Improved spatial prediction of soil properties and soil types combining semi-automated landform classification, geostatistics and mixed effect modelling

Omuto, Thine Christian
Dept. of Environmental and Biosystems Engineering
University of Nairobi
Nairobi, Kenya
cthine@uonbi.ac.ke

Paron, Paolo
Department of Water Engineering
UNESCO-IHE, Institute for Water Education
Delft, Netherlands
p.paron@unesco-ihe.org

Abstract — **Landforms classifications using a range of techniques from pure automatic classification to hybrid semi-automated model / expert opinion was tested as a base for a robust modelling of soil characteristics spatial distribution. The current approaches for producing soil maps use a single model which either blocks/controls the grouping effects or do not statistically recognize the natural landscape groupings. This study tested mixed-effects modelling technique for ingenious recognition of soil groupings and consequent improvement of the accuracy of the resultant soil maps. It further tested the various landscape classification for soil mapping. Mixed-effects modelling is a form of regression analysis for simultaneous modelling of the average landscape characteristics and individual units within the landscape. Hence, it can model a family of curves and potentially remove inadequacies inherent in the current models for soil mapping. Its potential in regression kriging of continuous and categorical soil attributes has been shown in this paper. Compared to the current application of a single model in regression kriging, mixed-effects modelling produced about five times improvement of the mapping accuracy. It is anticipated that its adoption will contribute to improved soil mapping.**

I. Introduction

In many studies there is an implicit equivalence between soil maps and the process of soil mapping. The latter, however, is a much more complex and long process than the final production of a map. In fact soil maps are only the final output in the process of mapping [1].

Historically soil classifications have remained largely stable over the past years, while field methodology, techniques and equipments have evolved at a much faster pace. An even greater excitement has been put in software development for soil mapping purposes. A pletora of automated procedures have been proposed [2], as well as new mathematical models developed for mapping soil units [3, 4,] together with remote sensing and digital terrain modelling analysis methods [5, 6, 7].

Recent research shows that opportunities still exist for improving the production of soil maps by accounting for more variability within- and -between soil mapping units [6].

Some of the methods in the literature tested so far for suitability in accounting for variability in the soil maps include numerical classification, multivariate statistical methods, geostatistics, fractal mathematics, etc. [4]. Geostatistics seem popular with many researchers perhaps because it is often easily implemented in numerous available GIS software [8]. Examples of geostatistical applications include regression kriging, kriging with external drift, universal kriging, etc. In regression kriging, the process of producing soil maps involves statistical modelling of the deterministic and stochastic components of the soil variables in the landscape. The deterministic component represents the large-scale trends while the stochastic component represents the small-scale autocorrelation trends. The large-scale trends are usually modelled using regression analysis while the autocorrelation trends are modelled with kriging analysis [9, 10,11,12]. The co-occurrence of regression and kriging analyses gives the name of regression kriging.

This study seeks to improve the regression part of the regression kriging method. The regression part is important since its results are the major inputs for the kriging part. Therefore, if it is substantially improved, it can potentially improve the performance of the regression kriging method. This study tested mixed-effects modelling approach for the improvement. Mixed-effects modelling is a form of regression analysis that can simultaneously model nested relationships. It is especially suitable in situations where unique relationships exist for certain individuals within groups and for different groups in a population. It's of particular interest in soil mapping because soil properties have unique relationships with soil forming factors in different catena and in the landscape in general. Even though these unique relationships have been recognized by pedologists, they have not been adequately represented in the modelling process for producing soil maps [2, 5]. Mixed-effects modelling approach presents the opportunity for recognizing such relationships and eventually contributing to accurate soil mapping. Mixed-effects modelling approach has been used in other studies with nested relationships, which is a promise for successful application in soil mapping. Some authors have used it to improve the modelling accuracy and efficiency [13, 14] while others have used it as a tool for incorporating environmental covariates in the modelling of continuous soil variables [15, 16]. These applications encourage the need to test its application in regression kriging of soil properties and soil types.

A recent study [17] has shown that mixed effect modelling has significantly improved the performance of regression kriging in soil mapping. This study takes that results further comparing the same methodology using different landform sources, spanning from pure geomorphometric terrain classification (TAS and Landserf) to an hybrid novel semi-automatic landform classification combining automated digital terrain analysis [18] with old-style physiographic province delineation [19] and geological attributes.

Landform classification has seen a new spring in recent years with both increasing interest in this topic and the partnership of geographical information, modeling and physical geography.

A very good recent panorama of methods and applications of geomorphometry is account for in [7]. A more specific example of semi automated data extraction for geomorphological mapping is provided in [20]. In this case study we have adopted and adapted existing packages and dataset [18, 21, 22] to the Kenyan topography

II. Study Area, Data and Methodology

A. *A.Study area*

his study focused on the Kenyan territories to test the various landform classification as a source for the combination of regression kriging and mixed effect modeling in soil mapping. Kenya is a semi-arid country, with highly contrasting topography, spanning from almost 6,000 metres above sea level, to the coastlines of the western Indian Ocean, and is cut through by the Eastern Branch of the Great Rift Valley. It has a bi-modal semi-arid climate, determined by the swift of the Monsoon in this part of the Indian Ocean. Vegetation wise it is dominated by shrubs and savannas plains with spotted and increasingly reducing tropical and open forests. The soils in Kenya are also highly variable, spanning from vertisols, andisols, black cotton to lithosols [23].

B. *Data*

Three sources of digital terrain classification have been used: a) a landform classification performed under TAS [24] and the recent WhiteBox environment [21]; b) a terrain classification using Landserf [22]; c) a combination of the global terrain model presented by [18] and the physiographic provinces of Kenya delineated by [19] together with a vector geological map derived from a the national geological map of Kenya [25].

374 georeferenced observations of classified soil profiles and clay content were used deriving from various sources like Tana River Development Authority [26], the Ministry of Livestock Development [27].

Besides the georeferenced soil data, this study also used rainfall data, Digital Elevation Model (SRTM, 90 metres DEM), land use, Normalized Difference Vegetation Index (NDVI) image, and the landform classification mentioned above as auxiliary information for spatial prediction of the soil attributes. All vector data were converted into raster, and after this all re-sampled to 1km resolution to match the smallest scale landform map [18].

C. Methodology

Regression kriging is a spatial prediction method that utilizes georeferenced observations to model the spatial pattern of an attribute and then using the model to predict the attribute in unsampled locations within a given study area. Its modelling of the spatial patterns entails the determination of large-scale trends present in the input data using regression analysis and small-scale autocorrelation using kriging analysis [30, 31, 32, 33, 11].

Mixed-effects modelling is a unique regression analysis that can simultaneously model nested hierarchical relationships. Its parameters for the high hierarchy (e.g. at landscape-scale) relationship are known as *fixed-effects* while the parameters associated with individual groups with the landscape are known as *random-effects* [13]. Fixed- and random-effects together form the mixed-effects modelling. For a detailed description of the method used in this study we refer to [17].

The application of mixed-effects model in regression kriging of continuous soil attributes was tested with the mapping of clay content in Kenya. The choice of predictors to use in the regression analysis was obtained from the correlation analysis between clay content and its potential predictors and between the predictors themselves. The predictors tested were DEM, NDVI, annual rainfall, and geographic coordinates. Since clay content and these predictors were positively skewed, they were first normalized with Box-Cox transformation [33] before the correlation analysis.

Spatial prediction of soil types used the classes in the classified soil profiles as the dependent variables and mean annual rainfall, NDVI, land use, landform, and geology as the predictors. These predictors were chosen to conform to the soil forming factors in the Jenny's Equation [34, 5]. They were modelled with a generalized linear mixed-effects (GLME) model for the deterministic part of the regression kriging [14].

The potential of mixed-effects modelling in regression kriging was assessed in two ways: one, by producing soil maps using regression kriging with and without mixed-effects and comparing the prediction accuracies of the two methods on holdout samples; and two, by assessing the magnitude of the nugget variance of the variograms for the resultant residuals from single and mixed-effects methods. In this study, the magnitude of the nugget variance was used to assess the potential of mixed-effects model in reducing modelling inadequacies in regression kriging of soil attributes.

III. Conclusions

The performance of mixed-effects in modelling the regression part of the regression kriging process was compared to the performance of the single model currently in application in the literature. The mixed-effects model produced higher correlation (r^2) between predicted and observed soil attributes and low residual standard errors (RSE) compared to the single model (Figure 1) using any of the three landform classifications.

The best results have been obtained with the hybrid landform classification which combines an automated landform analysis with pre-digital information and expert opinion.

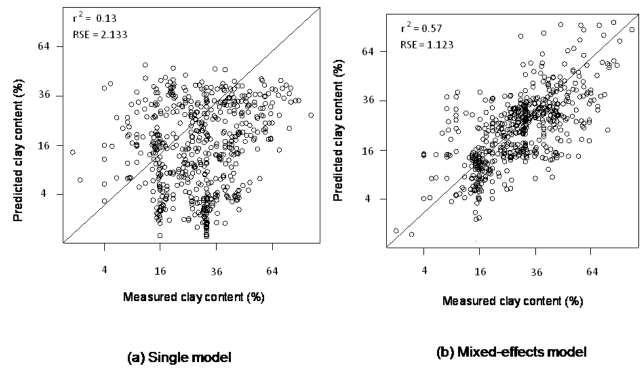

Figure 1: Comparison of performance of mixed-effects and single models in modelling regression part of regression kriging, using the Iwahashi & Pike landform classification (from [17])

Acknowledgment

A special acknowledgement is given to different authors who produced the R packages used in this study. The data used in this study was obtained from ISRIC[1], Kenya Meteorological Department[2], Geography Department of the University of Maryland[3], Regional Centre for Mapping of Resources for Development[4], Tana River Development Authority (TARDA), and the Government of Kenya.

References

[1] Buol, S.W., Hole, F.D. & McCracken, R.W. 1997. Soil Genesis and Classification. Iowa State University Press, USA.

[2] Zhu, A.X., Hudson, B., Burt, J., Lubich, K. & Simonson, D. 2001. Soil mapping using GIS, expert knowledge and fuzzy logic. Soil Science Society of America Journal, 65: 1463 – 1472.

[3] Heuvelink, G. & Webseter, R. 2001. Modelling soil variation: Past, present and future. Geoderma, 100, 269-301

[4] Lagacherie, P., McBratney A.B. & Voltz, M. 2007. Digital soil mapping. An Introductory Perspective. Elsevier, Amsterdam.

[5] McBratney, A.B, Santus M.L.M. & Minasny, B. 2003. On digital soil mapping. Geoderma, 117, 3 – 52.

[6] Brus, D.J., Bogaert, P. & Heuvelink, G.B.M. 2008. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. European Journal of Soil Science, 59, 166-177.

[7] Hengl, *T.,* Reuter*,* H.I. (eds) 2008. *Geomorphometry*: Concepts, Software, Applications. Developments in Soil Science, vol. 33, Elsevier, 772 pp

1 www.isric.org
2 www.kmd.ac.ke
3 http://pekko.geog.umd.edu
4 www.rcmrd.org

[8] Hengl, T., Heuvelink, G.B.M. & Rossiter, D.G. 2007. About regression kriging from equations to case studies. Computers & Geosciences, 33, 1301 – 1315.

[9] Lee, Y.M. & Ellis, J.H. 1997. On the equivalence of kriging and maximum entropy estimators. Mathematical Geology, 29, 131-152.

[10] Goovaerts, P. 1999. Geostatistics in soil science: State-of-the-art and perspectives. Geoderma, 89, 1-45.

[11] Hengl, T., Heuvelink, G.M.B. & Stein, A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, 122, 75-93.

[12] Douaika, A.T., Van Meirvennea M. & Toth, T. 2005. Soil salinity mapping using spatio-temporal Kriging and Bayesian maximum entropy with interval soft data. Geoderma, 128, 234-248.

[13] Pinheiro, J.C. & Bates, D.M. 2000. Mixed–Effects Model in S and S–Plus. Springer–Verlag, New York.

[14] Faraway, J.J. 2006. Extending the Linear Model with R: Generalized Linear Models, Mixed Effects and Nonparametric Regression Models. Taylor and Francis, Boca Raton, FL.

[15] Omuto, C.T., Minasny, B., McBratney, A.B. & Biamah, E.K. 2006. Nonlinear mixed–effects modelling for improved estimation of infiltration and water retention parameters. Journal of Hydrology, 330, 748 – 758.

[16] Omuto, C. T. & Gumbe, L.O. 2009. Estimating water infiltration and retention characteristics using a computer program in R. Computers & Geosciences, 35, 579-585.

[17] Omuto, C.T., & Paron, P. (in press). Combining mixed-effects modelling and regression kriging for improved spatial prediction of soil properties and soil types. European J. of Soil Sc.

[18] Iwahashi, J. & Pike, R.J. 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology, 86, 409-440.

[19] Ojany, F.F., 1966. The physique of Kenya. A contribution in landscape analysis. Annals Assoc. American Geographers, vol. 56: 183-196.

[20] Seijmonsberger, A.C., Hengl, T. & Anders, N.S., 2011. Semi-automated identification and extraction of geomorphological features using digital elevation data. In: Smith, MJ, Paron, P., & Griffiths, J.S. (2011) Geomorphological mapping methods and applications. Elsevier, Amsterdam.

[21] Lindsay, J., 2009. Whitebox GAT (ver. 1.0.7)[5]

[22] Wood, J. 2010. LandSerf: A Geographic Information System for the Visualization and Analysis of Surfaces, ver 2.3. City University of London[6]

[23] Sombroek, W.G., Braun, H.M.H. & van der Paow, B.J.A. 1982. Exploratory soil map and agro-climatic zone map of Kenya. Scale 1:1000000. Kenya Soil Survey, Ministry of Agriculture. Nairobi.

[24] Lindsay, J., 2005. The Terrain Analysis System: a tool for hydro-geomorphic applications. Hydrological Process., 19: 1123-1130

[25] Ministry of Energy and Regional Development of Kenya, 1987. Geological Map of Kenya with Bouger Gravity Contours, scale 1:1,000,000

[26] Tana River Development Authority, 1987. Development Plan. Report Vol II. TARDA, Nairobi.

[27] Touber, L. 1991. Wet season restrictions in accessibility of livestock in Mandera and Wajir Districts. Ministry of Livestock and Development, Government of Kenya, Nairobi, Kenya.

[28] Batjes**,** N.H. & Gicheru, P., 2004. Soil data derived from SOTER for studies of carbon stocks and change in Kenya (ver. 1.0). Report 2004/01, ISRIC - World Soil Information, Wageningen

[29] Batjes, N.H. 2008. ISRIC-WISE Harmonized Global Soil Profile Dataset. ISRIC-World Soil Information, Wageningen.

[30] Webster, R., 1989. Is regression what you really want? *Soul Use & Management*, **5,** 47-53.

[31] Odeh, I.O.A, McBratney, A.B. & Chittleborough, D.J. 1995. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression kriging. *Geoderma,* **67,** 215 – 226.

[32] Nielsen, D. & Wendroth, O. 2003. Spatial and Temporal Statistics: Sampling Field Soils and their Vegetation. Catena-Verlag, Reiskirchen.

[33] Eagleson, G.K. & Muller, H.G. 1997. Transformations for smooth regression models with multiplicative errors. *Journal of Royal Statistical Society*, **59,** 173-189.

[34] Jenny, H. 1964. Factors of Soil Formation: A System of Qualitative Pedology. Dover Publications Inc., New York.

5 http://www.uoguelph.ca/~hydrogeo/Whitebox/download.shtml
6 http://www.landserf.org/