean temperatures of the coldest month between -3ºC and 13ºC, mean temperature of the warmest month lower than 22ºC, and precipitation well distributed along the year (Moreno, 1961). A detailed soil survey of the study area at scale 1:10,000 was produced according to conventional soil survey procedures, including extensive field work, airphoto interpretation, and soil taxonomic classification according to SiBCS - Brazilian Soil Classification System (Embrapa, 2006). The final map was produced with a legend comprising six soil map units (Table 1).

For evaluating the use of decision tree for predictive soil mapping, two soil maps were evaluated: i) a soil map with original six soil map units legend, and ii) a soil map with modified legend, grouping soil map units by major soil groups (order) in the Brazilian Soil Classification System (Embrapa, 2006), resulting in a four map units legend, named as follows: MU1 (Ultisols or Argissolos - according to Brazilian Classification), MU2 (Inceptisols/Cambissolos), MU34 (Molisols/Chernossolos, grouping of MU3 and MU4), and MU56 (Entisols/Neossolos, grouping of MU5 and MU6).

A digital cartographic base was created by aerophotogrametric survey with digital restitution of planimetry and altimetry with 5 m between contour lines. A Triangular Irregular Network (TIN) digital elevation model (DEM) was produced by linear interpolation of contour lines, and a parabolic function was used to adjust the relief representation on valleys and elevations. The TIN was converted to a raster DEM with a spatial resolution of 5 m. The raster DEM was used for calculating nine predictive variables: slope gradient, profile curvature, planar curvature, curvature (combination of planar and profile curvature), flow direction, flow accumulation, flow length, Stream Power Index (SPI), and Topographic Wetness Index (TWI) (Wolock and McCabe, 1995). Each of these hydrologic or landform parameters was selected to be used as predictive variable because they may represent changes on soil-forming factors and, therefore, are believed to be informative on the occurrence of soil map units.

Data sampling for training consisted of 1,333 points (one observation per each 0.005 km^2) distributed randomly among soil map units as test points or training
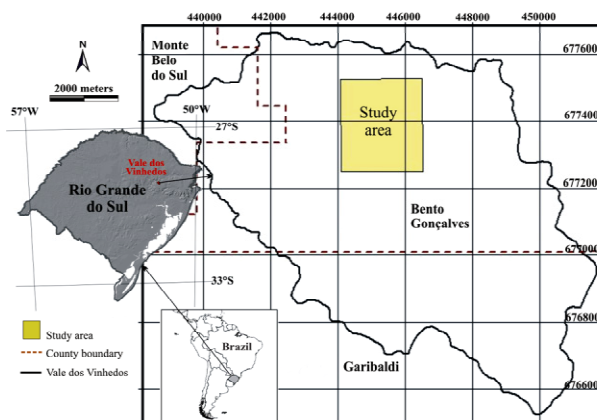

Figure 1 – Location of the study area in the Vale dos Vinhedos region of the Rio Grande do Sul State, Brazil. Map projection UTM Zone 22J, datum SAD69.

Table 1 – Soil map units of the conventional soil map.

| Symbol | Taxonomic classification | | Area | |
	SBCS (Embrapa, 2006)	Soil Taxonomy	km^2	%
MU1	Argissolo Vermelho	Typic Hapludults	1,016	15.1
MU2	Cambissolo Háplico	Typic Dystrudepts	2,813	41.8
MU3	Chernossolo Argilúvico	Typic Argiudolls	0.452	6.7
MU4	Chernossolo Háplico	Typic Hapludolls	1,836	27.3
MU5	Neossolo Litólico	Lithic Udorthents	0.275	4.1
MU6	Neossolo Regolítico	Typic Udorthents	0.344	5.1
Total			6,735	100.0

points (67% and 33%, respectively). The option for using random points intended to eliminate subjectivity, to allow simple reproducibility (Hengl and Rossiter, 2003), and to have sample points distributed proportionally to areas occupied by each soil map unit. In each random point all data layers were sampled. The attribute table of the layer of points received fields with values of elevation, DEM derived parameters, and soil map units (using both original and simplified legend). The data sampled in ArcView 3.2 environment (ESRI, 1999) was exported as tables and converted into a comma delimited file (CSV format) used for decision tree analysis, using the software Weka (Witten and Frank, 2005).

Decision tree analysis tested combinations of eight minimum number of elements in terminal nodes (150, 125, 100, 75, 50, 25, 10, or 5) and five decision tree algorithms (J48, BF Tree, Rep Tree, Random Tree, and Simple Chart). All of these algorithms are decision tree techniques which use a supervised machine learning approach. J48 is an implementation of a decision tree technique that is based on the C4.5 algorithm which was originally proposed by Quinlan (1986). REPtree method is also based on C4.5 algorithm and can produce classification (discrete outcome) or regression trees (continuous outcome). It sorts numeric attributes only once. SimpleCart method is a decision tree analysis based on Breiman et al. (1984). BFTree is a best-first decision tree learner and it is a learning algorithm for supervised classification learning. Best-first decision trees represent an alternative approach to standard decision tree techniques such as the C4.5 algorithm since they expand nodes in best-first order instead of a fixed depth-first order.

Decision tree structures were generated by partitioning the data recursively into a number of groups, each division being chosen to differentiate the response variable in the resulting nodes. Results were evaluated individually for each resulting tree of a set of 80 generated trees using error matrices. Additionally, accuracy statistics were computed for weighted error matrices, used because it was considered that not all mapping errors are equally serious for soil map users. Therefore, weights were assigned for calculating weighted error matrices using subjective criteria based on the degree of the importance of the maping mistake for land use planning. In the weight matrices, diagonals and the off-diagonals range from 0 to 1, with a value of 0 indicating that the mistake is more serious, and as the values increase towards 1, the mistake is considered decreasingly serious. A value 1 means that two classes are considered identical for accuracy assessment.

Criteria for evaluation of better response and selection of best tree were number of soil map units predicted, number of correctly classified instances, kappa index, mean absolute error, and size of the tree. For each legend scenario (original and grouped legend), the best combination of algorithm and number of objects in terminal nodes were selected. The output was imported in ArcView (ESRI, 1999) by implementing the generated

decision rules and creating a map for each soil map unit. These maps were overlayed, resulting in predicted soil maps.

Accuracies of the produced soil maps were determined by error matrices (Congalton, 1991). These are matrices with columns representing reference soil map units (i.e. the original soil survey, with the original or aggregated legends) and rows representing predicted digital soil map units. Each cell in the matrix contains the proportion in the mapped class of its row that was in fact observed in the class of its column. The diagonal represent agreement between the original map and predicted map and off diagonals represent misclassifications. In our study, error matrices comparing all pixels of the whole study area were used.

Four map accuracy matrices were calculated: i) overall accuracy, which is the proportion of correctly-classified pixels compared to total number of pixels; ii) producer accuracy, which is the probability of a pixel in a given unit of soil to be classified correctly; iii) user accuracy, which is the probability of a pixel classified as a given soil mapping unit to be correctly classified; and iv) Kappa index which compares the agreement between original and predicted soil maps against that which might be expected by chance alone (Cohen, 1960). Accuracy statistics were additionally calculated for weighted error matrices, which consider that not all mapping errors are equally serious for soil map users. Weights for calculating weighted error matrices were assigned subjectively, using as criterium the authors´ perception of the importance of soil characteristics for changing soil behavior when used for different agricultural land use types.

Results and Discussion

In algorithm runs reproducing the soil map with the original legend, the best performance (overall accuracy = 66.4%, Kappa index = 0.518, mean absolute error = 0.094, 145 terminal nodes) was observed using algorithm J48 with a minimum number of elements at the terminal node of N = 5 (Table 2). Although other combinations produced greater overall accuracy (up to 71.7%) or kappa indices (up to 0.575), they could only estimate five soil classes. Using an aggregated legend (legend of four classes), the best combination was algorithm J48 and minimum number of elements at terminal node of N = 25, which produced an overall accuracy of 73.6%, kappa index of 0.583, mean absolute error of 0.160, with a tree of 29 terminal nodes. Other combinations generated greater overall accuracy (up to 71.7%) or kappa index (up to 0.575), but only three soil map units were predicted (Table 2).

Algorithm J48 performed best, predicting the occurrence of all soil map units with higher overall accuracy and higher kappa indexes (Table 2). Variations in the number of objects in terminal node influenced accuracy, size of trees, and number of predicted classes. Small

Giasson et al.

Table 2 – Results of decision tree analysis in function of variations of algotithms and size of terminal nodes.

	Original legend						Simplified legend				
Simple chart											
N	OA	Kappa	MAE	Size	MU	N	OA	Kappa	MAE	Size	MU
150	65.8	0.439	0.121	3	2	150	66.4	0.435	0.208	3	2
125	65.8	0.439	0.121	3	2	125	66.4	0.435	0.208	3	2
100	65.8	0.439	0.121	5	2	100	66.4	0.435	0.208	3	2
75	65.8	0.439	0.121	5	2	75	66.4	0.435	0.208	3	2
50	66.4	0.500	0.109	7	4	50	69.7	0.513	0.182	5	3
25	68.9	0.529	0.107	9	5	25	70.2	0.545	0.170	13	4
10	71.1	0.572	0.099	9	5	10	70.9	0.542	0.164	25	4
5	71.3	0.574	0.098	43	5	5	70.2	0.536	0.159	31	4
Random tree											
150	53.9	0.226	0.141	21	2	150	69.2	0.491	0.197	19	2
125	53.9	0.226	0.141	21	2	125	69.2	0.491	0.193	21	2
100	54.3	0.255	0.138	25	2	100	69.2	0.491	0.193	25	2
75	58.3	0.332	0.133	37	2	75	69.4	0.495	0.189	31	2
50	63.6	0.430	0.120	55	4	50	63.9	0.403	0.206	49	3
25	61.1	0.408	0.123	103	3	25	64.8	0.436	0.185	111	4
10	63.6	0.466	0.108	253	6	10	60.8	0.392	0.185	253	4
5	52.1	0.328	0.108	467	6	5	61.2	0.418	0.167	463	4
Rep Tree											
150	66.0	0.444	0.121	3	2	150	67.0	0.445	0.205	3	2
125	66.0	0.444	0.121	3	2	125	67.0	0.445	0.205	3	2
100	66.0	0.444	0.121	3	2	100	67.0	0.445	0.205	3	2
75	66.0	0.444	0.121	5	2	75	67.0	0.445	0.205	5	2
50	68.2	0.509	0.112	5	3	50	67.0	0.445	0.205	7	2
25	67.1	0.516	0.105	7	5	25	69.8	0.515	0.180	15	3
10	70.4	0.560	0.101	29	5	10	69.8	0.515	0.180	19	3
5	68.4	0.535	0.103	47	5	5	68.7	0.507	0.170	37	4
BF Tree											
150	65.8	0.439	0.121	3	2	150	66.4	0.435	0.208	3	2
125	65.8	0.439	0.121	3	2	125	66.4	0.435	0.208	3	2
100	65.8	0.439	0.121	5	2	100	66.4	0.435	0.208	3	2
75	65.8	0.439	0.121	5	2	75	66.4	0.435	0.208	3	2
50	64.2	0.430	0.118	7	3	50	65.3	0.433	0.202	9	3
25	68.9	0.529	0.102	19	5	25	70.2	0.545	0.168	25	4
10	70.2	0.543	0.106	21	5	10	71.7	0.552	0.163	33	4
5	70.2	0.543	0.106	41	5	5	71.0	0.543	0.163	47	4
J48											
150	65.8	0.439	0.121	5	2	150	67.4	0.467	0.197	5	2
125	66.4	0.493	0.113	5	3	125	68.9	0.489	0.194	5	2
100	67.1	0.496	0.112	5	3	100	70.3	0.508	0.190	5	2
75	68.0	0.506	0.111	5	3	75	70.9	0.518	0.188	5	2
50	65.6	0.506	0.105	15	5	50	69.8	0.514	0.183	13	3
25	71.7	0.575	0.099	25	5	25	73.6	0.583	0.160	29	4
10	71.3	0.571	0.095	79	5	10	72.9	0.565	0.159	57	4
5	66.4	0.518	0.094	145	6	5	70.0	0.536	0.151	133	4

N = size of terminal node; OA= overall accuracy; Kappa = kappa index; MAE = mean absolute error; Size = size of the tree; MU = number of soil map units predicted.

number of objects in terminal nodes reduced mean absolute error and increased overall accuracy, kappa index, size of the tree, and number of predicted soil map units. When using same algorithms, overall accuracy was equal or larger producing maps with grouped map legend (four map units legend), although generating maps with higher mean absolute error. When using the grouped legend, more combinations of algorithms and sizes of terminal node were able to estimate the correct number of soil classes (Table 2).

Original soil maps (with original and aggregated legends) and estimated soil maps prepared using these optimal combinations described above are presented in Figure 2. A comprehensive evaluation of the reproducibility of the conventional soil map with digital soil mapping techniques using error matrices, in this case considering the agreement between conventional and estimated soil maps, is presented in Tables 3 and 4.

Overall accuracies were 68% and 69%, and Kappa indexes were 0.54 and 0.56, respectively for original and grouped legend maps. Based on the weight matrices (Tables 5 and 6), accuracy statistics were computed for weighted error matrices. Error evaluation using weighted error matrix for the estimated map with original legend (Table 7) presented an overall accuracy of 68% and kappa index of 0.571. For the map estimated with grouped legend (Table 8), overall accuracy was 80% and kappa index was 0.574

Error matrices show user's accuracies from 0.28 to 0.78 for unweighted error matrices (Tables 3 and 4) and from 0.52 to 0.88 for weighted error matrices (Tables 7 and 8), with higher user accuracies for weighted error matrices occurring for MU4 (which represents 27.3% of the area) (Table 7) for the original legend map and MU34 (which represents 34.0% of the area) (Table 8) for the grouped legend map. Producer's per-class reliability was

Table 3 – Error matrix for evaluation of maps with original map legend.

Predicted	Reference						Total	User Accuracy
	MU1	MU2	MU3	MU4	MU5	MU6		
MU1	0.06	0.04	0.01	0.01	0.00	0.01	0.13	0.47
MU2	0.07	0.35	0.02	0.02	0.00	0.02	0.48	0.72
MU3	0.00	0.01	0.02	0.01	0.00	0.00	0.05	0.54
MU4	0.01	0.01	0.01	0.21	0.02	0.01	0.27	0.78
MU5	0.00	0.00	0.00	0.01	0.02	0.00	0.03	0.63
MU6	0.01	0.01	0.00	0.01	0.00	0.01	0.04	0.28
Total	0.15	0.42	0.07	0.27	0.04	0.05	1.00	
Producer Accuracy	0.41	0.84	0.37	0.77	0.50	0.21		

Table 4 – Error matrix for evaluation of maps with grouped map legend.

Predicted	Reference				Total	User Accuracy
	MU1	MU2	MU34	MU56		
MU1	0.05	0.09	0.01	0.00	0.15	0.45
MU2	0.03	0.36	0.02	0.00	0.42	0.68
MU34	0.02	0.05	0.26	0.01	0.34	0.79
MU56	0.01	0.03	0.03	0.02	0.09	0.67
Total	0.11	0.53	0.32	0.03	1.00	
Producer Accuracy	0.32	0.87	0.76	0.23		

Table 5 – Weights assigned for the weighted error matrix for original legend map evaluation.

Predicted	Reference					
	MU1	MU2	MU3	MU4	MU5	MU6
MU1	1.00	0.66	0.66	0.66	0.33	0.33
MU2	0.66	1.00	0.66	0.66	0.33	0.33
MU3	0.66	0.66	1.00	0.9	0.33	0.33
MU4	0.66	0.66	0.9	1.00	0.33	0.33
MU5	0.33	0.33	0.33	0.33	1.00	0.8
MU6	0.33	0.33	0.33	0.33	0.8	1.00

computed similarly, using the table column totals. Moving down the columns (predicted classes), errors of omission occur when the mapper fails to correctly identify this reference site in its true class, i.e., the mapper failed to map the site in its correct class, leading to a lower producer accuracy. Producer accuracies ranged from 0.23 to 0.87 for unweighted error matrices and from 0.44 to 0.91 for weighted error matrices. For weighted error matrices (Tables 7 and 8), higher producer accuracies were associated to MU2, which covers 41.8% of the study area, both for the original and aggregated legend maps. For weighted error matrices (Tables 7 and 8), lower producer's accuracies were found for MU6 (which represents 5.1% of the area) and MU1 (which represents 15.1% of the area), for the original and aggregated legend maps, respectively. For unweighted error matrices (Tables 3 and 4), lower producer accuracies were found for MU6 which represents 5.1% of the study area

for the original legend map (Table 3) and MU56 (which represents 9.2% of the area) for the aggregated legend map (Table 4). When predicting soil classes using the selected combination of algorithms and number of elements on terminal node, both user and mapper accuracies were higher for soil classes that associated with larger mapped areas. These accuracies increased when

Table 6 – Weights assigned for the weighted error matrix for grouped legend map evaluation.

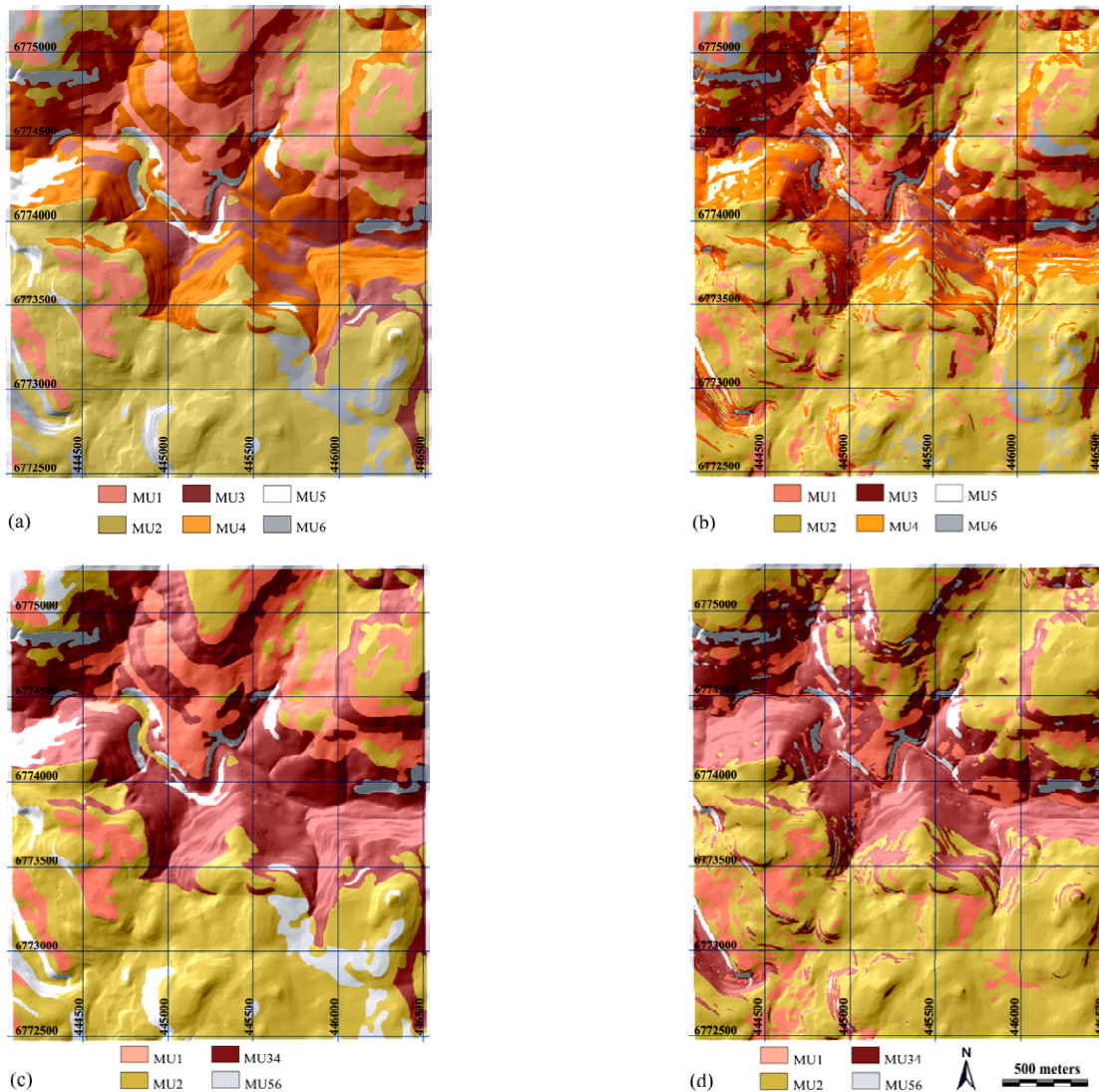| Predicted	Reference			
	MU1	MU2	MU34	MU56
MU1	1.00	0.66	0.66	0.00
MU2	0.66	1.00	0.66	0.33
MU34	0.66	0.66	1.00	0.33
MU56	0.00	0.33	0.33	1.00

Figure 2 – Soil maps of the study area: (a) soil map with six class legend, (b) estimated soil map with six class legend, (c) soil map with four class legend, and (d) estimated soil map with four class legend.

Table 7 – Weighted error matrix for evaluation of maps with original map legend

Predicted	Reference						Total	User Accuracy
	MU1	MU2	MU3	MU4	MU5	MU6		
MU1	0.06	0.02	0.01	0.01	0.00	0.00	0.10	0.59
MU2	0.05	0.35	0.01	0.01	0.00	0.01	0.43	0.81
MU3	0.00	0.00	0.02	0.01	0.00	0.00	0.04	0.61
MU4	0.01	0.01	0.01	0.21	0.01	0.00	0.24	0.86
MU5	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.81
MU6	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.52
Total	0.12	0.39	0.06	0.25	0.03	0.02	1.00	
Producer Accuracy	0.52	0.89	0.45	0.85	0.72	0.44		

Table 8 – Weighted error matrix for evaluation of maps with grouped map legend.

Predicted	Reference				Total	User Accuracy
	MU1	MU2	MU34	MU56		
MU1	0.05	0.02	0.01	0.00	0.08	0.58
MU2	0.06	0.36	0.03	0.01	0.46	0.78
MU34	0.01	0.01	0.26	0.01	0.29	0.88
MU56	0.00	0.00	0.00	0.02	0.02	0.86
Total	0.12	0.40	0.31	0.04	1.00	
Producer Accuracy	0.42	0.91	0.83	0.50		

using weighted error matrices as more importance is attributed to separation of soil units that are considered more distinct.

The use of random sampling generated different number of samples per soil map unit and represents map units that occupy larger areas with more samples. Because of this, lower user's and producer's accuracies were obtained for map units less represented in the random sampling, such as MU3, MU5, and MU6. The use of stratified sampling could reduce errors in the classification of these map units that occupy smaller areas and are underrepresented in the random sampling.

Conclusions

The algorithm J48 performed better than other tested algorithms for decision tree analysis. Soil map legend simplification by class aggregation resulted in only a small increase in both overall accuracy and Kappa index. This map legend simplification reduced precision and is not recommended, as not all soil map units could be discriminated. Application of weighted error matrices for evaluating estimated maps with original legend did not increase the overall accuracy, but increased kappa indices slightly, while the use of weighted error matrices for the evaluation of maps estimated with aggregated legend increased slightly the overall accuracy and kappa index. The use of weighted error matrices may be beneficial in some cases, although assigning weights is a critical step due to their subjective nature.

Acknowledgements

References

Bell, J.C.; Grigal, D.F.; Bates, P.C. 2000. A soil-terrain model for estimating spatial patterns of soil organic carbon. p. 295-31. In: Wilson, J.P.; Gallant, J.C., eds. Terrain analysis: principles and applications. Wiley, New York, NY, USA.

Bui, E.N.; Moran, C.J. 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. Geoderma 103: 79-94.

Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. 1984. Classification and regression trees. Wadsworth, Monterey, CA, USA.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37-46.

Congalton, R.G. 1991. A review of assessing the accuracy of classification of remotely sensed data. Remote Sensing of Environment 37: 35-46.

Empresa Brasileira de Pesquisa Agropecuária [Embrapa]. 2006. Brazilian System of Soil Classification. Centro Nacional de Pesquisa em Solos. Rio de Janeiro, RJ, Brazil. (in Portuguese).

Environmental Systems Research Institute [ESRI]. 1999. ArcView 3.2. Redland, CA, USA.

Figueiredo, S.R.; Giasson, E.; Tornquist, C.G.; Nascimento, P.C. 2008. Multiple logistic regression applied to soil survey in Rio Grande do Sul State, Brazil. Revista Brasileira de Ciência do Solo 32: 2779-2785. (in Portuguese, with abstract in English).

Giasson, E.; Clarke, R.T.; Inda Júnior, A.V.; Merten, G.H.; Tornquist, C.G. 2006. Digital soil mapping using logistic regression on terrain parameters in Southern Brazil. Scientia Agricola 63: 262-268.

Giasson, E.; Figueiredo, S.R.; Tornquist, C.G.; Clarke, R.T. 2008. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in Southern Brazil. p. 225-231. In: Hartemink, A.E.; McBratney, A.; Mendonça-Santos, M.L., eds. Digital soil mapping with limited data. Springer, New York, NY, USA.

Hengl, T.; Rossiter, D.G. 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Science Society of America Journal 67: 1810-1822.

Instituto Brasileiro de Geografia e Estatística [IBGE]. 1986. Map sheet SH.22 Porto Alegre and part of map sheets SH.21 Uruguaiana and SI.22 Lagoa Mirim:. IBGE, Rio de Janeiro, RJ, Brazil. (Levantamento de Recursos Naturais, 33). CD-ROM. (in Portuguese).

Knotters, M.; Brus, D.J.; Oude Voshaar, J.H. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67: 227-246.

Lagacherie, P.; Holmes, S. 1997. Addressing geographical data errors in a classification tree soil unit prediction. International Journal of Geographical Information Science 11: 183-198.

Moreno, J.A. 1961. Climate of the Rio Grande do Sul State. Secretaria da Agricultura, Porto Alegre, RS, Brazil. (in Portuguese).

Pavlik, H.F.; Hole, F.D. 1997. Soilscape analysis of slightly contrasting terrains in southeastern Wisconsin. Soil Science Society of America Journal 41: 407-413.

Quinlan, J.R. 1986. Induction of decision trees. Machine Learning 1: 81-106.

Voltz, M.; Lagacherie, P.; Louchart, X. 1997. Predicting soil properties over a region using sample information from a mapped reference area. European Journal of Soil Science 48: 19-30.

Witten, I.; Frank, E. 2005. Data Mining: Practical Machine Learning; Tools and Techniques. 2ed. Morgan Kaufmann, Burlington, MA, USA.

Wolock, D.M.; McCabe, G.J. 1995. Comparison of single and multiple flow-direction algorithms for computing topographic parameters in TOPMODEL: Water Resources Research 31: 1315-1324.

Zhu, A.X. 1997. A similarity model for representing soil spatial information. Geoderma 77: 217-242.