Digital soil mapping using reference area and artificial neural networks

Gustavo Pais de Arruda[1], José A. M. Demattê[2]*, César da Silva Chagas[4], Peterson Ricardo Fiorio[3], Arnaldo Barros e Souza[2], Caio Troula Fongaro[2]

[1]APagri Agronomic consultancy, Av. Adauto Pinheiro, 401 – 15040-368 – São José do Rio Preto, SP – Brazil.

[2]University of São Paulo/ESALQ – Dept. of Soil Science, Av. Pádua Dias, 11, PO Box 9 – 13418-900 – Piracicaba, SP – Brazil.

[3]University of São Paulo/ESALQ – Dept. of Biosystems Engineering.

[4]Embrapa Soils – R. Jardim Botânico, 1024 – 22460-000 – Rio de Janeiro, RJ – Brazil.

*Corresponding author <jamdemat@usp.br>

Edited by: Dionysis Bochtis

ABSTRACT: Digital soil mapping is an alternative for the recognition of soil classes in areas where pedological surveys are not available. The main aim of this study was to obtain a digital soil map using artificial neural networks (ANN) and environmental variables that express soil-landscape relationships. This study was carried out in an area of 11,072 ha located in the Barra Bonita municipality, state of São Paulo, Brazil. A soil survey was obtained from a reference area of approximately 500 ha located in the center of the area studied. With the mapping units identified together with the environmental variables elevation, slope, slope plan, slope profile, convergence index, geology and geomorphic surfaces, a supervised classification by ANN was implemented. The neural network simulator used was the Java NNS with the learning algorithm "back propagation." Reference points were collected for evaluating the performance of the digital map produced. The occurrence of soils in the landscape obtained in the reference area was observed in the following digital classification: medium-textured soils at the highest positions of the landscape, originating from sandstone, and clayey loam soils in the end thirds of the hillsides due to the greater presence of basalt. The variables elevation and slope were the most important factors for discriminating soil class through the ANN. An accuracy level of 82% between the reference points and the digital classification was observed. The methodology proposed allowed for a preliminary soil classification of an area not previously mapped using mapping units obtained in a reference area.

Keywords: map extrapolation, pedological survey, landscape attributes, pedological classes, data mining

Introduction

Knowledge of the properties and attributes of soil is extremely important for adequate soil management programs whose aim is environmental sustainability and food production efficiency. However, factors such as a decline in the number of soil scientists, high cost and time-consuming execution of the soil survey have contributed to a shortage of such information.

Data mining methods can provide solutions to assist in the automatic extraction of information from a set of available data (Behrens et al., 2005). Allied to this, strategies such as extrapolation of soil-landscape relationships from a reference area for physiographically similar regions where these relationships are not yet known, can help reduce the time required for and the high cost of soil surveys.

The reference area method aims to characterize in detail the soil from a small representative area, called a reference area, a natural region where the main soil classes are identified and established in terms of soil-landscape relationships (mapping rules). The knowledge acquired is used to facilitate and accelerate the investigation of soils in other areas in the same natural region. The importance, advantages, and disadvantages of this method were highlighted by Favrot (1989) and by Lagacherie et al. (1995).

The use of the reference area as a digital mapping strategy of soils associated with a certain predictive method can be found in studies such as Grinand et al. (2008) - classification trees; Kempen et al. (2009) - multinomial logistic regression; Yigini and Panagos (2014) - regression-kriging. Among the data mining methods, artificial neural networks are characterized by the possibility of efficient handling of large amounts of data together with an ability to generalize, as found in the literature (Zhu, 2000; Behrens et al., 2005; Boruvka and Penizek, 2007).

However, extrapolation of soil classes from a reference area using ANN has not yet been reported in the literature. Such a technique would help to diminish the soil map shortage, by generating at least one set of prior information in areas not mapped based on a smaller pedological mapping. It could be a quick and inexpensive alternative for obtaining soil information for use in projects that require this type of data. Thus, the present study aimed to evaluate the efficiency of ANN in the extrapolation of soil classes for an unmapped area from a reference area mapped in detail, based on environmental variables that express the soil-landscape relationship. The hypothesis is that the digital soil map would match the field soil distribution, supporting the use of ANN to identify soils.

Materials and Methods

Characterization of field area

The research was carried out in an area of 11,072 ha in Barra Bonita municipality , State of Sao Paulo, Brazil, located between the Universal Transverse Mercator

267

Arruda et al.   Artificial neural network and soil classes

(UTM) coordinates 750 530.3 and 764 287.8 m E and 7 524 287.8 and 7 514 029.1 m N (Fuse 22, *Datum* SAD 69). The climate is Cwa (Köppen system), subtropical of altitude with dry winters. Maximum average temperature is 30 °C during the hottest month and 12.2 °C during the coldest month. Average annual rainfall is 1,471 mm. The geology consists of sandstones cemented by clay of Itaqueri Formation and basalts of Serra Geral Formation, São Bento group (IPT, 1981). The original vegetation was deciduous broadleaf forest which has now been superceded by sugar cane.

Environmental variables

In the present study, the following environmental variables were considered: elevation, slope, slope profile, slope plan and convergence index, all derived from a digital elevation model (DEM) with a spatial resolution of 20 m, geology and geomorphic surface. The DEM was obtained by applying the TOPO to RASTER function in the ArcGIS 9.2 software (ESRI) to the digitalized level curves (at each 5 m), hydrography and reference points mapped on topographical maps from the IGC (Geographic and Cartographic Institute of São Paulo), on a spatial scale of 1:10,000. Geomorphological surfaces were identified based on relief variations using stereoscopic analysis on aerial photographs on a 1:30,000 scale. These surfaces were bounded by discontinuities and changes in slope gradients, using the analysis of terrain attributes derived from the DEM. The level curves obtained from the digitalization of planialtimetric charts give an interpretation of the local stratigraphy, according to the criteria proposed by Ruhe (1969) and Daniels et al. (1971).

The use of the "geomorphic surface" variable is justified because such information expresses a relationship between the soil distribution on the landscape and soil age. Thus, the soil factor "time" was indirectly considered together with the "relief" and "parent material" soil factors.

Digital soil mapping strategy

For a reference area (Favrot, 1989) using the ANN approach about 500 ha cultivated almost exclusively with sugar cane were utilized, located in the center of the study area (Figures 1 and 2), for which there was a soil survey with a level of detail on a scale of 1: 10,000.

The use of a reference area is based on the principle that the informations and classifications obtained in small areas that represent a set of physiographic features in particular can be extrapolated to areas with similar soil-landscape relationship.

To clarify the soil landscape relationship, three toposequences were also analyzed using point sequences and all soil units were included. For each toposequence, graphs of topography, clay and sand content, Al^{3+}, Sum of Bases (SB) and Cation Exchange Capacity (CEC) were generated.

Consequently, the data about the environmental variables used to feed the ANN originated from the ref-
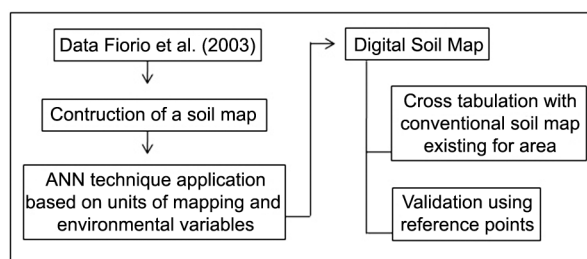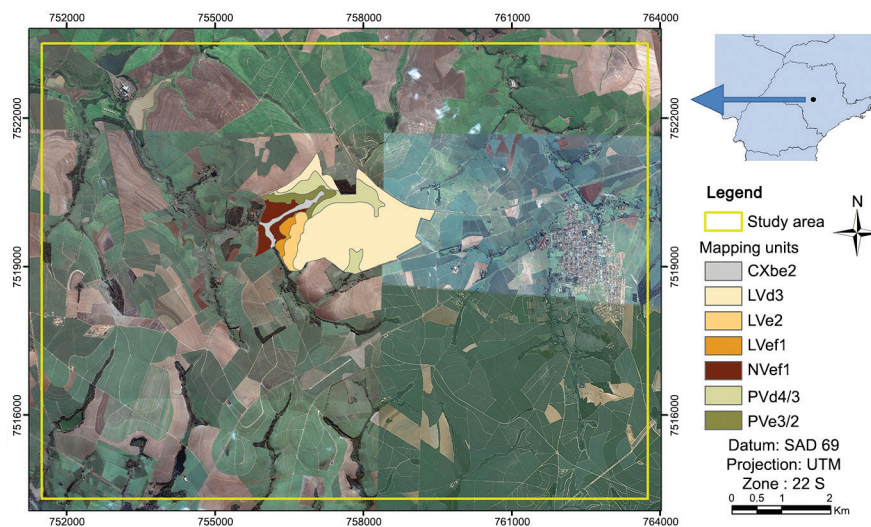

Figure 1 – Flowchart of the study.

Figure 2 – Convencional soil map of the reference area and study area location.

erence area. The process started with the creation of files with: (i) data for training each architecture of the neural network and (ii) data for the validation of each one of them. These training files allow the algorithm to learn the relationship between the environmental variables (input data) used by the network and the mapping units (output data). The validation data are used for evaluating the generalization performance of the learned mapping. All variable data were rescaled to a range from 0 to 1, in accordance with Chagas et al. (2010).

Next, information in each pixel of the reference area was extracted to create training files and validation files for each set of environmental variables tested. These files were independent from each other. An effort was made to follow the recommendations of Zhu (2000): the number of pixels collected for ANN training in each soil mapping unit should be 30 times the number of mapping units studied. As regards validation, the number of pixels was 50 % of that obtained for training. In a few cases, the total number of pixels obtained was lower than that. In such cases, 70 % of the pixels were used for training and 30 % for validation. After the collection, the files were converted to "Java Neural Network Simulator" format (JavaNNS), which is based on the "Stuttgart Neural Network Simulator 4.2 Kernel" (Zell et al., 1996).

The data set of entry data in ANN used in this study were as follows: All Variables (AV), All Variables except Slope (AV-S), All Variables except Elevation (AV-E), All Variables except Geology (AV-G), All Variables except Convergence Index (AV-CI), All Variables except Slope profile (AV-SPf), All Variables except Slope Plan (AV-SPl), All Variables except Geomorphic Surface (AV-GS). This strategy allows for evaluating the role of each variable in the prediction of soil mapping units. The number of neurons in the single internal layer was adjusted empirically (Hirose et al., 1991), using the mean quadratic error (MQE) as a criterion for training continuity. The MQE measures the difference between the estimated values and the expected values for each training event. The number of neurons in each output layer was kept constant, being equal to the number of mapping units (seven units). During the training of ANNs, the "back propagation" learning algorithm was used, with random weighting of interneurons ranging from -1.0 to + 1.0 and learning rate of 0.2. The number of learning cycles was 10,000.

During the test events statistical analyses were performed for each architecture of the network trained, using the validation files. Such analyses were the Kappa index and the global exactness, obtained from a confusion matrix (Congalton and Green, 1999) and from the analysis of a significance matrix obtained from the Kappa and variance results from each network. The occurrence of significant differences between Kappa values from the neural networks was identified using the Z statistics.

Based on the statistical results, the best neural network was chosen to generate the Digital Soil Map

(DSM). To generalize the neural network information, the file of the chosen network and an image file with the *pixel* data from the *raster* files were processed using the JavaNNS software. At this stage, the chosen network already trained with the information from the reference area is used to process all pixels from the experimental area. The final image file obtained includes the classification of each pixel according to the relationships between mapping units and the environmental variables established by the network.

Validation of the Digital Soil Map

To validate the DSM, 22 random georeferenced points were collected in the field areas at depths of 0-20 cm and 80-100 cm, summing 44 samples. The particle size composition was determined and the color obtained using a colorimeter. Samples were classified up to the second level of SiBCS (EMBRAPA, 2013). A data base file was generated from this data and a "true-false" test run against the DSM. Finally, the punctual information in the field was compared with the classification of the map.

Results and Discussion

Conventional soil map and mapping unit characterization

The mapped area (Figure 2) has a topographic flat central portion with axis running from NW to W, medium textured soils at greater altitudes grading to clayey soils towards the drainage channel. These characteristics are related to the occurrence of sandstone at greater altitudes and basalt at the lower ones. From the center of the area towards the east, there was no soil texture gradient because there is only sandstone in that direction.

Table 1, soils are referred to as classified by the SiBCS (EMBRAPA, 2013) and the Soil Taxonomy classification (Soil Survey Staff, 2014).

The analyses of topographic profiles, clay and sand content, Al^{3+} and CEC and SB values at subsurface of the

Table 1 – Mapping units identified in the study area.

Simbol	SiBCS (EMBRAPA, 2013)	Texture	Soil Taxonomy	Area (%)
CXbe2	Eutrophic Haplic Cambisol	Clayey	Typic Eutrudepts	4
LVd3	Dystrophic Red Latosol	Loamy	Rhodic Hapludox	61
LVe2	Eutrophic Red Latosol	Clayey	Rhodic Eutrudox	6
LVef1	Eutroferric Red Latosol	Clayey	Rhodic Hapludox	3
NVef1	Eutroferric Red Nitosol	Clayey	Rhodic Kandiudalfs	8
PVe3/2	Eutrophic Red Argisol	Loamy	Typic Rhodudalfs	5
PVd4/3	Dystrophic Red Argisol	Sandy/ loamy	Typic Rhodudults	13
Total				100

three toposequences resulted in a better understanding of the soil distribution on the landscape (Figure 3).

Toposequence A is 1850 m in length and ranged from 610 m to 704 m of altitude. At the higher portions, the parent material is sandstone while basalt appears near the drainage channels (Figure 4A). The chemical and textural values discussed in this toposequence are from the 0.80-1.00 m layer. The LVd3 was the unit occupying the largest area and has a mean clay content of 160 g kg^{-1} and CEC of 24.3 mmol$_c$ kg^{-1}, occurring mainly in gently undulating to undulating relief. In general, it had a divergent curvature slope and concave - straight slope profile. It was the only unit that was mapped on geomorphic surface I and also occupied most of geomorphic surface II (Figure 4B).

Geomorphic surface I was assumed to be the oldest surface because it was uphill from the others. It was flatter and the slopes were less than 8 %, plan/divergent slope plan and straight/convex slope profile. The limit between geomorphic surfaces I and II occurred in places with transitions from small slope to greater slope. Geomorphic surface II cut geomorphic surface I in the downhill direction and was considered more recent. Geomorphic surface II had gently undulating to undulating relief and convex straight curvature. Geomorphic surface III had an inclination greater than 20 %, which was the greatest among the surfaces and occurred near the drainage channel. The contrast between geomorphic

surfaces II and III was clear due to the sharp change in slope. Geomorphic surface III was considered the youngest, and had, in general, a concave profile and convergent profile plan.

Following the LVd3 in toposequence A was the LVe2 whose parent material was a mixture of sandstone and reworked basalt materials. Such condition resulted in an increase in clay content to 490 g kg^{-1} (average) and in CEC (38.1 mmol$_c$ kg^{-1}). It had undulating to strongly undulating relief, divergent slope plan and concave slope profile.

At the final portion of this toposequence the LVef1 and NVef1 units were identified, having a clay content of 710 and 780 g kg^{-1} and CEC of 38.2 e 59.1 mmol$_c$ kg^{-1}, respectively, as a result of basaltic parent material. This portion of the toposequence had slopes greater than 8 % (where the NVef1 occurs) and concave slope profile. Geomorphic surfaces II and III were identified in such places.

The smallest mapping unit found in toposequence A was the CXbe2 near the drainage channel. This unit was mapped in small altitudes (603 m) in varying slope classes, in reliefs from flat to mountainous, convergent and divergent slope plan, convex slope profile and geomorphic surfaces II and III.

Toposequence B started at the same point as toposequence A and therefore, LVd3 was the first mapping unit identified. Following this, the PVd4/3 unit had undulating to strongly undulating relief, divergent slope
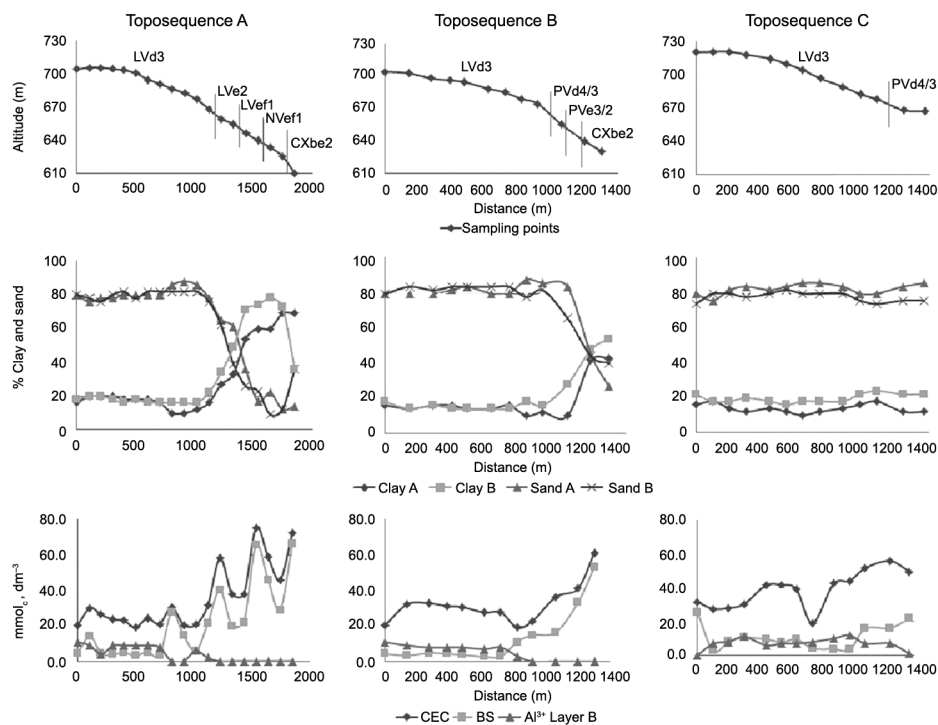

Figure 3 – Topographical profile, chemical and textural analyses of toposequences. Clay A = clay content at the depth 0 - 20 cm; Clay B = clay content at the 80 - 100 cm depth; Sand A = sand content at the 0 - 20 cm depth; Sand B = sand content in the 80 - 100 cm depth; CEC = Cation Exchange Capacity; BS = Sum of Bases; Al^{3+} = Content of exchangeable Al; Layer B = 80 - 100 cm depth.

270

Arruda et al.   Artificial neural network and soil classes

plan and concave and convex slope profile. This toposequence was almost entirely inside geomorphic surface III. Approaching the drainage channel with the concurrent increase in basalt contribution to the sandstone parent material, the PVe3/2 was found. The main difference between the Argisols was the parent material and the dominance of a more undulating relief in the PVe3/2. At the end of this toposequence, the CXbe2 was found again.

Comparing soil properties from these two toposequences, such as Al^{3+}, BS, CEC and clay and sand content at a 0.80-1.00 m depth (Figure 3), it was possible to conclude that parent material was the dominant factor in soil distribution. Samples with less BS, CEC, clay content and a greater amount of Al^{3+} and sand occurred at the higher positions and middle slope. At the end of the slopes of both toposequences A and B, this trend was inverted as a result of major basalt contribution to soil formation.

To better characterize the reference area, the length of toposequence C was 1370 m. LV3d and PVd4/3 were identified in this toposequence, developed mainly from sandstone. Sand and clay content in the subsurface was constant along the transect varying from 180 g kg^{-1} to 220 g kg^{-1} and from 760 to 800 g kg^{-1}, respectively. Compared to the other two toposequences, the values of Al^{3+} and SB were almost constant at the end section of toposequence C.
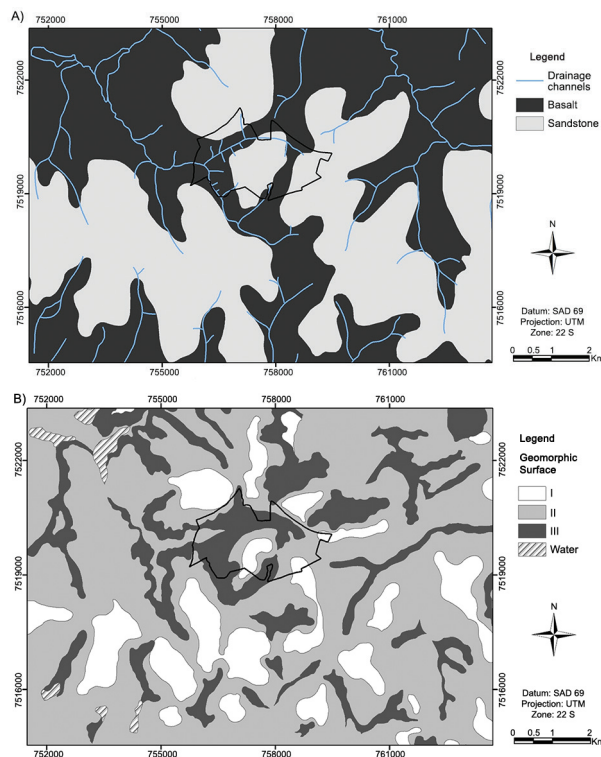
The distribution of mapping units along the toposequences was strongly influenced by the relief and parent material according to one of the two possible behaviors: (i) variation in chemical and textural values over the length of the toposequences in the transition sandstone-basalt and (ii) constant chemical and textural values over the length of top down sequences over sandstone.

Evaluation of artificial neural networks

The significance of the Kappa matrix (Table 2) was calculated from the results of the Kappa variance and network architecture obtained in the best performance which allowed for contrasting different results and choosing the neural network to be used to produce the DSM. The highest Kappa values were obtained from sets AV (0.809), AV-G (0.778) and AV-SPl (0.786), all using three neurons in the internal layer. A lower number of internal neurons yielded better results. Using several sets and from three to twenty neurons in the internal layer, Chagas et al. (2013) found using five to eight neurons in their study of two areas yielded the best results. Foody and Arora (1997) pointed out that greater and more complex networks are more efficient in characterizing a training set. Such behavior was not observed in the present study.

Lack of geological information or slope plan did not significantly interfere with the Kappa index, as compared to the use of all variables. In the first case, the geological information may have not contributed due to the small scale of available maps, as the situation described by Thomas et al. (1999) exemplifies. The similar average values for the slope plan values found in the mapping units (Figure 5) were the reason for such a small contribution.

The AV-S and AV-E yielded Kappa indexes of 0.635 and 0.4666 using seven and thirteen neurons in the internal layer. These were the sets with the worst performance. The greatest variation in elevation and slope would explain the greatest influence of these characteristics on soil discrimination (Figure 5). Gallant and Wilson (2000) highlighted the strong influence that these characteristics have on pedogenes and, as a consequence, on soil distribution on the landscapes. The same rationale can be established for the elevation.

The AV-GS yielded average values for the Kappa index (0.757) with nine neurons in the internal layer,

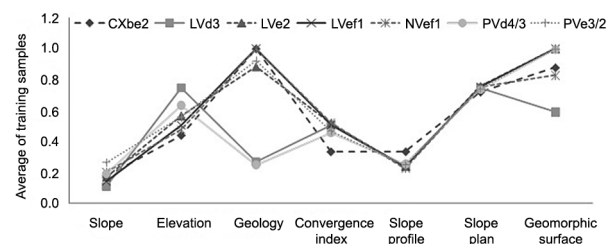Figure 4 – A) Parent material in the study area; B) Geomorphic surfaces in the study area.

Figure 5 – Environmental variables behavior according to soil classes.

meaning that the inclusion of the geomorphic surface enhanced soil identification. Therefore, landscape classification can contribute to the identification of soils in digital mapping even more than geology and slope plan which are the usual considerations. The use of geomorphic surface is a way of including the "time" variable as a soil formation factor in the analysis.

The neural network with the best performance during the training step (AV) was chosen to generate the DSM (Table 2). The confusion matrix (Table 3) generated by the application of AV, using the test samples, evaluated the capability of this neural network to discriminate soil classes as a function of the environmental variables.

The mapping unit LVef1 had the worst performance with agreement of 64 %. Eighteen (18) pixels of this unit were classified as NVef1. A closer look showed that this error had been caused by the small difference between the values of these two units (Figure 5). This mistake was due mainly to the values of the convergence index, slope profile and plan, which were very close between LVef1 and NVef1. On the other hand, slope and elevation showed a small but clear difference between the two units. The LVef1 had smaller slope and greater elevation compared to the NVef1. The main difference

between these two soil classes was morphological, such as structure and clay skins (Demattê and Terra, 2014). The digital mapping did not take into account such variables, and therefore was not efficient in discriminating these two units.

The LVe2 had the second worst performance, assigning 10 pixels wrongly to PVe3/2 and 6 to LVef1. At the end of the test step, 59 out of 81 pixels were assigned correctly to this mapping unit. The performance of the digital procedure in separating Latosols from Argisols was poor, but it was also difficult to make this separation in the field, due to the very similar landscape positions where they are found. The performance of the digital technique on the PVe3/2, LVd3, PVd4/3, and NVef1 mapping units was greater than 80 % and equal to 74 % for CXbe2.

Increase in the architectures' performance could be achieved if pixels near the border of units were eliminated or, if only field samples were used in the files, for training the architectures (Zhu, 2000).

Generation and characterization of digital soil map

The prediction of mapping units was made after training the neural network and validation of the architecture chosen. The digital soil map is pictured in Figure 6.

Table 2 – Kappa significance matrix for sets of environmental variables, using validation samples.

Set	AV	AV - S	AV - E	AV - G	AV - CI	AV - SPf	AV - SPl	AV - GS
	(3)[1]	(7)	(13)	(3)	(5)	(9)	(3)	(9)
Global exactness	83.7	69.1	54.5	81.1	77.0	76.5	81.4	79.3
Kappa	0.809	0.635	0.466	0.778	0.730	0.724	0.786	0.757
Variance[2]	0.305	0.445	0.512	0.341	0.390	0.391	0.338	0.361
AV	46.32							
AV - S	6.35*	30.10						
AV - E	12.00*	5.46*	20.59					
AV - G	1.22	5.10*	10.68*	42.13				
AV - CI	2.99*	3.28*	8.70*	1.77	36.97			
AV - SPf	3.22*	3.07*	8.58*	1.99*	0.22	36.61		
AV - SPl	0.91	5.39*	10.97*	0.31	2.07*	2.29*	42.75	
AV - GS	2.01*	4.29*	9.84*	0.79	0.99	1.20	1.10	39.84

*There is significant difference at the 95 % level; [1]Number of neurons in the internal layer; [2]Values times 1000; AV – all variables; S - slope; E - elevation; G - geology; CI – Convergence Index; SPf - Slope profile; SPl - Slope Plan; GS - Geomorphic Surface.

Table 3 – Confusion matrix obtained from the neural networks.

Unit	CXbe2	LVd3	LVe2	LVefl	NVefl	PVd4/3	PVe3/2	Total	User's	Z
CXbe2	61	2	0	0	0	1	0	64	95.3	31.14
LVd3	0	94	3	0	0	1	0	98	95.9	39.65
LVe2	2	0	59	1	0	3	4	69	85.5	17.34
LVefl	5	0	6	35	1	1	3	51	68.6	9.43
NVefl	8	0	3	18	104	0	0	133	78.2	17.97
PVd4/3	3	9	0	0	0	83	8	103	80.6	17.42
PVe3/2	3	0	10	1	0	3	74	91	81.3	16.83
Total	82	105	81	55	105	92	89	609		
Producer's	74.4	89.5	72.8	63.6	99	90.2	83.1			

Global exactness = 83.7; Kappa = 0.809; Variance = 0.000305; Z calculated = 46.319; Z table = 1.96

272

Arruda et al.   Artificial neural network and soil classes
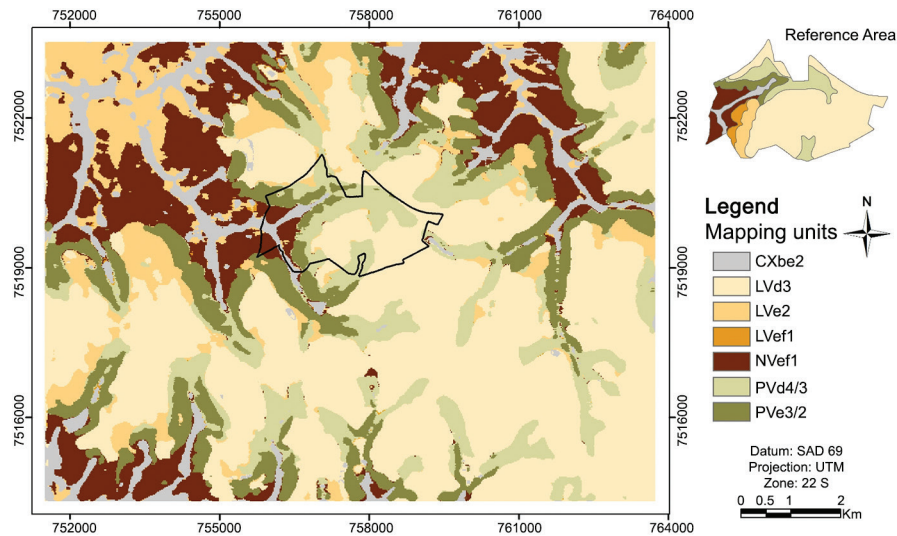

Figure 6 – Digital Soil Map.

The soil map made by the neural network was in agreement with the map made for the reference area, and follows the same soil-landscape relationship and soil-parent material. The LVd3 dominated the high altitudes (on top) down to the middle slope, grading to the Pvd4/3, PVe3/2, and LVe2 units. In areas with sandstone in the whole slope, the Argisols were assigned to a narrow strip at the lower third portion of the slope, grading to the CXbe2 near the drainage channel. This Cambisols pattern was the same in the entire area and can be explained by the fact that during the training stage of the neural network no output layer to represent the drainage channel and/or the river bed was created, and also due to the fact that these areas were mapped from these soils from the reference area.

The NVef1, LVe2, CXbe2, and a very small amount of pixels classified as LVef1 were found in the areas with basalt. The neural network was unable to discriminate between Nitosols and Latosols with clayey texture due to the great similarity in both the environmental data. The confusion matrix of AV network (Table 3) is in agreement with such an interpretation, because most of the pixels for validation of LVef1 were classified as being Nitosols. The digital classifier had difficulty in separating these two classes when, in certain cases, the morphological data was not taken into account.

Digital soil map validation

The percentage of agreement between the reference points and the digital soil map was 82 %, with eighteen true samples and four false samples. Two of the false points were related to difficulties in distinguishing between Latosols and Argisols, due to their similarity of distribution in the landscape. This agreement percentage is similar to that found by Zhu (2000) which found 77 % of true classification out of 64 field samples. On the other hand, areas in the south, with complex geology, can

generate lower readings. Vasques et al. (2015) observed a 57 to 67 % accuracy level using digital mapping, but in an area more complex than the present study.

Areas with considerably more complex geology and with a detailed (1:5000) scale digital soil mapping show decreases in yield with more accuracy as observed by Bazaglia Filho et al. (2013). The authors determined a 35 % agreement level between a digital map and a traditional one in an area of 182 ha with seven geology materials and 19 soil classes.

Among the factors that could be listed as causes of errors in digital mapping of soils are: soil variation due to factors other than those not measured by the environmental data collected (Chagas et al., 2010) and the challenge of learning by the neural networks which may wrongly assign mapping units due to the similarity in environmental variables collected (Zhu, 2000).

Conclusions

The elevation and slope showed the most contrasting behavior among the variables measured, and therefore contributed more to making distinctions between soil classes. The variable geomorphic surface showed potential for being used in digital mapping because it incorporates the soil forming "time" factor.

The similar behavior of environmental variables in Latosol and Argisol areas was a challenge in terms of distinguishing between these two soil classes by DSM. The same challenge was also noted in the fieldwork, due to the similarity in landscape where these two units were found.

The same soil distribution pattern observed in the reference area mapped by conventional method was also observed in the DSM. This study reported on digital techniques that have the potential to match the natural soil distribution in the landscape.

The application of the knowledge gained by the use of neural networks from a reference area to a nearby area and the use of environmental variables that express the soil landscape relationship were in agreement with the field mapping. Using this methodology, soil units were mapped in uncharted areas and provided preliminary information. This approach is interesting because, under certain conditions, it allows for less costly and faster soil mapping .

This type of approach is most important for areas where the geological configuration is similar. The proposed mapping system will be successful as long as the configuration of the greater area is similar to the smaller area. Once this similarity begins to disappear, and there is a different configuration, a new reference area should be provided to train a new ANN.

Acknowledgments

References

Bazaglia Filho, O.; Rizzo, R.; Lepsch, I.F.; Prado, H.; Gomes, F.H.; Mazza, J.A.; Demattê, J.A.M. 2013. Comparison between detailed digital and conventional soil maps of an area with complex geology. Revista Brasileira de Ciência do Solo 37: 1136-1148.

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

Boruvka, L.; Penizek, V. 2007. A test of an artificial neural network allocation procedure using the Czech soil survey of agricultural land data. p. 415-424. In: Lagacherie, P.; Mcbratney, A.B.; Voltz, M., eds. Digital soil mapping: an introductory perspective. Developments in Soil Science. Elsevier, Amsterdam, Netherlands.

Chagas, C.S.; Fernandes Filho, E.I.; 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).

Chagas, C. S.; Fernandes Filho, E. I.; Vieira, C. A. O. 2013. Comparison between artificial neural networks and maximum likelihood classification in digital soil mapping. Revista Brasileira de Ciência do Solo 37: 339-351.

Congalton, R.G.; Green, K. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers, New York, NY, USA.

Daniels, R.B.; Gamble, E.F.; Cady, J.G. 1971. The relation between geomorphology and soil morphology and genesis. Advances in Agronomy 23: 51-87.

Demattê, J.A.M.; Terra, F.S. 2014. Spectral pedology: a new perspective on evaluation of soils along pedogenetic alterations. Geoderma 217-218: 190-200.

Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. 2013. Brazilian System of Soil Classification = Sistema Brasileiro de Classificação de Solos. 3ed. Brasília, DF: Embrapa Solos. (in Portuguese).

Favrot, J.C. 1989. A strategy for large scale soil mapping: the reference areas method. Science du Sol 27: 351-368 (in French, with abstract in English).

Foody, G.M.; Arora, M.K. 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. International Journal of Remote Sensing 18: 799-810.

Gallant, J.C.; Wilson, J.P. 2000. Primary topographic attributes. p. 51-85. In: Wilson, J.P.; Gallant, J.C., eds. Terrain analysis: principles and applications. John Wiley, New York, NY, USA.

Grinand, C.; Arrouays, D.; Laroche, B.; Martin, M.P. 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143: 180-190.

Hirose, Y.; Yamashita, K.; Hijiya, S. 1991. Back-propagation algorithm which varies the number of hidden units. Neural Networks 4: 61-66.

Instituto de Pesquisas Tecnológicas [IPT]. 1981. Geological Map of the State of São Paulo. Escale - 1:500.000. = Mapa Geológico do Estado de São Paulo. Escala - 1:500.000. IPT, São Paulo, SP, Brazil. p. 126. (in Portuguese).

Kempen, B; Brus, D.J.; Heuvelink, G.B.M.; Stoorvogel, J.J. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma 151: 311-326.

Lagacherie, P.; Legros, J.P.; Burrough, P.A. 1995. A soil survey procedure using the knowledge on soil pattern of a previously mapped reference area. Geoderma 65: 283-301.

Ruhe, R.V. 1969. Quaternary Landscape in Iowa. Iowa State University Press, Ames, IA, USA.

Soil Survey Staff. 2014. Keys to Soil Taxonomy. 12ed. USDA-Natural Resources Conservation Service, Washington, DC, USA.

Thomas, A.L.; King, D.; Dambrine, E.; Couturier, A. 1999. Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges Mountains (Northeastern France). Geoderma 90: 291-305.

Vasques, G.M.; Demattê, J.A.M.; Viscarra Rossel, R.A.; Ramírez López, L.; Terra, F.S.; Rizzo, R.; Souza Filho, C.R. 2015. Integrating geospatial and multi-depth laboratory spectral data for mapping soil classes in a geologically complex area in southeastern Brazil. European Journal of Soil Science 66: 767-779.

Yigini, Y.; Panagos, P. 2014. Reference area method for mapping soil organic carbon content at regional scale. Procedia Earth and Planetary Science 10: 330-338.

Zell, A.; Mamier, G.; Vogt, M.; Mache, N.; Hübner, R.; Döring, S.; Herrmann, K.; Soyez, T.; Schmalzl, M.; Sommer, T.; Hatzigeorgiou, A.; Posselt, D.; Schreiner, T.; Kett, B.; Clemente, G.; Wieland, J.; Gatter, J. 1996. Stuttgart Neural Network Simulator v 4.2. University of Stuttgart, Stuttgart, Germany.

Zhu, A.X. 2000. Mapping soil landscape as spatial continua: the neural network approach. Water Resources Research 36: 663-677.