XXXV Congresso Brasileiro de Ciência do Solo
CENTRO DE CONVENÇÕES - NATAL / RN
O SOLO E SUAS MÚLTIPLAS FUNÇÕES
02 a 07 DE AGOSTO DE 2015

Knowledge-based digital soil mapping technique for predicting saturated hydraulic conductivity.

Michele Duarte de Menezes[(1)]; Phillip R. Owens[(2)] ; Carlos Rogério de Mello[(3)]; Sérgio Henrique Godinho Silva[(4)]; Nilton Curi[(3)].

[(1)]Professor; Federal University of Lavras; Lavras; Minas Gerais; michele.menezes@dcs.ufla.br; [(2)]Professor; Purdue University; [(3)]Professor, Federal University of Lavras; [(4)]Doctoral candidate, Federal University of Lavras.

ABSTRACT: The estimation of saturated hydraulic conductivity (Ksat) at non-sampled areas in a continuous way can be source of valuable information for land and water management. The objective of this work was assessing the spatial prediction of Ksat in two watersheds by means of knowledge-based digital soil mapping technique. The way of choose typical values as well as the role of land use as a covariate was tested. An independent validation data set was used for assessing the accuracy of predictions methods. The knowledge-based digital soil mapping technique is an accurate option for spatial prediction of Ksat considering: 1) non-linearity distribution of data, 2) the high cost of an intense sampling scheme; and 3) the scarce resources for that in Brazil. The use of a fuzzy membership maps as a guide for sampling Ksat at Lavrinha Creek Watershed (LCW) is an accurate alternative. At Marcela Creek Watershed (MCW) the mean of values inside the polygons provide better accuracy of spatial prediction, since the land use does not influence Ksat statistically.

Indexing terms: non-linear distribution; ANOVA test; fuzzy logics**.**

INTRODUCTION

The estimation of Ksat at non-sampled areas in a continuous way can be source of valuable information for land management, water yield, irrigation and drainage systems, and distributed hydrologic models. Ksat has been recognized for its high spatial variability (Moustafa, 2000), skewed frequency (Alvarenga et al., 2011) or non-normality of its distribution. Alson, the relationship between soil property and environmental covariates is often complex and non-linear (Zhu et al., 2010). Data normality can influence the estimation of certain spatial interpolation methods that assume the input data are distributed normally about the mean.

In the knowledge-based approach, the continuous variation of soils can be represented by continuous soil property maps derived from the similarity vectors and fuzzy logics (Zhu et al., 1997), and can be viewed as a non-linear transformation of environmental variables based on knowledge of soil-landscape relationships (Zhu et al., 2010). Besides to provide tools for the Pedologist to formalize the soil-landscape relationships, another advantage of this method is only one sampling point per soil type is required for the spatial prediction, which reduces the costs of mapping. Thus, the way of choose the typical value is a key point for the accuracy of spatial prediction.

This work aimed to assess the spatial prediction of Ksat, in two watersheds by means of knowledge-based digital soil mapping technique. The way of choose typical values as well as the role of land use as a covariate was tested.

MATERIAL AND METHODS

This study was conducted at MCW and LCW located in Alto Rio Grande Basin, Minas Gerais. Additional characteristics of study sites are listed in the **Table 1**.

Table 1 - Basic characteristics of the study sites.

	LCW	MCW
Area	676 m	470 ha
Elevation	1,151-1,780m	957-1,057m
Mean annual temperature	15˚C	19.7 ˚C
Annual precipitation	2,000mm	1,300mm
Native forest	Atlantic Forest	Cerrado

The Ksat was determined *in situ* by constant flow permeameter (Ghelph permeameter model 2800KI). This study departed from a total of 198 sampled points at LCW, following the regular grids of 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 m.

The knowledge about the soil-landscape relationships was qualitatively modeled using ArcSIE (Soil Inference Engine) version 9.2.402 (Shi et al., 2009). The Rule-Based Reasoning (RBR) inference was used to define the relationship between values of an environmental variables (relief and land use) and a given soil type. Digital elevation model with pixel of 10 m resolution were generated from a counter level in 1:50.000 (IBGE). The DEM derivatives (slope, altitude above the channel network, plan curvature, profile curvature and wetness index) were calculated using ArcGIS 10 (ESRI) and SAGA GIS 2.1.0.

Terrain attribute values and ranges associated with each soil map class were used to define membership or optimality functions (curves), which, in turn, define the relationship between the values of an environmental feature and soil type. The initial output from the inference process is a series of fuzzy membership maps in raster format, one for each soil type under consideration (Shi et al., 2009), representing similarities of each pixel in the landscape to the soil types. From these maps, two products can be generated: a hardened map that represents soil types or classes, and the spatially continuous soil property map (Ksat in this study).

The continuous variation of Ksat is derived from the similarity vectors, using the following formula (Zhu et al., 1997):

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} * V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}}$$

where V_{ij} is the estimated Ksat at location *(i,j)*, V^k is a typical value of soil type *k* (e.g. Udepts1), and *n* is the total number of prescribed soil classes for the area. The typical value consists of the central concept of soil property value for each soil type, which is generally a soil profile into a polygon from conventional soil survey. In this study, the data base does not coming from a conventional soil survey, thus it was tested the way of choose typical values, as follow:

a) mean Ksat value into each polygon of a soil type from the hardened map. The data set prediction was plotted in the soil type hardened map, and the mean value was calculated for each soil type, and then, used as V^k; b) mean Ksat value into each polygon that results from the soil type hardened map overlaid in land use raster map, if the ANOVA test shows that Ksat influence land use; c) point geographically located on the pixel with highest membership value for the correspondent soil type; d) point geographically located on the pixel with highest membership value for the correspondent soil type, but overlaid in land use raster map, if the ANOVA test shows that land use influenced Ksat.

If the local soil formative environment characterized by a GIS resembles the environment of a given soil category (Ksat), then property values of the local soil should resemble the property values of the candidate soil type. The resemblance between the environment for soil at (*i,j*) and the environment for soil type *k* is expressed by S_{ij}^{k} , which is used as an index to measure the level of resemblance between the soil property values of the local soil and those of soil category (Zhu et al., 1997). The higher the membership of a local soil in a given soil type, the closer the Ksat at that location will be to the typical property values (Zhu et al., 2010).

In order to assess if Ksat is significantly influenced by different types of land use, and not only relief, analysis of variance (ANOVA) were made by F test ($p<0.01$ or $p<0.05$). The statistical analyses were performed in SAS version 9.2 (Statistical Analysis System Institute).

If ANOVA test pointed out the influence of land use, in such cases, the typical value V^k came from the combination of soil and land use *k* to form soil type, e.g. Udepts1 under pasture. In ArcSIE, land use raster map was used as categorical data (data do not have quantitative meaning, values are only for labeling or categorizing different land uses) and overlaid with all soil classes, using the function type Nominal (Shi et al., 2009).

Of the total number of soils sampled, 25 points were used only 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 ($z^*(s_j)$) of Ksat:

$$MPE = \frac{1}{l}.\sum_{j=1}^{l}[\hat{z}\,(s_j) - z^*(s_j)]$$

and the root mean square prediction error (RMSPE):

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

where *l* is the number of validation points. The MPE measures the bias of prediction, and RMSE measures the precision of prediction.

RESULTS AND DISCUSSION

Histograms allow us to picture the distribution to see how it lies about the mean or median and to identify extreme values. It is possible to notice the frequency distribution of Ksat is positively skewed at both watersheds (**Figure 1**), in such way that the mean is greater them median of the distribution (**Table 2**). Also, the variability of this attribute is high, according to the coefficient of variation. The Ksat in both cases has non-normal distribution.

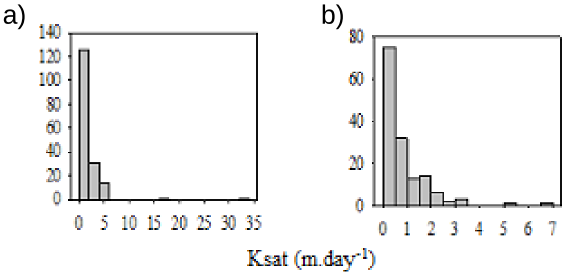

Figure 1 - Frequency histograms of LCW (a) and MCW (b).

According to **Table 3**, the variance between land uses was statistically significant only at LCW, which means the land use affected physical properties. In this case, different types of land use were joined, based on mean test for separation (**Table 4**). The values of Ksat encompass not only the topsoil properties, and therefore, it might be influenced by deeper pedological features more than the surface land use features at MCW. Oxisol is the main soil class at this watershed, which is recognized by its strong aggregate stability and the water moves through the soil readily.

At LCW, higher values of WI and low slopes were used for mapping hydromorphic soils in flatter alluvial areas (footslope). The Inceptisol occupies the well-drained portions of landscape with lower values of WI (summit, shoulder and backslope) formed by different combinations and ranges of slope, plan and profile curvature (Menezes et al., 2014). These soil-landscape relationships were converted into optmality curves, which the shapes are presented in the **Table 5**.

At MCW, the Acrudox usually occupy flatter summit positions in a more convex pedoform, expressed by higher values of AACHN, lower values of slope and negative values of plan curvature. The Hapludox is present on the shoulder, backslope and footslope positions (intermediate values of AACHN and gentle slopes). Two instances were applied for Inceptisol: one considering steeper slopes, and another for plan and profile curvatures. These instances were integrated using multiplication of functions. Hydromorphic soils (Endoaquent) are located in lower AACHN and higher values of WI. The shape of optimality curves are presented in the **Table 5**.

At both watersheds the pure nugget effect and low correlations between environmental covariates (digital terrain models and remote sensing NDVI transformation) and Ksat were found (Menezes, 2011). Thus, according to the Table 6, the validation indexes showed that knowledge-based digital soil mapping is an alternative with high accuracy to deal with highly-skewed data, since de MPE and RMSPE showed values closer to zero.

Among the methods, at LCW the use of a fuzzy membership map as a guide for choose typical value showed higher accuracy. The fuzzy membership values represent the similarities of each pixel in the landscape to the soil types. The higher the values, the closer are the central concept according to optimality curves stablished. At MCW the mean values within the polygons provided better accuracy. In this case, more points were necessary to the spatial prediction.

This method has an advantage of incorporating the qualitative knowledge of the pedologist into spatial prediction, and maps reflect it (**Figure 2**), bringing more details then polygon-based maps (Menezes et al., 2014). The ranges of predicted properties in the maps are somewhat different from the full data set (**Table 2**) due to only one typical value per soil type was used. But, in this study, some extreme values might be considered outliers.

CONCLUSIONS

The knowledge-based digital soil mapping is an accurate option for spatial prediction of Ksat considering: 1) non-linearity distribution of data, 2) the high cost of an intense sampling scheme; and 3) the scarce resources for that in Brazil. The use of a fuzzy membership maps as a guide for sampling Ksat at LCW is an accurate alternative. The mean of values inside the polygons provide better accuracy of spatial prediction at MCW, since the land use does not influence Ksat statistically.

ACKNOWLEDGEMENTS

The authors thank FAPEMIG, CNPq and CAPES.

REFERENCES

ALVARENGA, C.C.; MELLO, C.R.; MELLO, J.M. & VIOLA, M.R. Continuidade espacial da condutividade hidráulica saturada do solo na bacia hidrográfica do Alto Rio Grande, MG. Bras. Ci. Solo, 35:1745-1757, 2011.

MOUSTAFA, M. M. A geostatistical approach to optimize the determination of saturated hydraulic conductivity for large-scale subsurface drainage design in Egypt. Agricultural Water Management, 42: 291-312, 2000.

MENEZES, M.D.; SILVA, S.H.G.S.; MELLO, C.R.; OWENS, P.R.; CURI, N. Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches. Sci. Agric., 71:316-323, 2014.

MENEZES, M.D. Levantamento pedológico de hortos florestais e mapeamento digital de atributos físicos do solo para estudos hidrológicos. Tese (doutorado) – Universidade Federal de Lavras, 2011. 225p.

ZHU, A. X.; BAND, L.; VERTESSY, R.; DUTTON, B. Derivation of soil properties using a soil land inference model (SoLIM). Soil Science Society of American Journal, 61: 523-533, 1997.

ZHU, A.; QI, F.; MOORE, A.; BURT, J.E. Prediction of soil properties using fuzzy membership values. Geoderma, 158: 199-206, 2010.

SHI, X.; LONG, R.; DEKETT, R.; PHILIPPE, J. Integrating different types of knowledge for digital soil mapping. Soil Science Society of America Journal, 73: 1682-1692, 2009.

Table 2 - Descriptive statistics of saturated hydraulic conductivity of the full and validation dataset.

Ksat ($m\ day^{-1}$)	LCW						MCW					
	Mean	Median	STD	CV (%)	Min	Max	Mean	Median	STD	CV (%)	Min	Max
	1.65	0.95	3.02	183	0	32.35	0.80	0.46	0.93	117	0.03	6.99

STD – standard deviation; CV – coefficient of variation; Min – minimum; Max – maximum.

Table 3 - Summary of ANOVA performed to test the significance of land use effects on the variance of saturated hydraulic conductivity.

Watershed	Source of variance	df	Sum of squares	Mean square	F value
LCW	Land use effect	3	97.437301	32.479100	3.67**
	Residual	192	1696.962386	8.838346	
MCW	Land use effect	3	0.82612445	0.27537482	1.03
	Residual	1158	42.18503751	0.26699391	

** Significant at the 0.01 level.

Table 4 - Statistics of hydraulic saturated conductivity.

Land use	Mean (LCW)	Standard deviation (LCW)	Land use	Mean (MCW)	Standard deviation (MCW)
Native forest	2.53a	4.15	Native forest	0.31a	2.87
Natural regeneration	0.98b	0.94	Pasture	0.41a	3.33
Pasture	1.20b	2.32	Maize	0.48a	3.47
Wetland	0.76ab	0.76	Eucalyptus	0.64a	2.81

Means followed by the same letter do not differ significantly (p<0.05).

Table 5 - Optimality curves shapes for each soil type from digital terrain models.

Soil type	LCW					Soil type	MCW				
	Altitude	Slope	WI	Plan curv.	Profile Curv.		AACHN	Slope	WI	Plan curv.	Profile curv.
Fluvents	Z	Z	Bell	-	-	Endoaquent	Z	-	Bell	Bell	-
Udepts1	-	Bell	Bell	Bell	Bell	Acrudox	Bell	Bell	-	-	-
Udepts2	-	Bell	Bell	Bell	Bell	Dystrudept	Bell	S	-	-	Bell
Udepts3	-	Bell	Bell	Bell	Bell	Hapludox	Bell	Bell	-	-	-
Udepts4	-	Bell	Bell	Bell	Bell	Hapludox	Bell	Bell	-	-	-

Table 6 - Comparison of prediction methods.

Prediction method		LCW	MCW
Mean	MPE	0.385	-0.013
	RMSPE	1.932	0.057
Mean and land use	MPE	0.387	-
	RMSPE	1.933	-
Highest fuzzy membership	MPE	0.257	0.442
	RMSPE	1.283	1.876
Highest fuzzy membership and land use	MPE	0.299	-
	RMSPE	1.147	-

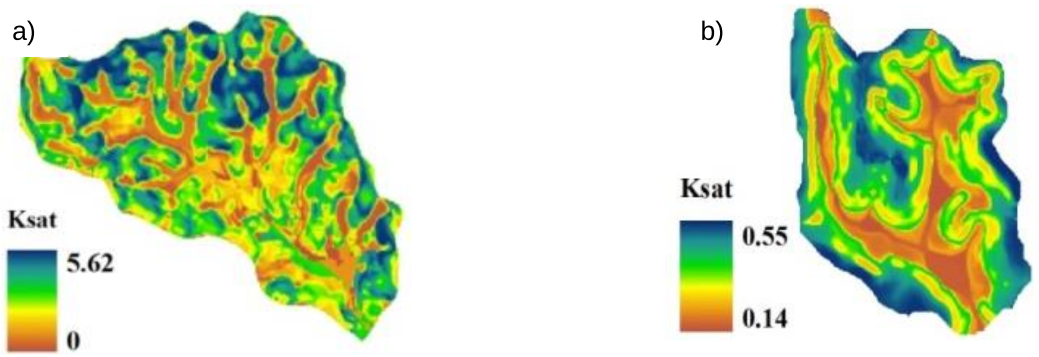

Figure 2 - Ksat prediction maps at LCW (a) and MCW (b).