XXXIV CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO
28 de julho a 2 de agosto de 2013 | Costão do Santinho Resort | Florianópolis | SC

Ordinary kriging and regression kriging for predicting soil physical-chemical properties in two watersheds at Minas Gerais State[1].

Michele Duarte de Menezes[2]; Phillip R. Owens[3]; Carlos Rogério de Mello[4]; Nilton Curi[4].

[1] Trabalho executado com recursos do CNPq.
[2] Professora; Universidade Federal Rural do Rio de Janeiro; Seropédica, Rio de Janeiro; michele_duarte@ig.com.br; [3] Professor; Purdue Univeristy; [4]Professores, Universidade Federal de Lavras.

ABSTRACT: The estimation of soil properties at non-sampled areas can be source of valuable information for land management. This study compared the performance of ordinary kriging (OK) and regression kriging (RK) for predicting soil physical-chemical properties in two watersheds with contrasting soil-landscape features. Mean prediction of error and root mean square of prediction error were used for assessing the prediction methods. A better linear correlation was found at Lavrinha Creek Watershed, suggesting some relationship between contemporaneous landforms and soil properties, and RK outperformed OK. In the most cases, RK were not performed at Marcela Creek Watershed due to lack of linear correlation between covariates and soil properties. Since alternatives of simple OK kriging have been sought, another prediction method should be tested, considering the systematic pattern of soil properties distribution over that landscape.

Index terms: spatial variability, geostatistics, environmental covariates.

INTRODUCTION

The estimation of soil physical-chemical properties at non-sampled areas can be source of valuable information for land management and water yield. Different interpolation techniques have been used, with varying degrees of success, and improved in order to create more accurate soil property maps.

OK depends on a weighting scheme dictated by the variogram, where closer sample locations have greater impact on the final prediction (Bishop & McBratney, 2001). As long as OK uses only observed data to map unsampled areas, more recent innovations have been preferred, such as hybrid geostatistical procedures. This technique accounts for environmental correlation, and has become increasingly used in recent years because they allow utilizing available secondary information and often result in more accurate local predictions (Goovaerts, 1999). One example is the regression kriging, in which the interpolation is not only based on observed data, but also regression of the target variable on spatially exhaustive auxiliary variables (Hengl et al., 2007). The auxiliary variables or environmental covariates use concepts of soil forming factors equation in scorpan's "Jenny-like" model, where s: soil, other properties of the soil at a point; c: climate, climatic properties of the environment at a point; o: organisms, vegetation or fauna or human activity; r: topography, landscape attributes; p: parent material, lithology; a: age, the time factor; n: space, spatial position (McBratney et al., 2003).

Landscape features, such as land use, topography, and parent material, are known to control different soil processes and the spatial distribution of soil properties. Thus, the selection of interpolation method can vary between contrasting landscapes, even for the same soil property. In this study, two watersheds located at different geomorphological units were studied, because pedogenesis has a strong dependence on geomorphic systems. Thereby, the objective of this study was a comparison of OK and RK for predicting soil physical-chemical properties in the State of Minas Gerais, Brazil.

MATERIAL AND METHODS

The watersheds are located in different physiographical regions: Mantiqueira Range region (LCW) and Campos das Vertentes region (MCW). Additional characteristics of study sites are listed in the **Table 1**.

Table 1. Basic characteristics of the study sites

	Lavrinha Creek Watershed	Marcela Creek Watershed
Area	676 ha	470 ha
Elevation	1151 - 1780 m	957 - 1057 m
Mean annual temperature	15°C	19.7°C
Annual Precipitation	2,000 mm	1,300 mm
Native forest	Atlantic Forest	Cerrado

The parent material at LCW is gneiss, whose alteration resulted in predominance of Cambisols. MCW is located in Campos das Vertentes Plateau geomorphological unit. The relief is represented by gentle slopes with intense soil development, where Latosols are the most expressive soil classes

The topsoil (0-15 cm) was sampled at both sites. A total of 198 points were sampled at LCW, following the regular grids 300 x 300 m and refined scale 60 x 60 m and 20 x 20 m, and two transects with the distance of 20 m between points. A total of 165 points were sampled at MCW, following the regular grids 240 x 240 m and refined scale 60 x 60 mSoil properties determined were bulk density; organic matter; saturated hydraulic conductivity (determined *in situ* by constant flow permeameter - Ghelph permeameter - model 2800K).

Terrain attributes were based on a 30 m resolution DEM, generated from counter lines freely available in Brazil (IBGE, 1973) with 1:50,000 scale. The hydrologic consistent DEM was created. The following terrain attributes were calculated: slope, profile curvature, plan curvature, and SAGA wetness index. SAGA GIS 2.0.6, ArcGIS spatial analyst, ArcSIE 9.2.402 (ArcGIS's extention) were used. The normalized vegetation index (NDVI) from LANDSAT is a difference ratio model of the near infrared (NIR) and red bands of multispectral image (NDVI = (NIR band - Red) / (NIR band + Red).

The first step in ordinary kriging is to calculate the experimental semivariogram using the equation: $\gamma^*(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[z(x_i) - z(x_i+h)]^2$, where $\gamma^*(h)$ is the estimated value of the semivariance for lag h; $N(h)$ is the number of experimental pairs separated by vector h; $z(x_i)$ and $z(x_i+h)$ are values of variable z at x_i and x_i+h, respectively; xi and x_i+h are position in two dimensions.

The regression kriging combines multiple linear regression and ordinary kriging (Bishop & McBratney, 2001). The final maps are an additive combination of both models in a regression kriging approach. The normal distribution is an ideal requirement for linear regression. Thus, the non-normal distributed data were log transformed. The kriging and statistical analysis were carried out in statistical software R (R Development Core Team, 2010).

Of the total number of soils sampled, 25 points were used as validation points at LCW and 20 points at MCW. Two indices were calculated from the observed and predicted values. The mean prediction of error (MPE) was calculated by comparing estimated values ($\hat{z}(s_j)$) with the validation points ($z^*(s_j)$): $MPE = \frac{1}{l}\cdot\sum_{j=1}^{l}[\hat{z}(s_j) - z^*(s_j)]$, and the root mean square prediction error (RMSPE):

$$\mathrm{RMSPE} = \sqrt{\frac{1}{l}\sum_{j=1}^{l}[\hat{z}(s_j) - z^*(s_j)]^2}$$

where l is the number of validation points. The relative improvement (RI) of RK over OK was assessed by using RMSPE:

$$\mathrm{RI} = \frac{RMSPE_{ok} - RMSPE_{RK}}{RMSPE_{ok}} * 100$$

RESULTS AND DISCUSSION

The parameters used to fit semivariograms of target (OK) and residuals of multiple linear regression are presented in **Table 1**. LCW showed stronger strength of spatial structure than MCW. As long as the nugget effect is high at MCW, there may be undesirable large estimates of variances, providing less smoother and reliable OK.

The stepwise linear regression models are presented in the **Table 2.** The correlations at LCW could be considered better. The steep relief at LCW, with predominance of erosional surfaces, where mostly Cambisols are formed and sediments have been deposited, suggest some relationship between contemporaneous landforms and soil properties. The low or lack of correlation at MCW is related to complexity of relationship between environment and soil variables. These relationships may be complex, unknown, often very noisy, and is not necessary linear, as assumed by the multiple regression (Hengl et al., 2004). The hypothesis is the pedogenesis processes are creating a systematic order or pattern of soil properties distribution. Latosols were formed in a preterit landscape, with distinct morphological features resulting from an environment of soil formation, which do not exist currently.

A key issue is whether the correlations could be used to improve the prediction performance of soil properties. The comparison of prediction methods are showed in **Table 3.** OK is known to be very sensitive to short-range variation (Laslett & McBratney, 1990), and the large RMSPE can be ascribed to this. In this study, if the predictive variables can explain even a small part of the variation in the target variable (higher values of R^2), the RK outperforms OK because it exploits the extra information. And also, the summation of kriged errors due to regression, lead to smoothing of the predicted values, hence the reduction of RMSPE (Odeh et al., 1994). According to Bishop & McBratney (2001), even when only a poorly correlated secondary attributes are available, the hybrid methods may still perform better than OK.

At MCW the prediction of organic matter and Ksat by RK were not performed, due the low coefficient of determination, because it results in pure kriging (no correlation).

Considering the improvement over the OK, the values of RI (%) showed that the prediction accuracy can be improved by incorporating ancillary variables into prediction. The **Figure 1** presents the OK and RK prediction maps, and **Figure 2** the land use map.

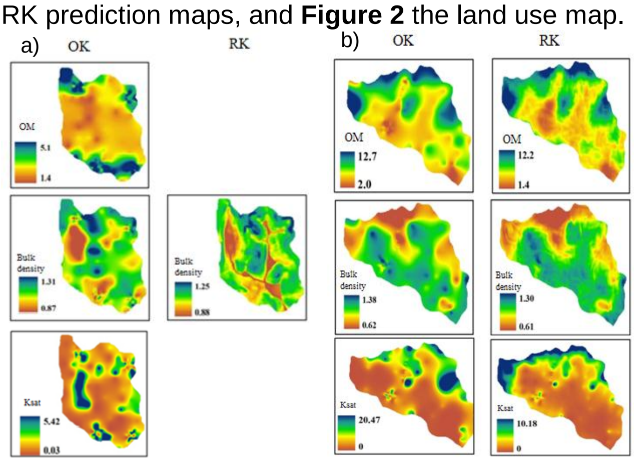

Figure 1 – Ordinary kriging (OK) and regression kriging (RK) prediction maps of soil physical-chemical properties at MCW (a) and LCW (b).

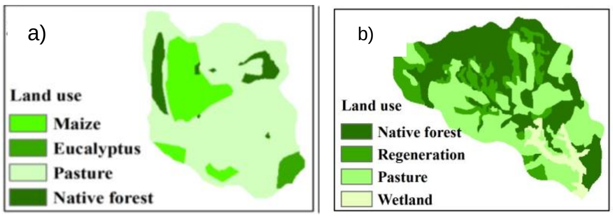

Figure 2 – Land use map at MCW (a) and LCW (b).

One limitation of OK is the exclusion of information on soil-landscape relationships. The maps show gradual transitions with fairly low level of detail. Another issue was the bulls-eyes features in some maps, probably due the short range of semivariogram and high nugget effect. The RK maps reflects changes of the terrain attributes and NDVI were kept on multiple regression. And also, the RK, which uses auxiliary variables, has a more smoothing effect on minimising the influence of outliers on prediction performance (Odeh et al., 1994).

At LCW, the relief seems to influence the forest cover indirectly, since pasture is preferably implanted in flatter and lower areas. The higher organic matter content detected at higher altitudes was probably due to lower temperatures. Vegetation distribution controls organic matter (Gessler et al., 2000), which in turn might explains the lower bulk density and higher Ksat in the same portions of the landscape (native forest or regeneration). At MCW were found higher values of organic matter in native forest and eucalyptus areas (east side). Lower values of organic matter were found under pasture areas, which is the predominant land use at this watershed. Latosols tend to have good physical properties influenced by aggregate stability. Nevertheless, would be expected higher values of Ksat in Latosols. But in accordance to the land use, lower values of Ksat were found due to the compaction by the cattle in pasture areas.

CONCLUSIONS

Better linear correlations were found at LCW, suggesting some relationship between contemporaneous landforms and soil properties, and RK outperformed OK.

Since alternatives of simple OK kriging have been sought, another prediction method should be tested at MCW, considering not only the linear relationships between covariates and soil properties, but also the systematic patter of soil properties distribution over that landscape.

Besides the relief and soil, the land use markedly explained the spatial variability soil properties at both watersheds.

REFERENCES

BISHOP, T. F. A. & McBRATNEY, A. B. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma, 103:149-160, 2001.

GESSLER, P. E.; CHADWICK, O.A.; CHAMRAN, F. et al. Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Science Society of American Journal, 64:2046-2056, 2000.

GOOVAERTS, P. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma, 89:1-45, 1999.

HENGL, T.; HEUVELINK, G.; ROSSITER, D.G. About regression-kriging: from equations to case studies. Computer and Geosciences, 33:1301-1315, 2007.

HENGL, T.; HEUVELINK, G.; STEIN, A. A. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, 122:75-93, 2004.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Carta do Brasil ao milionésimo. Rio de Janeiro, 1973.

LASLETT, G. M. & MCBRATNEY, A. B. Further comparison of spatial methods for predicting soil pH. Soil Science Society of American Journal, 54:1533-1558, 1990

McBRATNEY, A. B.; SANTOS, M. L. M.; MINASNY, B. On digital soil mapping. Geoderma, 117:3-52, 2003.

ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma, 63:197-214, 1994.

R DEVELOPMENT CORE TEAM. R: A Language for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2010

Table 1. Geostatistical parameters for the best-fitted semivariograms of the target variables and the residuals of multiple linear regression at LCW and MCW.

Soil property/1	Model	Target			Residuals			Nugget/sill/2	Strength of spatial structure/3
		Range (m)	Sill - Nugget	Nugget	Range (m)	Sill - Nugget	Nugget		
					LCW				
OM (%)	Gaussian	763,98	5,86	0,51	763,98	3,36	1,27	0,09	Strong
BD (g cm^{-3})	Exponential	623,63	0,06	0	374,19	0,029	0,012	0,00	Strong
logKsat (m day^{-1})	Exponential	436,5	0,33	0,15	218,27	0,24	0,15	0,45	Medium
					MCW				
OM (%)	Gaussian	311,80	0,52	0,39	279,6	0,49	0,39	0,75	Weak
BD (g cm^{-3})	Spherical	571,60	0,0044	0,0058	714,5	0,0038	0,0058	1,32	Weak
logKsat (m day^{-1})	Gaussian	332,60	0,11	0,20	291,04	0,09	0,21	1,82	Weak

/1 OM – organic matter; BD – bulk density; Ksat – saturated hydraulic conductivity;/2 Calculated from the target data set./3 Values <0,25 being strong, 0,25-0,75 being medium, and >0,75 being weak.

Table 2. Stepwise multiple linear regression models between soil physical-chemical properties and environmental covariates at Lavrinha Creek Watershed.

Soil Properties	Altitude	Slope	Wetness index	Plan curvature	Profile curvature	NDVI	R^2
				LCW			
OM (%)	0,010129	-	-	-0,115689	0,211202	4,734977	0,22
BD (g cm^{-3})	-0,001001	0,0028765	-	-	-0,0371671	-	0,20
Ksat (m day^{-1})	0,0039241	-	-	-	-	-	0,30
				MCW			
OM (%)	0,012492	-	-	-	-	-	0,05
BD (g cm^{-3})	-0,001071	-	-0,02264	-	-0,0141589	-	0,12
Ksat (m day^{-1})	0,006330	-	-	-	-	-	0,04

/1 OM – organic matter; BD – bulk density; Ksat – saturated hydraulic conductivity.

Table 3. Comparison of interpolation methods and relative improvement of physical-chemical properties at Lavrinha Creek Watershed.

Soil property	OK		RK		RI (%)
	MPE	RMSPE	MPE	RMSPE	
			LCW		
Organic matter (%)	-0,663	3,317	-0,482	2,408	27,4
Bulk density (g cm^{-3})	0,009	0,045	0,002	0,012	73,3
Ksat (m day^{-1})	0,439	2,196	0,242	1,208	45,0
			MCW		
Organic matter (%)	0,146	0,659	-	-	-
Bulk density (g cm^{-3})	-0,005	0,063	-0,015	0,023	63,5
Ksat (m day^{-1})	0,087	0,368	-	-	-

OK – ordinary kriging; RK – regression kriging; MPE: mean prediction of error; RMSPE: root mean square of prediction error; RI: relative improvement.