Larger geologic complexity implies larger uncertainty

Alessandro Samuel-Rosa[1], Ricardo Simão Diniz Dalmolin[2] and Pablo Miguel[3]

[1] Curso de Pós-Graduação em Agronomia-Ciência do Solo da Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, Brazil. E-mail: alessandrosamuel@yahoo.com.br.
[2] Departamento de Solos da Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brazil. E-mail: dalmolinrsd@gmail.com.
[3] Programa de Pós-Graduação em Ciência do Solo da Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brazil. E-mail: tchemiguel@yahoo.com.br.

Abstract

The prediction of soil attributes through predictive models became popular in the last decade. Land-surface parameters are the ones used at the largest extend as predictor variables. However, the relation between land-surface parameters and soil attributes is not evident in all land surfaces: it might be strong in one but weak in another. Besides, this relation usually varies across the landscape due to the other soil forming factors influence such as the parent material. Using the results of a study carried out in a geologically complex small catchment in Southern Brazil we show that soil particle-size distribution can be estimated from land-surface parameters. The prediction functions can explain more than half the overall data variance. The literature shows that such performance is superior to that of the conventional soil mapping approach. However, when we evaluate the spatial accuracy of the predictive models we see that in locations of larger geological heterogeneity the predictive models present large prediction errors. Thus, larger geological heterogeneity implies larger uncertainty. Unfortunately the geological information available is not sufficiently accurate to help improving significantly the spatial accuracy of the prediction functions.

Keywords: predictive models, linear regression models, geology, soil information user

1. Introduction

The concept of geologic complexity describes the distribution of faults and lithological boundaries as a function of scale (Hodkiewicz, 2003 apud Ford and Blenkinsop, 2008). Such feature is a result of the combined action of several different geologic actors (tectonic setting, deformation and igneous events and sedimentation patterns (Ford and McCuaig, 2010)). The larger the number of geologic actors playing in a given area, the more complex will be the distribution of geologic units (Ford and McCuaig, 2010).
Geologic complexity has been used a a 'positive' proxy to indicate the occurence of ore deposits (Gonzato et al., 1998; Weinberg et al., 2004; Hodkiewicz et al., 2005; Ford and Blenkinsop, 2008; Ford and McCuaig, 2010). However, geologic complexity is also a 'negative' proxy since it is directly related to the occurence of uncertainty in a geologic map. Such uncertainty arises from the fact that geologic

complexity affects our ability to accurately identify and understand the geologic features of a given area (Keefer, 2007). Again, the larger the geologic variability, the larger the geologic complexity (Bardossy and Fodor, 2001 apud Keefer, 2007) and thus the uncertainty in a geologic map (Keefer, 2007).
In circumstances where geology plays an important role on the spatial distribution os soil attributes, we expect the geologic complexity to affect the performance of prediction functions used to map soil attributes. We believe that raw geologic maps are not be the best alternative to be used as predictor variables in these cases. Therefore, our objective is to analyze the effect of geologic complexity on the accuracy of predictive models used to map soil attributes and, in the near future, test alternative proxies related to the geologic complexity to build the predictive models.

2. Methodology

The study was carried out in a small catchment (18.92 km^2) located in Santa Maria, Southern Brazil. Climate is classified as Cfa (humid subtropical climate without a defined dry season), with mean annual temperature of 19.2 °C and mean annual precipitation of 1708 mm well-distributed along the year (Maluf, 2000). The relief varies from plan to mountainous and elevations range from 147 to 478 m (Figure 1).

a) b)

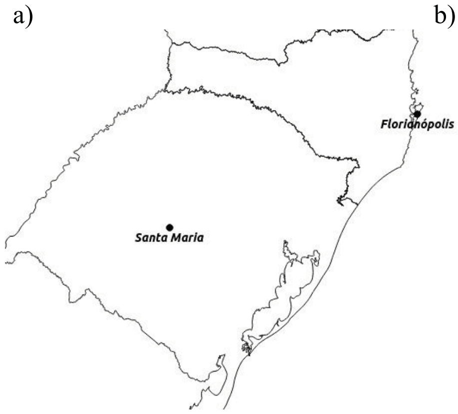

Figure 1. Location of the study area in the Brazilian territory (a) and its topographic features and the water reservoir (b).

The geology is complex (Maciel Filho, 1990; Sartori, 2009): in elevations above ±350 meters occurs the Superior Sequence of the Serra Geral Formation (igneous rocks – rhyolite-rhyodacite) and in elevations between ±200 and ±350 meters occurs the Inferior Sequence of the Serra Geral Formation (igneous rocks – basalt-andesite) and in its interior or bellow it the Botucatu Formation (sedimentary rocks - aeolian sandstone). In elevations bellow ±200 meters occurs the Caturrita Formation (sedimentary rocks – fluvial sandstone). Colluvial deposits (material from the Serra Geral and Botucatu Formations) occur in elevations between ±250 and ±300 meters. Colluvial deposit (material from the Caturrita and Botucatu Formations) can also be found in elevations ±200 meters. Recent fluvial deposits are common close to the drainage channels.

The best available geologic map is published at a scale of 1:25.000 and was prepared by Maciel.Filho (1990) using the topographic maps of the Brazilian Army's Geographic Service (DSG) published at a scale of 1:25.000. Therefore, the boundaries of the geologic units were drawn following the contour lines of the topographic maps. Many recent sedimentary bodies are not depicted in the geologic maps such as fluvial and colluvial deposits.

Areas of larger geologic complexity are those occurring in mid-slope inclined positions and comprise the Inferior Sequence of the Serra Geral Formation and the Botucatu Formation. Areas of small geologic complexity are those occurring in higher and lower positions with gently sloping relief and comprise the Superior Sequence of the Serra Geral Formation and the Caturrita Formation.

A set of 339 soil samples were used to build the models to predict the clay fraction of the soils (0 to 20 cm or the whole A horizon when the soils was shallower than 20 cm) were collected in earlier studies during the soil and land use survey of the area using expert knowledge (Miguel, 2010; Samuel-Rosa et al., 2011). The proportion of clay-size fraction in the soils was determined by the pipette method after removal of the organic matter with H_2O_2 (30 % v/v) in those samples containing more than 5% of organic matter (Embrapa, 1997) . In each of the 339 the parent material of the soils was identified.

Multiple linear regression models were built using land-surface parameters (elevation, slope, catchment area, convergence index, wetness index were derived from a 10-m resolution DEM using SAGA GIS) to estimate the particle-size distribution of soils. Before building models the three size fractions (sand, silt and clay) were transformed to additive log-ratios according to Aitchison (1986). Model validation was performed through repeated (100 times) 10-fold cross-validation. All statistical analysis were carried out in R (R Development Core Team, 2011).

3. Preliminary results

The predictive models built to predict the particle-size distribution of the soils are described bellow. The land-surface parameter ELEV explains the largest proportion of the variance (60%). Such predictive variable reproduces the parent material influence on soil particle-size distribution, since igneous rocks prevail in upslope and sedimentary rocks in downslope positions. Soils derived from igneous rocks posses a finer particle-size distribution. Since ELEV explains the largest part of the data variance the influence of the parent material is stronger than that of the topography. Such close relation between ELEV and parent material was expected since previous models used ELEV as a predictor of geology (Maciel Filho, 1990).

$$\ln(clay/sand) = -1.0027474 + 0.0081838\ ELEV - 0.0413414\ SLOPE - 0.9702653\ WET - 0.0107230\ CONV$$

$$\ln(silt/sand) = -0.9136230 + 0.0090393\ ELEV - 0.0278960\ SLOPE - 0.0151896\ CONV - 0.9682048\ WET$$

Then why not use the geologic map as a predictor variable? Table 1 bellow shows the comfusion matrix calculated using all the 339 observations made in the study area. Reference values indicate the information observed in the field while prediction values are those as depicted on the geologic map of Maciel Filho (1990). The results are very satisfactory with an accuracy of 0.85 and a kappa index of 0.70. Such values of accuracy and kappa index have been established as the primary goal of many soil mapping projects (ten Caten, 2011).

Table 1. Confusion matrix for the whole study area.

Prediction	Reference		User accuracy
	Sedimentary	Volcanic	
Sedimentary	132	16	0.8000
Volcanic	33	158	0.9080
Mapper accuracy	0.8919	0.8272	
Accuracy	0.8555 (95% CI: 0.8134 – 0.8911)		
NIR	0.5133 (P-Value [Acc > NIR]: < 2e-16)		
Kappa	0.7099 (Mcnemar's Test P-Value: 0.02227)		

However, when we look at table 2, we are able to see that the accuracy of the geologic map of Maciel Filho (1990) is not as accurate as suggested Table 1. In the areas of small geologic complexity all field observations match with the geologic map, giving an accuracy estimate of 100%. On the other hand, in areas of large geologic complexity, the accuracy estimate is only 63%, yielding a kappa index of 0.27. Note that the accuracy estimate is very close to the no-information rate estimate, suggesting that the information content of the geologic map in such areas is very small. The use of such information to build predictive models shall be avoided since it can introduce large errors in the predictions.

Table 2. Confusion matrix for areas of small and large geologic complexity. Reference values were observed in the field and prediction values are as depicted on the geologic map.

Prediction	Small complexity areas			Large complexity areas		
	Reference		User accuracy	Reference		User accuracy
	Sedimentary	Volcanic		Sedimentary	Volcanic	
Sedimentary	95	0	1.0000	37	16	0.5286
Volcanic	0	109	1.0000	33	49	0.7538
Mapper accuracy	1.0000	1.0000		0.6981	0.5976	
Accuracy	1.0000 (95% CI: 0.9821 – 1.0000)			0.6370 (95% CI: 0.5499 – 0.7180)		
NIR	0.5343 (P-Value [Acc > NIR]: < 2.2e-16)			0.5185 (P-Value [Acc > NIR]: 0.003605)		
Kappa	1.0000 (Mcnemar's Test P-Value: NA)			0.2798 (Mcnemar's Test P-Value: 0.022271)		

NIR – no information rate.

When we calculate some summary statistics of the residuals of the particle-size distribution predicted in the study area we are able to draw the effect of geologic complexity on the uncertainty about the predictions. Table 3 shows the descriptive statistics of the absolute residuals. Since the clay content and range in the soils is reduced the prediction residuals are also small and orbitate around 5% in both small and large geologic complexity areas. However, the prediction residuals of the silt and sand content in areas of large geologic complexity is two times larger than in areas of small geologic complexity.

Table 3. Summary statistics of absolute prediction residuals in areas of smal and large geologic complexity.

Geologic complexity	Clay (%)		Silt (%)		Sand (%)	
	Mean	SE	Mean	SE	Mean	SE
Small	5.20	0.31	6.79	0.37	7.60	0.44
Large	5.19	0.35	12.55	0.67	15.80	0.91

Number of observations: small geologic complexity = 204 observations; large geologic complexity = 135 observations.

4. Conclusion

Our preliminary results show that particle-size distribution can be estimated from land-surface parameters using linear regression models. However, when we evaluate the residuals of the predictions in locations of larger geological complexity the predictive models have a poor performance. As a general rule, larger geologic complexity implies larger uncertainty in the predictive models built to predict the particle-size distribution of the topsoil. Unfortunately the geological information available is not sufficiently accurate to help improving significantly the accuracy of the predictive models. Neither there are environmental variables available to allow using the geologic information in the predictive models. Therefore, next step of the study will account for the quantification of the geologic complexity which can be done using fractal analysis. Besides, new environmental variables will be developed to be used as explanatory variables taking into the predictive model the effects of geologic complexity.

Acknowledgments

We want to thank the CNPq and CAPES.

References

Aitchison, J. (1986), *The statistical analysis of compositional data*. London: Chapman & Hall. 416p.

Bardossy, G., Fodor, J. (2001), Traditional and new ways to handle uncertainty in geology. *Natural Resources Research*, 10:179-187.

Empresa Brasileira de Pesquisa Agropecuária – Embrapa (1997), *Handbook of methods of soil analysis*. 2.ed. Rio de Janeiro, Centro Nacional de Pesquisa de Solos. 212p.

Ford, A., Blenkinsop, T.G. (2008), Evaluating geological complexity and complexity gradients as controls on copper mineralisation, Mt Isa Inlier. *Australian Journal of Earth Sciences*, 55:13-23.

Ford, A., McCuaig, T. (2010), The effect of map scale on geological complexity for computer-aided exploration targeting. *Ore Geology Reviews*, 38:156-167.

Gonzato, G., Mulargia, F., Marzocchi, W. (1998), Practical application of fractal analysis: problems and solutions. *Geophysical Journal International*, 132:275-282.

Hodkiewicz, P. (2003), The interplay between physical and chemical processes in the formation of world-class orogenic gold deposits in the Eastern Goldfields Province, Western Australia. PhD thesis, University of Western Australia, Austrália.

Hodkiewicz, P.F., Weinberg, R.F., Gardoll, S.J., Groves, D.I. (2005), Complexity gradients in the Yilgarn Craton: fundamental controls on crustal-scale fluid flow and the formation of world-class orogenic-gold deposits. *Australian Journal of Earth Sciences*, 52:831-841.

Keefer, D.A. (2007), *A framework and methods for characterizing uncertainty in geologic maps.* Proceedings of the 3D Workshop, University of Illinois. 5p.

Maciel Filho, C.L. (1990), Carta geotécnica de Santa Maria. 21p.

Maluf, J. (2000), A new climatic classification for the state of Rio Grande do Sul, Brazil. *Revista Brasileira de Agrometeorologia*, 8:141-150.

Miguel, P. (2010), *Pedological characterization, land use and modeling of the soil loss in hillslope areas the Plateau Border of RS*. Master's thesis in soil science. Santa Maria, Universidade Federal de Santa Maria. 112p.

R Development Core Team (2011), *R: A language and environment for statistical computing*.

Samuel-Rosa, A., Miguel, P., Dalmolin, R.S.D., Pedron, F.A. (2011), Uso da terra no Rebordo do Planalto do Rio Grande do Sul. *Ciência e Natura*, 33:161-173.

Sartori, P. (2009), Geologia e geomorfologia de Santa Maria. Ciência e Ambiente, 38:19-42.

ten Caten, A., Dalmolin, R.S.D., Pedron, F.A., Mendonça-Santos, M.L. (2011), Multiple Logistic Regressions: controlling factors in applications to soil class prediction. *Revista Brasileira de Ciência do Solo*, 35:53-62.

Weinberg, R.F., Hodkiewicz, P.F., Groves, D.I. (2004), What controls gold distribution in Archean terranes? *Geology*, 32:545-548.