Prediction of Soil Properties Using Fuzzy Membership

Zhu, A.X.[1,2*]; Moore, A.[3]; Burt, J. E.[2]

[1]State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, axing@lreis.ac.cn

[2]Department of Geography, University of Wisconsin, Madison, Wisconsin, U.S.A. Email: azhu@wisc.edu

[3]USDA-NRCS-National Geospatial Development Center, 157 Clark Hall Annex, Prospect Street, West Virginia University, Morgantown, WV 26506

Keywords: Fuzzy Set, Geographic Information System, Soil Properties, Expert System, Soil-Landscape Model, SoLIM

Abstract

This research explores the use of fuzzy membership values generated by the Soil Landscape Inference Model (SoLIM) to predict detailed spatial variation of soil properties. Two fuzzy membership based approaches were used to predict soil property values over space. The first is a fuzzy membership weighted average model with which the soil property value at a location is the weighted average of the fuzzy membership values and the typical soil property values of the soil types. This approach has two models: one based on the typical values from soil description and the other based on the property values at the locations with maximum fuzzy membership values. The second approach is a multiple linear regression with fuzzy membership values where by the soil property value at a location is predicted using a regression between the soil property and fuzzy membership values. These models were then compared with a predictive model based on existing soil survey data and a predictive model based on multiple linear regression with terrain attributes. A case study in the Driftless Area of southwestern Wisconsin showed that that over flat areas where relationships between soil property values and terrain attributes approach linear, linear regression with topographic variables would work well, but over areas of stronger relief where relationships between soil property values and terrain attributes are non-linear, regression with fuzzy membership values is an improvement. However, from the perspectives of data requirement and able to handle non-linearity, the weighted average model would have clear advantages over the other two.

1. Introduction

Soil property maps generated from conventional soil survey maps are no longer sufficient because they often do not represent the spatial variability of soil properties at the level of detail needed for many applications. Statistical/geostatistical methods have been used to provide detail spatial variability of soil properties (McBratney and Webster, 1986; Webster, 1991; Moore et al., 1993; Gessler et al., 2000). However, these techniques rely too heavily on the assumption of linearity and stationarity. It is unlikely that a direct linear relationship exists

* Corresponding author. Tel: +86-(0)10-6488-8961; Fax: +1-608-265-3991
Email address: axing@lreis.ac.cn

between terrain attributes and soil property values; in fact, the relationships between soil property variation and underlying terrain variables can be very complex (Lark, 1999). The linearity and stationarity assumption and the data requirements of these techniques present stiff challenges to their application over large and diverse landscapes.

This research explores the possibilities of using fuzzy membership values generated by the Soil Landscape Inference Model (SoLIM) (Zhu, 1997; Zhu et al., 2001) to predict soil property values in areas where the relationship between soil property values and terrain attributes is perceived to be non-linear. The soil similarity vector of a local soil derived using the SoLIM approach can be viewed as a non-linear transformation of environmental variables based on expert knowledge of soil-landscape relationships. Variation of soil properties over landscapes with well-defined landscape positions and distinct soil types may be considered highly non-linear. Soil properties on these landscapes tend to change gradually within landscape positions and quickly in transition zones between landscape positions. The premise of this research is that the inherent non-linearity of soil similarity vectors can be used to describe and model non-linear variation in soil property values.

2. Material and Methods

2.1 Soil similarity vector and SoLIM

SoLIM is a predictive approach to soil mapping. It consists of two major components: a similarity model for representing soil spatial variation and a set of inference techniques for populating the similarity model (Zhu, 1997). Under the similarity model a given area is represented as a raster layer. The size of each grid (pixel) in the raster layer is often very small (such as 10 meters or 30 meters on each side). The soil at a given pixel (*i,j*) is then represented by an *n*-element similarity vector (referred to as soil similarity vector), $\boldsymbol{S}_{ij} = (S_{ij}^{1}, S_{ij}^{2}, \ldots S_{ij}^{k} \ldots S_{ij}^{n})$, where *n* is the number of prescribed soil classes (such as taxonomic units) over the area and S_{ij}^{k} is the similarity value of the soil at pixel (*i,j*) to the prescribed soil class *k*. It must be pointed out that S_{ij}^{k} is an index which measures the similarity between the local soil at (*i,j*) to soil class *k*. The more similar a soil is to a prescribed soil class, the higher its similarity value (fuzzy membership). Thus, a similarity value of 1.0 means that the soil at (*i,j*) is a typical instance of the prescribed class while a similarity value of 0.0 means that the local soil does not belong to the prescribed soil class at all. With this similarity model, spatial variation of soil can be described at a very detailed level. The coupling of a raster representation in the spatial domain with a similarity representation in the attribute domain allows the spatial variation of soils to be expressed and retained at much greater details than it can be in conventional soil maps.

The SoLIM approach for populating the similarity model is based on the classic concept that soil is a product of interaction among climatic factors, landform, parent material, organism, and hydrological factors over time (Jenny, 1980; Hudson, 1992). In other words, there exist relationships between soils and the environmental conditions under which they formed. We may infer the soil type at a given location if we have local environmental conditions and the knowledge of how these environment conditions are related to the soils. Zhu and Band (1994) and Zhu (2000) predict soil series distribution with the use of artificial intelligence (AI) and GIS/remote sensing (RS) techniques. AI techniques can be used to extract knowledge on soil-environment relationships (Zhu, 1999; Zhu, 2000; Qi and Zhu, 2003). GIS/RS techniques were used to characterize soil formative environmental conditions (Zhu et al., 1996). The extracted knowledge and the characterized environmental conditions can then be linked through a set of inference techniques to derive the soil similarity vector for each location (pixel) (Zhu and Band, 1994).

2.2 Methods

Two fuzzy membership based approaches were used to predict soil property values over space. One is a fuzzy membership weighted average model with which the soil property value at a location is the weighted average of the fuzzy membership values and the typical soil property values of the soil types (Equation 1).

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}} \qquad (1)$$

where V_{ij} is the predicted soil property value at location (i,j), $S^{k}_{i,j}$ is the fuzzy membership value in soil type *k* for the soil at the given location, and v^k is the typical soil property value for soil type *k*. This approach has two models: one uses the representative value of the given soil property from soil description as v^k (referred to as the ***weighted average model***) and the other uses the property value at the location where the fuzzy membership value in the given soil series is the highest (***maximum membership model***). The other approach is a multiple linear regression with which the soil property value at a location is predicted using a regression between the soil property and fuzzy membership values in each of the soil classes (***membership regression model***).

The above three models were then compared with a predictive model based on existing soil survey data (***soil map model***) and that based on multiple linear regression with terrain attributes (***terrain regression model***). The soil map model uses the typical soil property values based on the existing soil survey of the area to approximate the soil property values at local sites. The typical values of the soil properties for the soil series in each of these map units were determined based on the Map Unit Interpretation Record (MUIR) database (Soil Survey Staff, 1997).

For the terrain regression model, elevation, slope, aspect, planform curvature, and profile curvature were used as the independent variables. While other topographic variables likely have some influence on soil formation in this study area, only these five were used in the process of creating the SoLIM-generated soil map (Smith, 2004).

The following soil properties were used for each of the models: A-horizon sand and clay content, Bt_1-horizon sand and clay content, depth to Bt_1-horizon, loess thickness, and depth to weathered bedrock.

2.3 Study area and research data

The study was conducted in a watershed in the Driftless Area of southwestern Wisconsin (Figure 1). The study area consists of two distinct but related areas: the first has gently rolling terrain consisting of a thin layer of loess over clayey residuum underlain by fractured dolomite, and soil classes that differ primarily in depth to bedrock; the second has steeper terrain, more variable soil types, and occurs in places where stream channels have cut through the dolomite to expose the sandstone below.

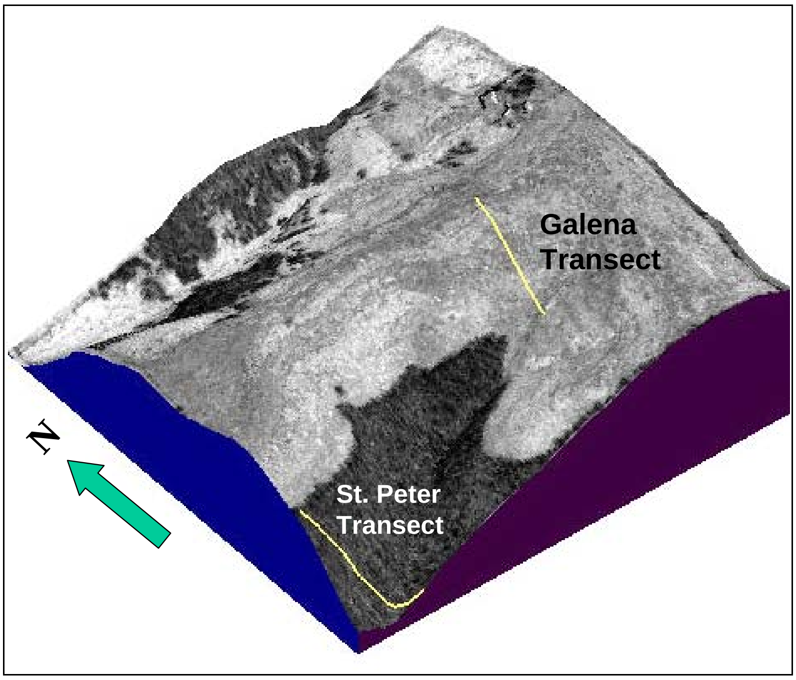

Figure 1: Topography of the study area and the location of field transects

In order to characterize the landscape, two representative transects were established: one on the gently rolling summit (the "Galena transect"), and one on the steeply sloping backslope (the "St. Peter transect") (Figure 1). These transects were based on preliminary field investigations and existing soil maps and designed to capture the maximum amount of soil variation possible. The Galena transect starts from a convex position on the summit and extends across the summit, down the shoulder and into a concave drainage way and consists of 32 sample locations. The St. Peter transect starts on a shoulder, extends down the steep south-facing backslope to a concave footslope and into the drainage way. It resumes at the base of the north-facing slope and extends through the footslope, up the backslope, and terminates 10 meters past the transition from backslope to shoulder. 43 observations were made on the transect: 28 on the south facing slope and 15 on the north-facing slope.

Fuzzy membership maps of soil series for the area created in a different study (Smith, 2004) were used to create the soil similarity vectors for locations along these two transects. These soil similarity vectors were then used in three membership based models.

3. Results and Discussion

Tables 1 and 2 compare the mean absolute errors of predictions from all five models across the 7 soil properties used along the two transects. Over the gently rolling area (along the Galena transect), the terrain regression model produces mean absolute error (MAE) values ranging from 1.1 to 9.0 times lower than other groups of models. In additions, R^2 values for this model range from 0.1 to 0.3 higher than other groups of models. Over the steep area (along the St. Peter transect), the membership regression model had the MAE values ranging from 1.5 to 17 times lower than other models and the R^2 values for the model range from 0.1 to 0.8 higher than other groups of models.

From the above we can observe that in predicting soil property linear regression models based on the terrain attributes may be limited to areas with gentle landscapes. For steep landscapes, the relationships between soil property and terrain attributes can be highly non-linear and non-linear transformation of the terrain variables would be required if satisfactory prediction is to be made.

The maximum membership model produced reasonably good performance measures comparing to the two statistical models. One must realize that the two statistical models (terrain-based and membership based) use all field points in model development. This means that the error measures (or performance measures) are those of model development, not those of model validation. The maximum membership model only used one field sample per soil series (the sample with the maximum membership in that series) for model development. In addition, the performance measures for the maximum membership model are those of model validation (that is, only the field samples, not used in the model development, were used to compute the performance measures). In this sense the maximum membership model may have out performed both regression models. From the field data requirement perspective, the maximum membership model has clear advantages over the statistical models.

Table 1: MAE for all selected models - Galena transect

Property	Soil Map Model	Weighted Average Model	Maximum Membership Model	Terrain Regression Model	Membership Regression Model
A-horizon Sand	5.53	5.56	6.68	0.74	0.73
A-horizon Clay	1.73	2.00	1.69	1.31	1.90
Bt1-horizon Sand	6.12	2.62	1.15	1.02	1.14
Bt1-horizon Clay	4.83	5.67	4.95	3.75	4.83
Depth to Bt1	10.16	9.37	8.68	7.67	8.02
Loess Thickness	30.32	24.35	14.81	9.46	14.29
Depth to Cr	29.37	24.26	16.23	11.56	15.36

Note: Cells highlighted in yellow have the lowest MAE values for each soil property. Cells highlighted in green have the second lowest MAE for each soil property.

Table 2: MAE for all selected models – St. Peter transect

Property	Soil Map Model	Weighted Average Model	Maximum Membership Model	Terrain Regression Model	Membership Regression Model
A-horizon Sand	18.35	15.38	7.62	6.72	4.81
A-horizon Clay	5.63	5.33	3.04	3.57	1.63
Bt1-horizon Sand	15.89	16.21	10.98	10.97	4.3
Bt1-horizon Clay	6.99	5.61	17.12	3.67	2.28
Depth to Bt_1	13.35	12.24	10.86	9.11	9.28
Loess Thickness	56.29	59.78	5.98	4.88	3.45
Depth to Cr	17.72	20.22	17.94	15.48	9.37

4. Conclusions

The implications are that linear regression models based on topographic variables might be appropriate over gentle landscapes where the relationships between soil property values and terrain attributes approach linear. For areas with steep landscape, linear regression models based on topographic variables might break down. Non-linear transformation of the topographic variables is needed for the linear regression models to be effective. This study further suggests that weighted average model using maximum fuzzy membership values as a way to define the representative soil property values would have clear advantages over the statistical models from the perspectives of field data requirement and ability in handling non-linearity of the relationships between soil properties and terrain variables.

Acknowledgment

The research reported here was supported by Natural Resources Conservation Services, United States Department of Agriculture through its Graduate Studies Program to Amanda Moore and by the Chinese Academy of Sciences through its "One-Hundred Talents" program to Prof. A-Xing Zhu.

References

Gessler, P.E., O.A. Chadwick, F. Chamran, L. Althouse, and K. Holmes. 2000. Modeling soil-landscape and ecosystems properties using terrain attributes. *Soil Sci. Soc. Am. J.* 64:2046-2056.

Hudson, B.D., 1992. The soil survey as paradigm-based science. *Soil Science Society of America Journal*, 56, pp. 836-841.

Jenny, H., 1980. *The Soil Resource: Origin and Behaviour*, Springer-Verlag, New York, 377 pp.

Lark, R.M. 1999. Soil-landform relationships at within-field scales: An investigation using continuous classification. *Geoderma* 92: 141-165.

McBratney, A.B. and Webster, R., 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates, *Journal of Soil Science*, Vol. 37, pp. 617-639.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* 57:443-452.

Qi, F. and A.X. Zhu, 2003. "Knowledge discovery from soil maps using inductive learning", *International Journal of Geographical information Science*, Vol. 17, No. 8, pp. 771-795.

Smith, M.P. 2004. *Optimal DEM Resolution and Neighborhood Size for Soil Resource Inventory Using the SoLIM Approach.* pp 85. Master's thesis, University of Wisconsin-Madison.

Soil Survey Staff. 1997. *National Map Unit Interpretation Record* (1997) [Online WWW]. Available URL: http://soils.usda.gov/soils/survey/nmuir/index.html [Accessed 10 February 2004]. Natural Resources Conservation Service, USDA, National Soil Survey Center, Lincoln, NE.

Webster, R., 1991. Local disjunctive kriging of soil properties with change of support, *Journal of Soil Science*, Vol. 42, pp. 301-318.

Zhu, A.X., 1999. A personal construct-based knowledge acquisition process for natural resource mapping. *International Journal of Geographical Information Science*. 13, 119-141.

Zhu, A.X., 2000. Mapping soil landscape as spatial continua: the neural network approach. *Water Resources Research*, 36, 663-677.

Zhu, A.X. and Band, L.E.,1994. A knowledge-based approach to data integration for soil mapping. *Canadian Journal of Remote Sensing*, 20, 408-418.

Zhu, A.X., Band, L.E., Dutton, B., Nimlos, T.J., 1996. Automated soil inference under fuzzy logic. *Ecological Modelling*. 90, 123-145.