

ASPECTS REGARDING THE USEFULNESS OF GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) FOR DIGITAL MAPPING OF SOIL PARAMETERS

C. PATRICHE¹, I. VASILINIUC²

¹ Romanian Academy, Iași
e-mail: pvcristi@yahoo.com

² University „Al. I. Cuza” Iași

An alternative to classical soil mapping, using the limits of the soil spatial units, is the digital mapping, which makes use of statistical methods and GIS-derived spatial layers for a more accurate and continuous display of soil parameters' spatial distribution. Our study attempts to test the usefulness of a relatively new statistical method, namely the Geographically Weighted Regression (GWR). In order to reveal its superiority, the results were compared with those achieved by means of other digital mapping methods (global multiple regression, ordinary kriging, cokriging). The methods were applied in the area of Horoiata Basin (Tutovei Hills) for soil reaction values, using the data from 133 soil profiles. For validation purposes, the statistical models were applied to an independent, quasi-random sample of soil profiles, which was not taken into account for models elaboration, and we compared the real and estimated pH values.

Key words: *Geographically Weighted Regression, soil digital mapping*

Our study attempts to investigate the possibilities of deriving digital maps of soil parameters using statistical methods in GIS. The research is motivated by the discontinuous character of the classical soil mapping, due to the assignment of a unique value for a certain Ecological Homogeneous Territory (TEO).

MATERIAL AND METHOD

The study region is represented by Horoiata Basin, situated in Tutovei Hills and covering a surface of about 86km² (fig. 1). We analyzed the soil reaction spatial distribution, starting from a sample of 133 soil profiles. The initial sample was first divided into a *working sample* of 123 soil profiles, which was used for the elaboration of statistical models, and a randomly selected *validation sample* of 10 soil profiles, which was used for the validation of the results achieved by means of the working sample (fig. 2).

Several spatialisation methods were applied, using ArcGIS 9.3 Geostatistical Analyst extension, namely: *ordinary kriging*; *cokriging* with altitude as auxiliary variable; *cokriging* with altitude and slope as auxiliary variables; *multiple stepwise global regression* with altitude and slope as predictors; *Geographically Weighted Regression (GWR)* with altitude and slope as predictors.

The outcomes of all these methods were compared in order to select the best approach. The validation of the spatial models was carried out using different

procedures. Cross-validation is the “leave one out” validation procedure and it was performed on the working sample. For a certain point of this sample, the procedure implies the comparison of the real value with the one estimated when the point value was not included in the analysis. The independent sample validation was applied to the validation sample (not included in models elaboration) and it implies the comparison of the real and the estimated values. Statistical validation implies testing the statistical significance of quality parameters (correlation coefficients, partial regression coefficients, intercept), testing the residuals normality, computation of standard errors.

The predictors we used for regression analysis and cokriging were the Digital Elevation Model (DEM) with a 10x10m resolution and the slope of the terrain, derived from DEM at the same spatial resolution.

The first major applications of ordinary kriging in soil studies in the early 1980's [2] was mainly at the field scale. Since then it has been extended to larger areas, upscaling to the regional extent. Previous studies [11], [12], [7] have demonstrated the superiority of residual kriging to other spatialisation methods such as ordinary kriging, universal kriging, multiple-linear regression and cokriging. The methodology of all these methods along with examples of applications can be found in many scientific publications related to the use of (geo)statistical methods in geosciences [9], [5], [8], [3], [6]. The methodology of the Geographically Weighted Regression, on which our research is focused, is described by its promoters [4].

The main advantage of using the regression as a global spatialisation method is that of quantifying the roles played by the predictors in explaining the spatial distribution of the analyzed parameter. The main disadvantage is the smoothing of the real spatial variation along with the modification of the real values in the known points. Moreover, a global regression model is unable to render spatial anomalies. Consequently, the regression analysis is more appropriate for those parameters showing weaker spatial variation. In contrast, the local approach by means of kriging has the advantage of preserving the values in the known points and of being capable to render the spatial anomalies. On the other hand, kriging does not explain the spatial distribution of the analyzed parameters, the effect of the controlling factors being implicit. Consequently, the kriging approach is better suited for parameters displaying a more accentuated spatial variation. We may notice that the two spatialisation methods are complementary and their combination (known as residual kriging or regression-kriging) should produce a superior spatialisation method.

The GWR approach eliminates the disadvantage of the global regression method of being unable to render spatial anomalies. This happens because the method applies the multiple regression locally, inside circular moving windows, with fixed or variable dimensions.

RESULTS AND DISCUSSIONS

According to the soil profiles sample, the soil reaction values in our region of interest varies from 5.6 to 8.8, with a mean value of 7.4 and a standard deviation of 0.769. The frequency distribution of pH values is bimodal (*fig. 2*). One peak is situated at lower pH values (around 6.7) and it corresponds to higher terrain altitudes, from the northern, eastern and western part of our region, where the pedogenesis is more advanced and soils display clay accumulation and little carbonates along their profiles. The other peak is located at higher pH values (around 7.9) and it corresponds to lower altitude areas, with alluvial and colluvial

soils, where materials detached from slopes by erosion, having higher carbonates concentrations, accumulate.

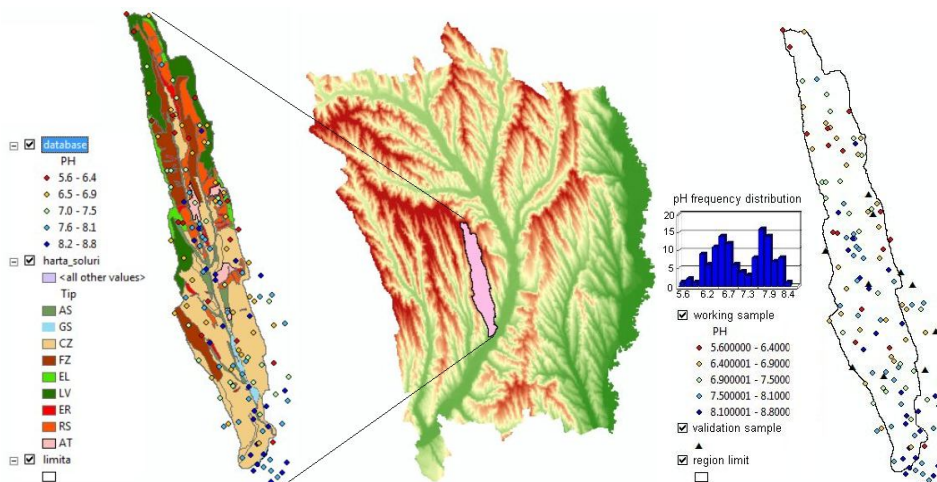


Figure 1 Position of Horoiata Basin and distribution of soil units

Figure 2 Distribution of working and validation samples

Therefore, we may notice that the spatial distribution of soil reaction is partially related to terrain configuration, mainly to altitude and slope. These spatial trends can be statistically reproduced using regression models. The global regression model (fig. 3) shows that there is an inverse relationship between soil reaction values and altitude, meaning that soil reaction becomes more acid as the terrain altitude grows, due to the transition towards more developed soil profiles. The regression coefficient of -0.00569 for altitude means that the pH values decrease, on average, with 0.569 units when the altitude increases with 100m . The positive regression coefficient for slope (0.0293) indicates a direct relationship between soil reaction and slope values, meaning that, generally, steeper slopes are associated with higher pH values, due to more intense erosion which brings to surface higher carbonated soil materials. The average variation in this case is 0.0293 pH units / 1° of slope.

If regression models are able to render continuously spatial trends, they are unable to account for local variations, induced by local factors. These local variations can be rendered by means of interpolation methods such as kriging (fig 3). Ordinary kriging produces a spatial model depending only on point values. The values at any unsampled location is computed as a weighted average of the values from the neighboring points, therefore ignoring any influence of potential predictors. The cokriging approach corrects partially this weakness, by including in the model the influence of auxiliary factors (altitude and slope in our case), but this influence is only taken into account at sample locations and not continuously at any location.

To sum up, we may notice that the two broad categories of spatialisation methods have important drawbacks. The regression models include spatial trends but are unable to render spatial anomalies, while the kriging models render local variations, but do not include spatial trends. The classical approach to solve this problem is the regression-kriging (residual kriging) approach, which combines the 2 methods.

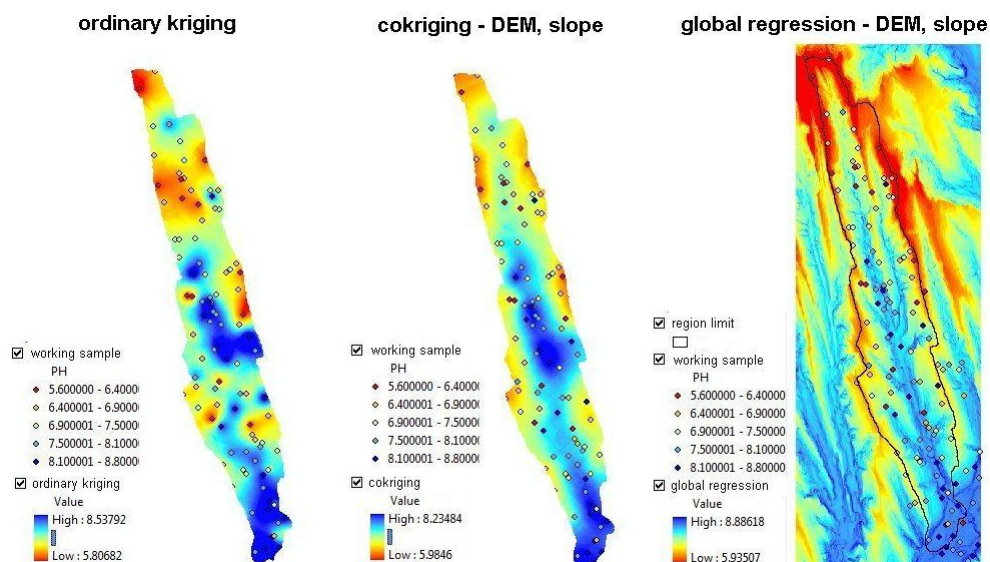


Figure 3 **Spatial models of soil reaction resulted form application of different methods**

Another possible solution, tested in our study, is the application of regression as local interpolator. This is exactly what the Geographically Weighted Regression does, by computing regression models at each location based on the values from the neighboring points and their associated predictors' values.

The way GWR works and the outcome for soil reaction is shown in *figure 4*. GWR produces, apart from the digital map of the analyzed parameter, a series of raster layers storing in each pixel the values of the parameters defining the regression equations, namely the intercept and the regression coefficients. A unique regression model is applied in each pixel. The quality of these models vary significantly from one location to another, meaning that there are locations where soil reaction depends greatly on altitude and slope, while in other location these spatial relations fade considerably.

By applying regression locally, GWR is able to account for both spatial trends and local variations, which makes it, theoretically, superior to the other methods we tested. This is also proved by the quality parameters for the validation sample (*tab. 1*). From all the tested methods, GWR produces the smallest root mean square error (RMSE) and the highest slope for the regression line defining the relation between the observed and the predicted pH values. Unfortunately, the

method implemented in ArcGIS 9.3 does not produce a cross-validation analysis for the working sample.

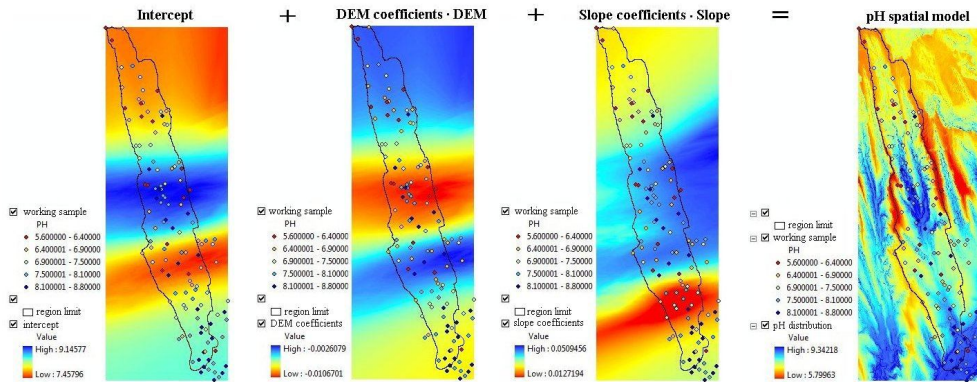


Figure 4 Application of GWR for modeling the spatial distribution of soil reaction, using DEM and slope as predictors

Table 1
Root mean square error (RMSE) and slope of regression line for cross-validation of the working sample and for the validation sample

Method	Working sample		Validation sample	
	RMSE	Slope of regression line	RMSE	Slope of regression line
Ordinary kriging	0.692	0.368	0.761	0.090
Cokriging – altitude	0.697	0.513	0.631	0.312
Cokriging – altitude, slope	0.657	0.566	0.686	0.252
Global regression – altitude, slope	0.636	0.331	0.621	0.299
GWR - altitude	-	-	0.638	0.329
GWR – altitude, slope	-	-	0.613	0.418

CONCLUSIONS

From all the methods tested, the best results were achieved using the Geographically Weighted Regression (GWR) with DEM and slope as predictors. Nevertheless, the models obtained so far have a modest degree of explanation for the spatial distribution of pH values. Some of the **possible ways of improving** their performances could be: the combined application of GWR with kriging of the residuals; the testing of other potential predictors, of non-linear models; the integration of the discrete information regarding the spatial distribution of soil units, geological substratum, land use etc.

Our analysis reveals some of the advantages and disadvantages of using statistical methods in GIS for modeling the spatial distribution of soil parameters.

Among **the advantages**, we may mention: the fast and accurate processing of information; the possibilities of estimating the errors; the continuous rendering of the spatial distribution of the analyzed parameter; the explication of the spatial distribution; the possibility to integrate and perform a complex analysis using a various spatial information. Among **the disadvantages**, we may mention: the need of a large database; the need of complex statistical and GIS software; the complexity of the methods: it is imperative for the researcher to know very well all the aspects of the applied methods in order to avoid misinterpretations.

BIBLIOGRAPHY

1. Bishop, T.F.A., Minasny, B., 2005 - *Digital Soil-Terrain Modeling: The Predictive Potential and Uncertainty*. S. Grunwald (Editor), Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics. CRC Press, Boca Raton, Florida, pp. 185-213.
2. Burgess, T.M., Webster, R., 1980 - *Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging*. Journal of Soil Science 31, 315-331.
3. Burrough, P.A., McDonnell, R. A., 1998 - *Principles of Geographical Information Systems*, Oxford University Press.
4. Fotheringham, S., Brunsdon, C., Charlton, M., 2002 - *Geographically Weighted Regression. The analysis of spatially varying relationships*, Wiley.
5. Goovaerts, P., 1997 - *Geostatistics for natural resources evaluation*, Oxford University Press.
6. Hengl, T., 2007 - *A Practical Guide to Geostatistical Mapping of Environmental Variables*, JRC Scientific and Technical Research series, Office for Official Publications of the European Communities, Luxembourg, EUR 22904 EN, 143 pp.
7. Hengl, T., Heuvelink, G.B.M., Stein A., 2004 - *A generic framework for spatial prediction of soil variables based on regression-kriging*, Geoderma 120, 75–93.
8. Isaaks, E.H., Strivastava, R. M, 1989 - *An Introduction to Applied Geostatistics*, Oxford University Press, 561 pp.
9. Johnston, R. J., 1978 - *Multivariate Statistical Analysis in Geography*, Longman, New York.
10. McBratney, A.B., Mendonça Santos, M.L., Minasny B., 2003 - *On digital soil mapping*, Geoderma 117, p. 3-52.
11. 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, 215-225.
12. Odeh, I. O. A., McBratney, A. B., 2000 - *Using AVHRR images for spatial prediction of clay content in the lower Namoi valley of eastern Australia*, Geoderma, 237-254.
13. Patriche, C.V., 2007 - *Spatialisation methods for (agro)climatic and soil parameters at regional and local scales*, Poster Papers Proceedings, 6th ECPA Conference, Skiathos, Greece, Fountas, S., Gemtos, T.A., Aggelopoulou, K.D., Blackmore, B.S. (editors).
14. Smith, M.J., Goodchild, M.F., Longley, P.A., 2007 - *Geospatial Analysis. A Comprehensive Guide to Principles, Techniques and Software Tools. Second Edition*, Matador, Troubador Publishing Ltd., 2007, online version: <http://www.spatialanalysisonline.com/>.
15. *** - *ArcGIS Desktop 9.3 Help*, Environmental Systems Research Institute (ESRI), <http://webhelp.esri.com/arcgisdesktop/9.3/>.