Land-Surface Segmentation as sampling framework for soil mapping

Lucian Drăguţ
Department of Geography
West University of Timişoara
Timişoara, Romania
lucian.dragut@cbg.uvt.ro

Andrei Dornik
Department of Geography
West University of Timişoara
Timişoara, Romania
dornik_andrey@yahoo.com

Abstract—**Current sampling methods require a large number of samples to account for spatial variation of environmental covariates, which often conflicts the available financial resources. Thus, efficient sampling strategies are desirable. The aim of this study was to evaluate the potential of land-surface segmentation in stratifying a landscape into homogeneous areas, which can be used as support in optimizing soil sampling. The experiments were carried out in a study area where soil samples were available. Land-surface variables were derived from DEMs and segmented with a multiresolution segmentation (MRS) algorithm, into objects considered as homogeneous soil-landscape divisions. Thus, one sample was randomly selected within each segment, based on which predictions of the A-horizon thickness and soil types, were made. Predictions based on the land-surface segmentation sampling schemes outperformed predictions based on simple random sampling and conditioned Latin hypercube, respectively.**

INTRODUCTION

Spatial resolutions of soil maps for about 70 % of the Earth's ice-free land surface are too low to help with practical land management [1]. Conventional survey methods involve too much resources to be cost-effective in high-resolution soil mapping. Digital Soil Mapping (DSM) is an appropriate framework for producing detailed soil maps based on quantitative relationships between soil properties or types and their 'environment' [2]. Efficient sampling designs play an important role in DSM [2], as they have a significant impact on the accuracy of the maps [3].

Classical sampling methods (e.g. simple random sampling, systematic sampling and stratified sampling) as well as the model-based sampling strategy require a large number of samples to account for the spatial variation of environmental variables [4]. As sampling is constrained by financial resources, efficient sampling strategies are desirable [5]. Increasingly available geospatial information (e.g. satellite imagery, geology maps, Digital Elevation Models (DEMs) can be exploited as environmental covariates to optimize sampling locations [5] within the framework of a soil-landscape model [6]. However, sampling with support of environmental covariates has not been fully developed in DSM [5]. A number of recent papers [e.g. 4, 7, 8] demonstrated the value of purposive mapping based on such covariates in producing more accurate predictions by using fewer, but more representative samples.

Land-surface segmentation (LSS) is a relatively new technique to partition land-surface variables (LSVs) obtained from DEMs into contiguously homogeneous areas in multivariate feature space [9]. The most popular segmentation algorithm is Multiresolution Segmentation (MRS) as implemented in the eCognition® software [10]. The algorithm merges spatially contiguous pixels or cells into segments based on local homogeneity criteria [10]. The resulting land-surface objects incorporate scale, spatial autocorrelation, anisotropy and non-stationarity in their definition of homogeneity [11]. There have been only a few attempts to map soils based on LSS. The only approach of segmentation to optimize soil sampling [12] showed that a segmentation-based sampling (SBS) scheme produced better distribution of sampling locations over the area of interest, as compared to simple random sampling and regular sampling schemes.

It is clear that the potential of LSS to DSM has not been fully employed and the applicability of this technique to optimize soil sampling has only been touched upon. Therefore, we aimed at evaluating the potential of LSS in stratifying a landscape into homogeneous areas, which can be used as support in optimizing soil sampling.

METHODS

The experiments were carried out in the administrative territory of Branisca, which is located in western part of

Romania. The study area extends over 78 km^2 in a hilly region with altitudes between 165 and 670 m. Over 95 % of the area has slope gradients below 30 degrees. The bedrock consists of limestone, sandstone, igneous and sedimentary rocks. Three soil classes are widely developed, namely cambisols, luvisols and protisols. A 30 m SPOT-based DEM and a spatial database containing 93 records were available.

LSVs were derived from DEMs and segmented with a MRS algorithm, into objects considered as homogeneous soil-landscape divisions. Thus, one sample was randomly selected within each segment, based on which prediction of the dependent variables (thickness of A-horizon and soil classes) were made. Results were compared with predictions based on other sampling schemes, namely simple random sampling (SRS) and conditioned Latin hypercube (cLHS) [5]. The methodology is shown in Fig. 1.

Land-surface variables

Four LSVs, i.e. slope, plan curvature, profile curvature and topographic wetness index (TWI) were selected as potential soil covariates. For slope and curvatures, scale optimization was conducted according to [13]. Regressions between dependent variables and LSVs derived in increasing windows (using LandSerf) were conducted. The scales where the regression peaks emerged were retained for further analyses. The LSVs scaled as above and TWI were used for identification of the best predictors, with linear forward stepwise regression for thickness of A-horizon and logistic forward stepwise regression for soil class respectively.

Land-surface segmentation

For both dependent variables, six segmentations were performed with the MRS algorithm, in the eCognition 8.8 software. Three of them represent the three segmentation levels obtained from the application of the tool (named PT throughout the paper) presented in [14], using only the elevation layer as input.

The other three schemes were obtained using an improved version of the Estimation of Scale Parameter (ESP) tool [15], based on the best predictors identified as above. Thus, for the A-horizon thickness, segmentation process was based on slope derived in a 5x5 window and TWI. For soil classes, segmentation was performed on profile curvature (9x9 window) and slope (5x5 window).

Sampling schemes

The dataset was divided into two parts: 20 % of samples as control points for validation and 80 % as training points. The SBS schemes resulted by randomly selecting one sample per object out of the training points for each dependent variable and segmentation scheme. The other two sampling schemes, SRS and cLHS, were created using the same number of samples as in the object based-sampling.

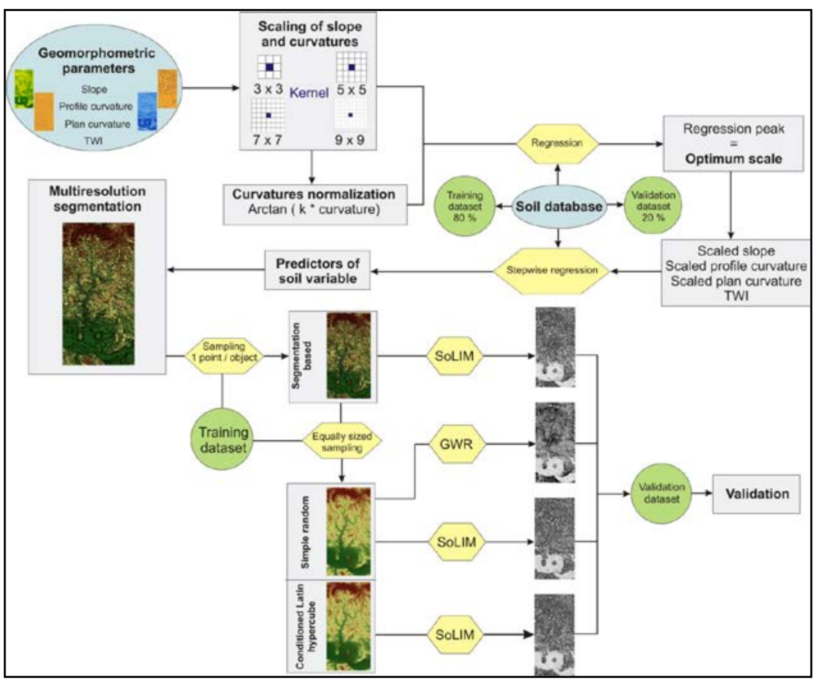

Figure 1. Flow chart showing the main steps of the experiment.

Soil mapping and accuracy assessment

Segmentation schemes were evaluated through their ability of accurately mapping soils. Thus, for each sampling scheme in Tables 1 and 2, maps of soil variables and classes were created using the sample-based inference engine implemented in the SoLIM Solutions 2010 software [16]. The agreement between maps and the reference data were assessed with standard methods: 1) overall accuracy and kappa index of agreement for soil classes; and 2) RMSE, agreement coefficient, mean absolute error and mean error for the A-horizon thickness.

RESULTS AND DISCUSSION

Segmentation produced between 11 and 40 homogeneous areas with the PT tool, and 6 to 22 with the ESP tool (Tables 1 and 2). As expected, accuracies improved with the number of samples for all sampling methods. However, the SBS performed consistently better than SRS and CLHS.

Predictions of the A-horizon thickness with SBS yielded the highest agreement coefficients (AC in Table 1), and the lowest RMSEs (except for PT 1, where cLHS was slightly better) and MAEs (except for PT 1 and 2, where cLHS was better). All SBSs produced results comparable with predictions that used the entire population of training samples. This includes the scheme with 6 samples (about 11 % of total). The other two methods produced less reliable results at the same number of samples, as shown by considerably lower AC values. The best results were achieved with the SBSs at the finest levels (PT3 and ESP3) of the two segmentation methods (Table 1).

Predictions of the soil classes with SBS outperformed SRS and cLHS in all cases but ESP3, where cLHS gave better results (Table 2). PT3 was the only sampling scheme that achieved results similar to those obtained with the entire population of training samples. These results were obtained with about 62 % of training samples. The other two methods produced significantly poorer predictions (kappa of 0.2 and 0.3 respectively). Interestingly, segmentations on the DEM alone (PT1 to 3) produced significantly better results (Table 2) than segmentation on the slope and profile curvature (ESP1 to 3), which predicted well the distribution of cambisols, luvisols and protisols in the study area. This might be due to the conflicts between homogeneity in profile vs. gradient (e.g. convexities/concavities have homogeneous profile, but heterogeneous gradient).

The good performance of SBS in optimizing soil sampling stems from the ability of LSS of delineating objects that maximize internal homogeneity and external difference. Conceptually, SBS can be seen as a particular case of purposive mapping [7], where the appropriate samples can be estimated within areas of homogeneous LSVs, which are delineated with the aid of segmentation instead of fuzzy c-mean classification. The results presented here agree with previous findings [4, 7] on improving the accuracy of soil mapping with limited samples.

When guided by local variance, LSS self-adapts to the scale of local variability in the LSVs, with additional benefits of accounting for spatial autocorrelation, anisotropy and non-stationarity [11]. To check the impact of non-stationarity on analyses, an additional test was performed with Geographically Weighted Regression (GWR) [17]. The samples obtained with SRS where employed for mapping the A-horizon thickness, using the same covariates as in the other predictions. Results always improved significantly (Table 1), which clearly shows the importance of local models in predicting soil properties.

TABLE I. ACCURACY ASSESSMENT OF THE A-HORIZON PREDICTIONS BASED ON DIFFERENT SAMPLING METHODS

Sampling method		***No.*[a]**	***RMSE***	***AC*[b]**	***MAE***	***ME***
PT1[c]	SRS	11	14.82	0.47	11.38	5.35
	GWR[d]	11	10.64	0.72	9.01	3.99
	SBS	11	11.95	0.75	9.65	0.19
	cLHS	11	11.52	0.63	9.33	- 1.85
PT2	SRS	14	13.92	0.72	11.45	3.41
	GWR	14	10.33	0.77	8.71	- 0.92
	SBS	14	10.41	0.79	9.10	- 3.17
	cLHS	14	11.03	0.74	8.35	3.21
PT3	SRS	31	11.54	0.73	10.27	- 1.36
	GWR	31	9.82	0.77	8.59	- 0.32
	SBS	31	8.84	0.85	6.61	- 2.97
	cLHS	31	10.28	0.84	8.67	3.00
ESP1	SRS	6	15.79	0.37	12.75	2.52
	GWR	6	10.72	0.53	8.60	0.84
	SBS	6	12.65	0.69	10.00	2.37
	cLHS	6	15.88	0.56	11.57	3.27
ESP2	SRS	8	15.67	0.27	11.87	1.76
	GWR	8	12.65	0.61	11.78	- 1.28
	SBS	8	11.25	0.71	9.37	- 1.16
	cLHS	8	12.21	0.69	9.63	- 3.46
ESP3	SRS	15	12.29	0.74	8.70	- 0.65
	GWR	15	8.83	0.79	7.77	- 2.65
	SBS	15	8.73	0.86	7.49	- 1.15
	cLHS	15	10.79	0.82	7.77	5.88
All training samples (80 %)		56	11.11	0.72	9.24	0.38

a. Number of samples. b. Agreement Coefficient. c.PT1 to 3- the three segmentation levels obtained with the PT tool; ESP1 to 3- the three segmentation levels obtained with the ESP tool. d. GWR was used for mapping based on the SRS samples

TABLE II. ACCURACY ASSESSMENT OF THE SOIL CLASSES PREDICTIONS BASED ON THREE SAMPLING METHODS

Sampling method		*No.*	*OA*	*KIA*
PT1	SRS	12	0.25	- 0.05
	SBS	12	0.38	0.09
	cLHS	12	0.25	- 0.01
PT2	SRS	15	0.31	0.05
	SBS	15	0.50	0.23
	cLHS	15	0.38	0.11
PT3	SRS	40	0.44	0.20
	SBS	40	0.63	0.45
	cLHS	40	0.50	0.30
ESP1	SRS	10	0.25	- 0.06
	SBS	10	0.31	0.07
	cLHS	10	0.25	- 0.04
ESP2	SRS	13	0.25	- 0.05
	SBS	13	0.31	0.05
	cLHS	13	0.31	0.01
ESP3	SRS	22	0.38	0.09
	SBS	22	0.38	0.15
	cLHS	22	0.44	0.14
All training samples (80 %)		65	0.63	0.45

In this case, GWR successfully substituted a poor sampling design by accounting for non-stationarity, which is a built-in capability of LSS.

In conclusion, SBS showed a high potential in optimizing soil sampling in the study area. The two SBS methods performed better than SRS and cLHS in predicting the A-horizon thickness and the soil classes. SBS would enable the reduction up to 11 % in the number of samples neccesary to predict the A-horizon thickness, and up to 62 % to predict soil classes. This methodology could be effective in reducing costs of soil surveys. The analyses presented here further highlight the importance of considering locally adaptive techniques in optimization of sampling schemes and predictions of soil properties.

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS - UEFISCDI, project number PN-II-ID-PCE-2011-3-0499. The Romanian dataset was generously provided by OSPA Hunedoara. We thank to three anonymous reviewers for providing comments and suggestions that greatly improved the paper.

REFERENCES

[1] Sanchez, P.A., S. Ahamed, F. Carré, A.E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A.B. McBratney, N.J. McKenzie, M.d.L. Mendonça-Santos, B. Minasny, L. Montanarella, P. Okoth, C.A. Palm, J.D. Sachs, K.D. Shepherd, T.-G. Vågen, B. Vanlauwe, M.G. Walsh, L.A. Winowiecki, and G.-L. Zhang, 2009. "Digital soil map of the World." Science, 325: 680-681.

[2] McBratney, A.B., M.L. Mendonça Santos, and B. Minasny, 2003. "On digital soil mapping." Geoderma, 117: 3-52.

[3] Brus, D.J., and J.J. de Gruijter, 1997. "Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion)." Geoderma, 80: 1-44.

[4] Yang, L., A.X. Zhu, F. Qi, C.-Z. Qin, B. Li, and T. Pei, 2012. "An integrative hierarchical stepwise sampling strategy for spatial sampling and its application in digital soil mapping." International Journal of Geographical Information Science, 27: 1-23.

[5] Minasny, B., and A.B. McBratney, 2006. "A conditioned Latin hypercube method for sampling in the presence of ancillary information." Computers & Geosciences, 32: 1378-1388.

[6] MacMillan, R.A., A. Torregrosa, D. Moon, R. Coupé, and N. Philips, 2009. "Chapter 24 Automated predicitve mapping of ecological entities", In Geomorphometry-Concepts, Software, Applications. Developments in Soil Science, vol. 33, Edited by: Hengl, T. and Reuter, H.I., Elsevier, Amsterdam.

[7] Zhu, A.X., L. Yang, B.L. Li, C.Z. Qin, E. English, J.E. Burt, and C.H. Zhou, 2008. "Purposive sampling for digital soil mapping for areas with limited data", In Digital Soil Mapping with Limited Data, Edited by, Springer, New York.

[8] Zhu, A.X., F. Qi, A. Moore, and J.E. Burt, 2010. "Prediction of soil properties using fuzzy membership values." Geoderma, 158: 199-206.

[9] Drăguţ, L., and C. Eisank, 2011. "Object representations at multiple scales from digital elevation models." Geomorphology, 129: 183-189.

[10] Baatz, M., and A. Schäpe, 2000. "Multiresolution Segmentation-an optimization approach for high quality multi-scale image segmentation", In Angewandte Geographische Informationsverarbeitung, Edited by: Strobl, J., Blaschke, T. and Griesebner, G., Wichmann-Verlag, Heidelberg.

[11] Drăguţ, L., C. Eisank, and T. Strasser, 2011. "Local variance for multi-scale analysis in geomorphometry." Geomorphology, 130: 162-172.

[12] Häring, T., and B. Schröder, "Sampling optimization using image segmentation," presented at the 4th International Workshop on Digital Soil Mapping, Rome, 2010.

[13] Drăguţ, L., T. Schauppenlehner, A. Muhar, J. Strobl, and T. Blaschke, 2009. "Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling." Computers & Geosciences, 35: 1875-1883.

[14] Drăguţ, L., and C. Eisank, 2012. "Automated object-based classification of topography from SRTM data." Geomorphology, 141–142: 21-33.

[15] Drăguţ, L., D. Tiede, and S. Levick, 2010. "ESP: a tool to estimate scale parameters for multiresolution image segmentation of remotely sensed data." International Journal of Geographical Information Science, 24: 859-871.

[16] Zhu, A.X., L. Band, R. Vertessy, and B. Dutton, 1997. "Derivation of soil properties using a soil land inference model (SoLIM)." Soil Science Society of America Journal, 61: 523-533.

[17] Brunsdon, C., A.S. Fotheringham, and M.E. Charlton, 1996. "Geographically weighted regression: A method for exploring spatial nonstationarity." Geographical Analysis, 28: 281-298.