

## Note

### Digital soilscape mapping of tropical hillslope areas by neural networks

Waldir de Carvalho Junior<sup>1\*</sup>, César da Silva Chagas<sup>1</sup>, Elpídio Inácio Fernandes Filho<sup>2</sup>, Carlos Antonio Oliveira Vieira<sup>2</sup>, Carlos Ernesto Gonçalves Schaefer<sup>2</sup>, Silvio Barge Bhering<sup>1</sup>, Marcio Rocha Francellino<sup>3</sup>

<sup>1</sup>Embrapa Solos, R. Jardim Botânico, 1024 - 22460-000 - Rio de Janeiro, RJ - Brasil.

<sup>2</sup>UFV - Depto. de Ciência do Solo, Av. P.H. Rolfs, s/n - 36570-000 - Viçosa, MG - Brasil.

<sup>3</sup>UFRRJ/Instituto de Florestas - Depto. de Ciências Ambientais, Rod. BR 465, km 47 - 23890-000 - Seropédica, RJ - Brasil.

\*Corresponding author <waldircj@cnpq.embrapa.br>

Edited by: Miguel Cooper / Jussara Borges Regitano

**ABSTRACT:** Geomorphometric variables are applied in digital soil mapping because of their strong correlation with the disposition and distribution of pedological components of the landscapes. In this research, the relationship between environmental components of tropical hillslope areas in the Rio de Janeiro State, Brazil, artificial neural networks (ANN), and maximum likelihood algorithm (MaxLike) were evaluated with the aid of geoprocessing techniques. ANN and MaxLike were applied to soilscape mapping and the results were compared to the original map. The ANN architectures with seven and five neurons in the hidden layer produced the best classifications when using samples obtained systematically. When random samples were applied, the best neural net architectures were within 22 and 16 neurons in the hidden layer. In conclusion, the ANN can contribute to soilscape surveys, making map delineation faster and less expensive. The digital elevation model (DEM) and its derived attributes can contribute to the understanding of the soil-landscape relationship of tropical hillslope areas; the use of artificial neural networks and MaxLike is feasible for digital soilscape mapping. The systematic sampling method provided a global accuracy of 70 % and 65.9 % for the ANN and the MaxLike, respectively. When the random sampling method was applied, the ANN had a global accuracy of 69.6 %, and the MaxLike had an accuracy of 62.1 %, considering the total study area in relation to the reference map.

**Keywords:** geomorphometric attribute, digital soil mapping, digital elevation model

### Introduction

The study of the relationship between environmental components can contribute to a better understanding of our environment. Improvements in computational performance, information technology, GIS techniques, and the ever-increasing volume and sophistication of geographical data have led to extraordinary advances in digital soil mapping (Hempel et al., 2006). Digital soil mapping (DSM) has been actively investigated for more than a decade (MacMillan, 2006). However, it has not yet achieved widespread acceptance for operational mapping by mainstream mapping agencies.

In Brazil, apparently DSM is still in its early stage. Soil researchers have just begun to work on the DSM field. In fact, only a few Brazilian published scientific papers in this field were presented in Mendonça-Santos and McBratney (2006), such as the research done by Carvalho Júnior et al. (2006), Giasson et al. (2006) and Valladares and Hott (2006). The use of artificial neural networks (ANN) in DSM provides the means to automatically quantify the spatial distribution of the tropical hillslope soils. Brown et al. (1998) used the ANN to classify glaciated landscapes. Chang and Islam (2000) used the ANN to estimate soil physical properties, and Giles and Franklin (1998) have studied an automatic digital mapping approach to classify hillslope units. Moreover, Behrens et al. (2005) have utilized the ANN to predict soil units.

The main goal of this research is to utilize the ANN to classify soils in hillslope areas (Domain of "Mar de Morros") in the Rio de Janeiro State, identifying the disposition and

distribution of the geomorphometrics and pedological components. The research was carried out with the aid of geoprocessing techniques and supervised classification, focusing to some extent on the identification of soils by geomorphometric features. The goal is to compare ANN soilscape classification results with the original map of the area and the map generated by the maximum likelihood estimation classifier (MaxLike).

### Materials and Methods

The area under study has 16,470 ha and is located in the north of the northwest region of the Rio de Janeiro State in Brazil (47°48' to 47°54' W, 20°48' to 21°00' S). It can be seen in the "Varre Sai" sheet of the topographical chart (IBGE, 1991) and is in the Itabapoana and Muriaé watersheds (Figure 1). The digital cartographic base contains contour lines with 20-m resolution, elevations of some specific points and drainage networks (1:50,000 scale), utilizing the UTM projection system with the "Córrego Alegre" horizontal datum.

The defined soilscape units are groups in a high hierarchic level with soil classes of the Rio de Janeiro State Soil Survey (Carvalho Filho et al., 2003). Thus the soilscape units of the domain of the Oxisols (S1), Ultisols (S2), Aquent Entisols (S3), Inceptisols (S4), and Rock Outcrops (S5) (Figure 1) were generated.

The neural network simulator "Neural Java Network Simulator" (JavaNNS, 2001), developed by the Wilhelm-Schickard Institute for Computation Science in Tübingen (Germany), was utilized. This simulator is based on the Neural Stuttgart Networks Simulator 4.2 kernel (SNNS, 1998), with a new graphical interface.

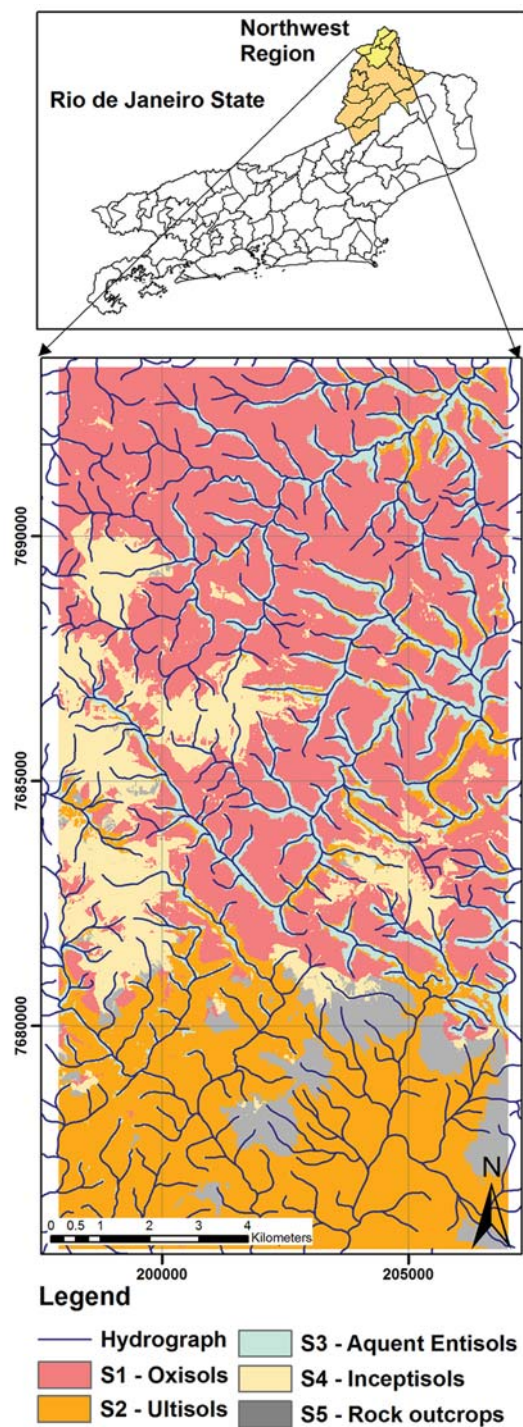


Figure 1 – Area location and the defined soilscape units in the Rio de Janeiro State, Brazil.

The ANN classification procedure includes: (i) definition of input parameters and output soilscape; (ii) collection of training and validation samples; (iii) generation of training files and validation files in the JavaNNS format; (iv) training of the ANN architectures; definition of the best ANN architectures; (v) validation of the trained ANNs; (vi) generation of the confusion matrix; (vii) comparison with the MaxLike results; (viii) application of the trained ANNs and MaxLike into the studied area; and (ix) genera-

tion of thematic images for comparisons. The DEM, from which the majority of ANN input attributes had been generated, has the following characteristics in meters: cell size of 20, maximum value of 1017, minimum of 190, average of 638 and standard deviation of 164, with 901 lines and 457 columns.

Among the geomorphometric parameters utilized in this research, the hydrologic corrected digital elevation model (DEM) is the most important one. This model is generated from contour lines, elevations of surface-specific points and a hydrograph, using Topogridtool (ArcInfo 9) and posterior removal of the spurious depressions, according to Hutchinson (1989) and Hutchinson and Gallant (2000). From the DEM and a map algebra, secondary attributes have been generated, such as: slope (S), aspect (A), curvature (C), plain curvature (PC), profile curvature (PRC), superficial flow accumulation (SFA), superficial flow direction (SFD), relative altimetry of sub-basins (RASB) (Carvalho Júnior et al., 2008) and Euclidean distance of the hydrograph (EDH). These geomorphological parameters compose one of the soil formation factors related to relief conditions (McBratney et al., 2003; Chagas et al., 2010), and are the most used formation factors in soil mapping because of their strong correlation with the spatial variability of soil attributes over landscape (Ceddia et al., 2009).

In this study the following softwares were used: (i) ARC/INFO version 8.2 - ESRI – to produce the DEM and other attributes; (ii) ArcView GIS version 3.2a - ESRI – to view and computed results; (iii) JavaNNS - Java Neural Network Simulator - Version 1.1 - University of Tübingen.; (iv) Executables funcpow, gerapat and Max\_like\_cof – (provided by Professor Carlos A. O. Vieira - Civil Engineering Department – Viçosa Federal University) – for data management; (v) ERDAS IMAGINE version 8.5 – ERDAS Systems – to collect data samples; and (vi) Microsoft Excel – 2000 – Microsoft Corporation – for data management.

A strategy to stop training the ANNs was assumed to be 20,000 cycles, in which the whole set of training samples is used for every cycle. A strategy was also used to reduce the learning rate ( $\eta$ ) with the following training progression in the form: (i) from 0 to 10,000 cycles  $\rightarrow \eta = 0.2$ ; (ii) from 10,000 to 15,000 cycles  $\rightarrow \eta = 0.1$ ; and (iii) from 15,000 to 20,000 cycles  $\rightarrow \eta = 0.075$ . The ANN architecture is composed of an input layer with a neuron for each input discriminating variable, one hidden layer and an output layer with as many neurons as there are informational classes. The project was developed using the backpropagation learning algorithm, in which all neurons in one layer are fully connected with the next layer of neurons.

To evaluate the classification accuracy, the kappa statistic and its variance were used. These measures were obtained from the confusion or error matrices. All the feature (discriminating variables) values were rescaled to a range between 0 (zero) and 1 (one) due to their magnitude and were stacked in one layer. The samples were acquired in two ways, e. g., systematic and random sampling, by retaining the values of the cells for each attribute and generating the training and validating files in the JavaNNS format. Both systematic and random sampling for training and validation were carried out to evaluate the ANN classification performance. In the first case, the systematic samples were selected using the fieldwork knowledge from the fieldwork

itself, and they represent each soilscape shown in Figure 1. The fieldwork sites were selected ‘a priori’. On the other hand, the samples randomly selected were collected independently to be representative of each soilscape unit without any constraints.

The validation samples for the ANNs were all obtained randomly without considering their location. Table 1 shows the number of samples acquired for training and validation of the systematic samples and the random samples. The accuracy evaluation was made using the soilscape map as reference data.

### Results and Discussion

#### Soilscape supervised classification with systematic samples

Classifications were carried out with the artificial neural network (ANN) and MaxLike algorithm. The ANNs were trained with ten neurons in the input layer, each neuron corresponding to one aspect, relative altimetry, curvature, slope, flow direction, flow accumulation, Euclidean distance of the hydrograph, altimetry, profile, and plain curvature. The informational classes are the five defined soilscape units (Table 1). Thus, the ANN output layer had five neurons representing the five informational classes. The amount of training and validation samples collected both randomly and systematically is shown in Table 1. For the systematically collected samples, various architectures were tested in search of one that presented the minimum SSE value (Sum of Squared Errors), and a lesser number of neurons in the hidden layer. Thus, some ANN architectures were trained sequentially, using ten neurons in the input layer and five neurons in the output layer.

After 20,000 cycles of training with the training samples by increasing the number of hidden neurons from one to ten (constructive method), the lower SSE values were obtained for the five neurons with SSE of 9.18 and for the seven neurons with SSE of 4.0. The same samples that were utilized in the ANN were also utilized in MaxLike to test and compare the results. Then the validation file was applied to two ANN architectures and to the MaxLike classifier, and the classification results of both were compared with the reference map in order to generate the ANN and MaxLike confusion matrices and derived accuracy measures (Table 2). In an overview, the ANN with seven neurons presented a global accuracy of 88.3 %, and a Kappa value of 0.851 (Table 2). The Oxisol soilscape (S1) had a better accuracy (99.7 %), followed by Aquent Entisol soilscape (S3, 97.7 %), and by the Ultisol soilscape (S2, 87.1 % accuracy).

Table 1 – Numbers of training and validation samples for five soilscape units.

	Number of samples and form of acquisition			
	Random		Systematic	
	training	validation	training	validation
S1 – Oxisols	125	272	350	359
S2 – Ultisols	350	250	332	350
S3 – Aquent Entisols	250	246	280	300
S4 – Inceptisols	257	315	300	345
S5 – Rocks Outcrops	272	126	135	130

The MaxLike classification presented a lower value of global accuracy, equal to 81.9 %.

A significance matrix of the kappa index was elaborated from previous results and presented in Table 3. The values out of the diagonal line (greater than 1.96) indicate significant differences between the classifiers. Thus, there is no significant difference between the ANN5 and ANN7, and both presented better and significantly different results from the MaxLike classification. The S1 and S3 classes had better accuracy results considering the ANN classifications (Table 3). On the other hand, the S2 and S3 classes had better classification in the MaxLike algorithm. Therefore, the ANN5 and ANN7 were utilized to classify the study area, and MaxLike was used just for comparison.

Table 2 – Confusion matrix, statistical Kappa coefficient and global accuracy from the validation sample classifications to the ANNs and MaxLike.

ANN with five neurons in the hidden layer – ANN5							
Class	S1	S2	S3	S4	S5	Total	Users (%)
S1	358	10	16	81	0	465	77
S2	0	296	3	3	19	321	92.2
S3	0	0	281	0	0	281	100
S4	1	1	0	259	5	266	97.4
S5	0	43	0	2	106	151	70.2
Total	359	350	300	345	130		
Producers (%)	99.7	84.6	93.7	75.1	81.5		
Global accuracy = 87.6				Kappa = 0.842			
ANN with seven neurons in the hidden layer – ANN7							
Class	S1	S2	S3	S4	S5	Total	Users (%)
S1	358	12	5	81	0	456	78.5
S2	0	305	2	1	22	330	92.4
S3	0	0	293	0	0	293	100
S4	1	4	0	256	9	270	94.8
S5	0	29	0	7	99	135	73.3
Total	359	350	300	345	130		
Producers (%)	99.7	87.1	97.7	74.2	76.2		
Global accuracy = 88.3				Kappa = 0.851			
Maximum likelihood Algorithm (MaxLike)							
Class	S1	S2	S3	S4	S5	Total	Users (%)
S1	282	0	32	49	363	363	77.7
S2	49	346	17	40	494	494	70
S3	0	0	251	0	251	251	100
S4	28	1	0	256	293	293	87.4
S5	0	3	0	0	83	83	96.4
Total	359	350	300	345			
Producers (%)	78.6	98.9	83.7	74.2			
Global accuracy = 81.9				Kappa = 0.767			

Where: S1 = oxisols, S2 = ultisols, S3 = Aquent Entsoils, S4 = Inceptisols and S5 = rocks Outcrops.

Table 3 – Significance Matrix of Kappa index to the ANN and MaxLike tests.

Classifier	ANN5	ANN7	MaxLike1
KAPPA	0.842	0.851	0.767
Variance	0.000117	0.000112	0.000163
ANN5	<b>77,843</b>		
ANN7	0.595	<b>80,412</b>	
MaxLike1	<b>4,482</b>	<b>5,065</b>	<b>60,076</b>

#### Soilscape supervised classification with random samples

Adopting the same procedure employed in the ANN with systematic samples, different architectures were tested and those with 22, 16, 17, 11, and 15 neurons in the hidden layer were selected. The significance of the kappa statistic matrix for evaluating each classification (Table 4) showed that there was a difference between the ANN14 results, the ANN16 and the ANN22. There was no difference between the other ANNs. By comparing the classification results of the MaxLike and all other ANNs, the MaxLike classifier was different from the others, indicating a worse performance. Thus, it was defined that the ANN22, ANN16, ANN17, ANN11 and ANN15 was used in the soilscape classification of the area. Even though MaxLike presented a bad result, it was used for comparison.

#### Study area classifications

To evaluate ANN and MaxLike classifications for the area, the confusion matrix results, global accuracy, and Kappa index were calculated and are presented in Table 5. The value of global accuracy shows that the best results were obtained by utilizing both ANN5 and ANN16, with global accuracy of 70 % and 69.6 %, respectively. The MaxLike had a lower value of global accuracy (65.9 %). The kappa coefficient of these three classifiers, ANN5, ANN16, and MaxLike, were 0.55, 0.55, and 0.49, respectively.

Figure 2 shows the ANN5 classification. It can be used in a visual comparison with Figure 1, which presents the reference map and on which a discussion of inclusion and omission errors is based. Considering the visual aspect of all classification results, it seems to have consistent forms. For example, the Aquent Entisol soilscape follows the drainage network. Moreover, little groups of cells spread out inside soilscapes were delineated, indicating that the smaller ANN architectures generalize the better soilscape classes. As shown in the ANN16 classification, the soilscape classes seem to be relatively more spread out throughout the study area, and even more so when the MaxLike classification is analyzed. In this visual evaluation, a good delineation of limits between the Ultisols (S2) and Oxisols (S1) is observed.

As an evaluation of the classifications, two Oxisol profiles described in the reference soil survey map of the Rio de Janeiro State (Carvalho Filho et al., 2003) were found and compared with the three maps generated by ANN5, ANN22, and MaxLike. It was observed that the two Oxisol profiles were both well delineated by the three classifiers according to their geographic positions. The values of the kappa coefficient, according to a proposal of Landis and Koch (1977), indicated good quality. Regarding the global accuracy, all were higher than 65.9 %, which is

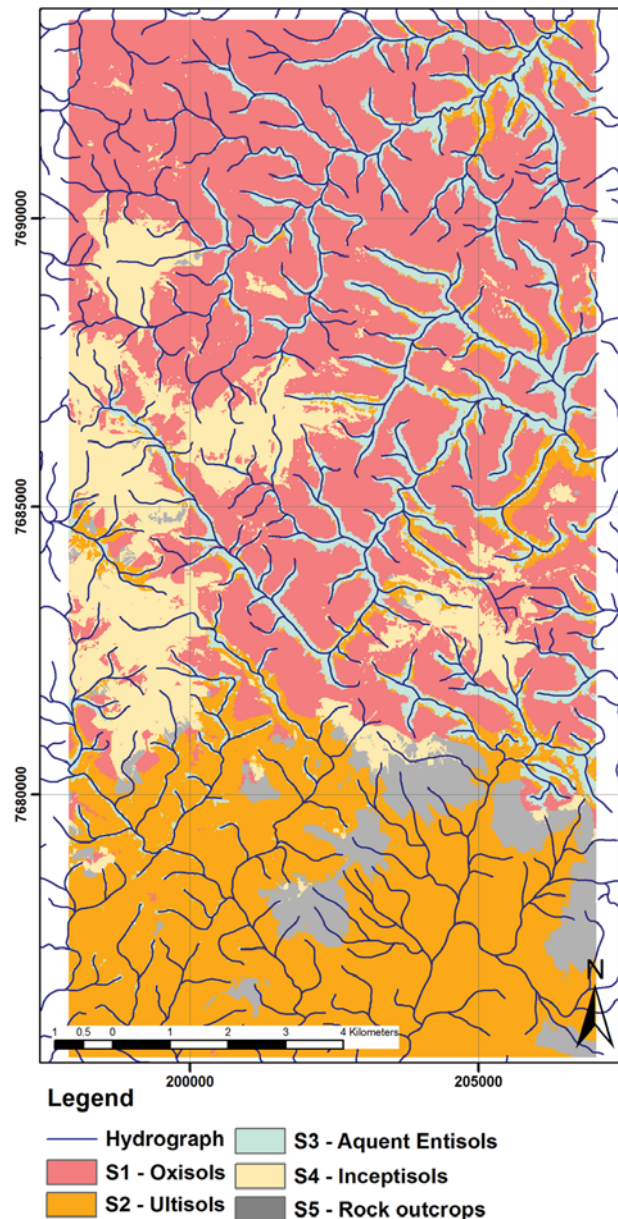


Figure 2 – Classification of the area by the ANN5.

considered a satisfactory result. The ANN5 and MaxLike classifications fit to the field knowledge, indicating that the curvature is an important attribute in the classification, even more so when associated with other attributes, such as Euclidean distance, relative altimetry, slope, and the altimetry (Figure 3).

#### Conclusions

The attributes derived from DEM can contribute to the understanding of the tropical hillslope area soil-landscape relationship, separately or in groups. The use of classifiers like ANNs and MaxLike is feasible for digital soilscape classification, assisting with map delineation, and the terrain attributes discriminate the considered soilscapes. The classifications of

Table 4 – Significance matrix of the kappa statistic to the validation samples.

Classifier	ANN11	ANN13	ANN14	ANN15	ANN16	ANN17	ANN22	MaxLike
Kappa	0.821	0.818	0.79	0.821	0.833	0.826	0.837	0.703
ANN11	<b>64,306</b>							
ANN13	0.166	<b>63,875</b>						
ANN14	1,659	1,497	<b>57,926</b>					
ANN15	0	0.166	1,657	<b>64,109</b>				
ANN16	0.675	0.842	2,335	0.674	<b>67,344</b>			
ANN17	0.279	0.445	1,938	0.278	0.396	<b>65,506</b>		
ANN22	0.903	1,071	2.56	0.901	0.229	0.625	<b>68,114</b>	
MaxLike2	<b>5,900</b>	<b>5,743</b>	<b>4,230</b>	<b>5,893</b>	<b>6,583</b>	<b>6,181</b>	<b>6,803</b>	45,665

Table 5 – Global accuracy and kappa coefficient of classification results of the area.

Classifier	Number of neurons in the hidden layer	Global accuracy	Kappa index
ANN7	7	68.5	0.54
ANN5	5	70.0	0.55
MaxLike1	--	65.9	0.49
ANN11	11	67.7	0.52
ANN15	15	66.1	0.51
ANN16	16	69.6	0.55
ANN17	17	66.3	0.51
ANN22	22	69.1	0.54
MaxLike2	--	62.1	0.46

the area by ANN classifiers were all better than the ones obtained from MaxLike, and the ANN simulator Java NNS was suggested because of its friendly interface and free software. The backpropagation method gives an acceptable classification error after training and learning processes, but the real global accuracy depends on obtaining ground control points during new field trips.

References

Behrens, T.; Forster, H.; Scholten T.; Steinrucken, U.; Spies E.; Goldschmitt, M. 2005. Digital soil mapping using artificial neural network. Journal of Plant Nutrition and Soil Science 168: 21-33.

Brazilian Institute of Geography and Statistics [IBGE]. 1991. Geoscience Directory. Cartography Department. Varre-Sai sheet SF.24-V-A-IV-3. IBGE, Rio de Janeiro, RJ, Brazil. (in Portuguese).

Brown, D.G.; Lusch, D.P.; Duda, K.A. 1998. Supervised classification of types of glaciated landscapes using digital elevation data. Geomorphology 21: 233-250.

Carvalho Filho, A.; Lumbreras, J.F.; Wittern, K.P.; Lemos, A.L.; Santos, R.D. 2003. Rio de Janeiro State Soil Survey. Rio de Janeiro: Embrapa Solos. 1 MAP. 1:250.000. Available at: <http://mapoteca.cnps.embrapa.br> [Accessed Jun. 15, 2008]. (in Portuguese).

Carvalho Júnior, W.; Fernandes Filho, E.I.; Vieira, C.A.O.; Chagas, C.S. 2006. Geomorphometric attributes used to soil-landscapes supervised classification of mountainous areas. In: Mendonça-Santos, M.L.; McBratney, A.B., eds. Proceedings of the Global Workshop on Digital Soil Mapping 2 - DSM for regions and countries with sparse spatial data infrastructure Embrapa, Rio de Janeiro, RJ, Brazil. (CD-ROM).

Carvalho Júnior, W.; Fernandes Filho, E.I.; Vieira, C.A.O.; Schaefer, C.E.G.R.; Chagas, C.S. 2008. Geomorphometric attributes applied to soil-landscapes supervised classification of mountainous tropical areas in Brazil: a case study. p. 357-365. In: Hartemink, A.E.; McBratney, A.; Mendonça-Santos, M.L. Digital soil mapping with limited data. Springer, Berlin, Germany.

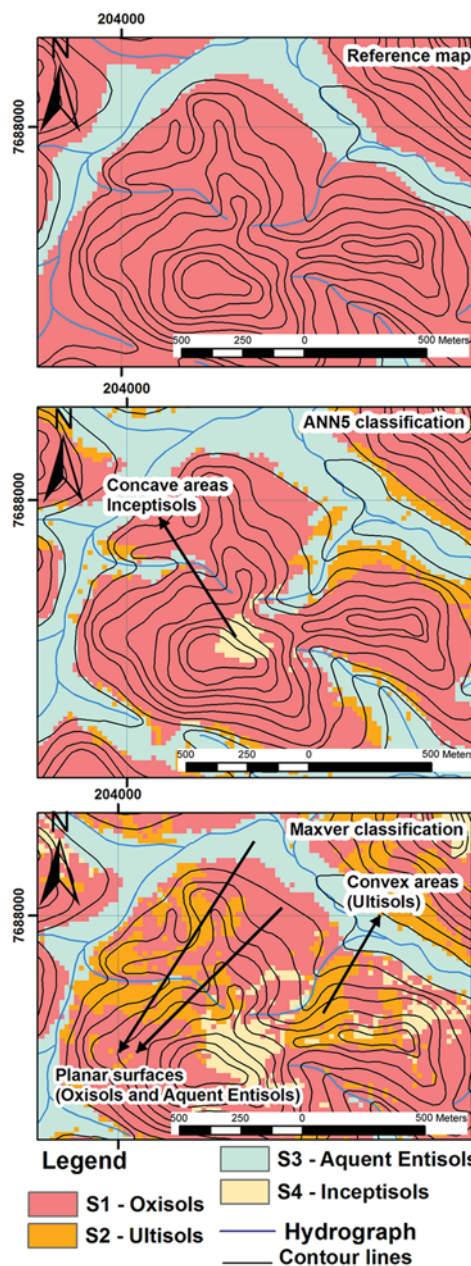


Figure 3 – Comparison between the reference map and the ANN5 and MaxLike classification.

- Ceddia, M.B.; Vieira, S.R.; Villela, A.L.O.; Mota, L.S.; Anjos, L.H.C.; Carvalho, D.F. 2009. Topography and spatial variability of soil physical properties. *Scientia Agricola* 66: 338-352.
- Chagas, C.S.; Fernandes Filho, E.L.; Vieira, C.A.O.; Schaefer, C.E.G.R.; Carvalho Júnior, W. 2010. Topographic attributes and Landsat7 data in the digital soil mapping using neural networks. *Pesquisa Agropecuária Brasileira* 45: 497-507. (in Portuguese, with abstract in English).
- Chang, D.; Islam, S. 2000. Estimation of soil physical properties using remote sensing and artificial neural network. *Remote Sense of Environment* 74: 534-544.
- Giasson, E.; Figueiredo, S.R.; Tornquist, C.G.; Clarke, R.T. 2006. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in Southern Brazil. In: Mendonça-Santos, M.L.; McBratney, A.B., eds. *Proceedings of the Global Workshop on Digital Soil Mapping 2. - DSM for regions and countries with sparse spatial data infrastructure*. Embrapa, Rio de Janeiro, RJ, Brazil. (CD-ROM).
- Giles, P.T.; Franklin, S.E. 1998. An Automated approach to the classification of the slope units using digital data. *Geomorphology* 21: 251-264.
- Hempel, J.W.; Hammer, R.D.; Moore, A.C.; Bell, J.C.; Thompson, J.A.; Golden, M.L. 2006. Challenges to digital soil mapping. In: Mendonça-Santos, M.L.; McBratney, A.B., eds. *Proceedings of the Global Workshop on Digital Soil Mapping 2. - DSM for regions and countries with sparse spatial data infrastructure*. Embrapa, Rio de Janeiro, RJ, Brazil. (CD-ROM).
- Hutchinson, M.F.; Gallant, J.C. 2000. Digital elevation models and representation of terrain shape. p. 29-50. In: Wilson, J.P.; Gallant, J.C. eds. *Terrain Analysis: Principles and Applications*. Wiley, Hoboken, NJ, USA.
- Hutchinson, M.F. 1989. A new procedure for gridding elevation and stream data with automatic removal of spurious pits. *Journal of Hydrology* 106: 211-232.
- Java Neural Network Simulator [JavaNNS]. JavaNNS: version 1.1. 2001. Tübingen, University of Tübingen, Germany. Available at: [http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/welcome\\_e.html](http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/welcome_e.html) [Accessed Dec. 6, 2008].
- Landis, J.R.; Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174.
- McBratney, A.B.; Santos, M.L.M.; Minasny, B. 2003. On digital soil mapping. *Geoderma* 117: 3-52.
- MacMillan, R.A. 2006. Experiences with applied DSM: protocol, availability, quality and capacity building. In: Mendonça-Santos, M.L.; McBratney, A.B., eds. *Proceedings of the Global Workshop on Digital Soil Mapping 2. - DSM for regions and countries with sparse spatial data infrastructure*. Embrapa, Rio de Janeiro, RJ, Brazil. (CD-ROM).
- Mendonça-Santos, M.L.; McBratney, A.B., eds. 2006. *Program and Full Papers for the 2nd Global Workshop on Digital Soil Mapping - DSM for regions and countries with sparse spatial data infrastructure*. Embrapa, Rio de Janeiro, RJ, Brazil. (CD-ROM).
- Stuttgart Neural Network Simulator [SNNS]. 1998. SNNS: version 4.2. Tübingen, University of Tübingen, Germany. Available at: <http://ftp.informatik.uni-tuebingen.de> [Accessed Oct. 6, 2005]
- Valladares, G.S.; Hott, M.C. 2006. GIS and geomorphometry to pedological mapping. In: Mendonça-Santos, M.L.; McBratney, A.B., eds. *Proceedings of the Global Workshop on Digital Soil Mapping 2. - DSM for regions and countries with sparse spatial data infrastructure*. Embrapa, Rio de Janeiro, RJ, Brazil. (CD-ROM).

---

Received May 21, 2010

Accepted March 13, 2011