Available online at
http://journals.usamvcluj.ro/index.php/promediu

ProEnvironment

ProEnvironment 6 (2013) 556 - 561

Original Article

Sampling Design Optimization Based on Soil-Land Inference Model (SoLIM)

SARMADIAN Fereydoon[1], Ali KESHAVARZI*[1], Antonia ODAGIU[2]

[1]*Laboratory of Remote Sensing and GIS, Department of Soil Science Engineering, University of Tehran, P.O.Box: 4111, Karaj 31587-77871, Iran*
[2]*University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Mănăştur St., no. 3-5, 400372 Cluj - Napoca, Romania*

Received 8 November 2013; received and revised form 21 November 2 013; accepted 30 November 2013
Available online 30 December 2013

Abstract

Soil is an essential source of ecosystem services such as food production and climate regulation. Soil information is of fundamental importance for decision making on adequate land use planning and management and environmental protection which is in fact the motivation behind soil surveys. One of the majors concerns in soil science relies on the fact that the conventional methods of soil analysis are too expensive and time-consuming, and soil legacy databases are often not adequate for assessing and mapping the soil condition. In this sense, producing relevant soil information for improving the current soil legacy databases is one of the big goals of soil sensing and digital soil mapping. 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. As sampling is constrained by financial resources, efficient sampling strategies are desirable. In this paper, we are focusing on sampling design optimization based on Soil-Land Inference Model (SoLIM) with respect to environmental covariates (soil and terrain attributes).

Keywords: sampling strategy, soil mapping, terrain characteristics, Soil-Land Inference Model.

1. Introduction

1.1. Soil Maps and Digital Soil Mapping (DSM)

Soil maps are used for many purposes. To name a few examples, they are used in planning to evaluate land allocation scenarios, in agronomy to assess the suitability of the land for growing crops or assess the faith of pollutants such as pesticides, in ecology to develop nature conservation plans, and in hydrology and climatology to describe the role of the soil in the hydrological cycle [1].

Since many years, national governments and international organisations have therefore put much effort in mapping the soil. Soil maps are also increasingly used to derive spatially distributed soil inputs to environmental and ecological process models. For instance, soil maps provide important information about physical, chemical and biological soil properties needed by acidification and groundwater flow models [1].

Digital soil mapping (DSM) is defined as the creation and population of spatial soil information systems by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems [2, 3]. Efficient sampling designs play an important role in DSM [4], as they have a significant impact on the accuracy of the maps [5].

* Corresponding author.
Tel & Fax: +982632231787
e-mail: alikeshavarzi@ut.ac.ir

Digital soil mapping does not just produce a paper map; it is a dynamic process in which geographically referenced databases are created at a given spatial resolution.

A digital soil map is essentially a spatial database of soil properties, based on a sample of landscape at known locations.

Field sampling is used to determine spatial distribution of soil properties, which are mostly measured in the laboratory.

These data are then used to predict soil properties in areas not sampled [3]. Digital soil maps describe the uncertainties associated with such predictions and, when based on longitudinal data, can provide information on dynamic soil properties. The process is summarized in fig. 1.

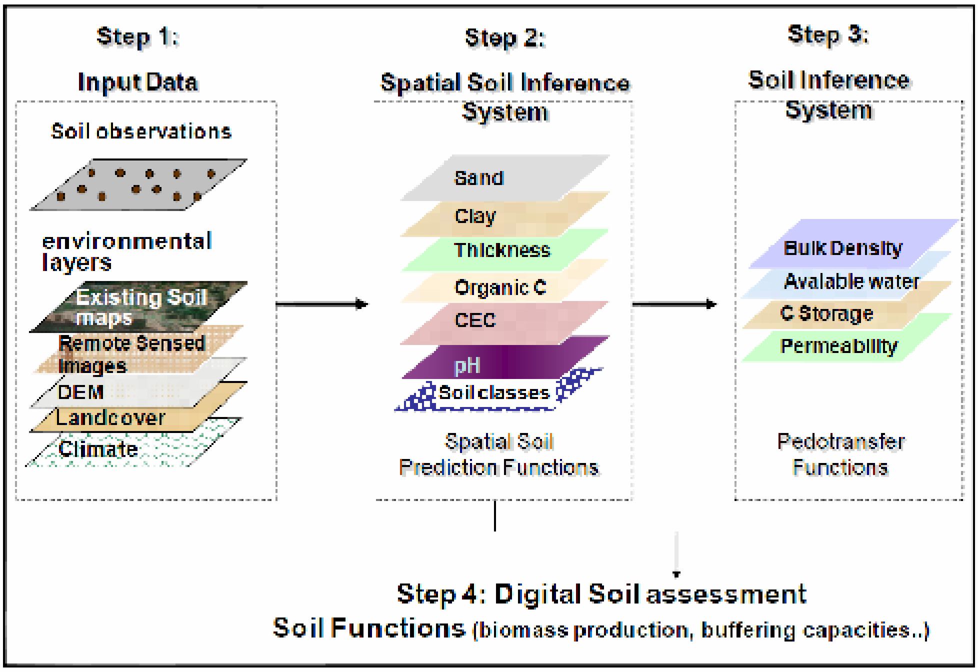

Figure 1. Digital soil mapping procedure [3]

1.2. Sampling design for soil mapping

Conventionally, soil sampling is carried out using different sampling designs. Common sampling designs are: Grid sampling: A grid with suitable spacing is placed on a landscape to be studied. Sites can be selected at intersections of the grid lines or within the grid cells. Grid sampling does provide equally spaced observations and it reveals any systematic variation across the tract under study. The drawback in geostatistical analysis is the equal distance between all sampling points. It should be noted that there is no randomization associated with grid sampling, therefore, the assumptions underlying several statistical analysis (e.g. ANOVA - analysis of variance) can not be fulfilled. Random sampling: Sample locations are selected at random, with equal probabilities of selection and independently from each other. The rationale is to exclude any form of bias, such as a conscious or even unconscious process of discriminatory selection on parts of the individuals. The technique has advantages of being statistically sound and unbiased, however, random samplings tend to cluster spatially (non-uniform density of observations per unit area and of dispersion of sites over the delineations) and are not likely to detect and measure systematic variation.

Random stratified sampling: The area is first divided into a number of sub-regions, called strata, and then random sampling is applied to each of the strata separately. The sample sizes in the strata may be chosen such that the probabilities of the locations of being sampled differ between strata. Transects : Soil samples are taken along straight lines across a landscape. The spacing between sampling points might be equal, nested, or random. Transect sampling reveals spatial variability along a line (often downhills), however, spatial variability in other directions is neglected.

Target sampling: Two or more attributes (e.g. topographic attributes such as slope, aspect, plan or profile curvature) are used to identify homogeneous and heterogeneous patterns. The goal is to identify 'representative sampling points'. This is a technique which minimizes the effort (costs) and maximizes the information content, on the assumption, that the sampling points are representative for the total data set (study area). It should be noted that there is no randomization associated with target sampling, therefore, the assumptions underlying several statistical analysis (e.g. ANOVA - analysis of variance) can not be fulfilled. Different sampling approaches must be used depending on the objectives, which are strongly influenced by scale. Each experimental design has constraints and strengths with regard to the analysis of data.

1.3. Effect of terrain characteristics on digital soil survey and sampling

Creating detailed soil information is necessary to meet the demands of ecological and environmental management systems [6, 7, 8]. The scale at which traditional soil surveys are created and the polygon data model used is often incompatible with other environmental data layers derived from digital terrain analysis and remote sensing techniques [7]. In addition, the process of manually creating conventional soil surveys is often a subjective one because it relies on the visual identification of landscape conditions through air photo interpretation for delineating soil–landscape units. Terrain characteristics, such as slope gradient, slope aspect, profile curvature, contour curvature computed from digital elevation model (DEM), are among the key inputs to digital soil surveys based on geographic information systems (GIS) (fig. 2).

The use of geographic information system based soil-mapping applications can resolve these limitations associated with traditional soil surveys by producing digital soil information at very fine scales, and by using an objective quantification of the landscape to characterize the soil formative environment [4]. In GIS-based soil-mapping applications, raster-based digital elevation models (DEM) are used to compute the terrain attributes, such as slope gradient, slope aspect, profile and contour curvature, which are required for characterizing a generalized soil-formative environment. Numerous authors have shown the need of terrain attributes derived from DEM for digital soil mapping [9, 10].

1.4. Soil-Land Inference Model (SoLIM)

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 [11, 12]. However, these techniques rely too heavily on the assumption of linearity and stationarity. It is unlikely that a direct linear relationship exists between terrain attributes and soil property values; in fact, the relationships between soil property variation and underlying terrain variables can be very complex [13]. The linearity and stationary assumption and the data requirements of these techniques present stiff challenges to their application over large and diverse landscapes.

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 [14].

As sampling is constrained by financial resources, efficient sampling strategies are desirable [15]. Increasingly available geospatial information (e.g. satellite imagery, geology maps, Digital Elevation Models) can be exploited as environmental covariates to optimize sampling locations [15] within the framework of a soil-landscape model [16, 17].

A number of recent papers [e.g. 14, 18] demonstrated the value of purposive mapping based on such covariates in producing more accurate predictions by using fewer, but more representative samples.

SoLIM, a soil land inference model, was developed to address the limitations of conventional soil survey [19]. The SoLIM approach employs recent developments in geographic information science (GISc), artificial intelligence (AI), and information representation theory to overcome these limitations. While the methods for deriving soils data are new, the model has its foundations in the soil factor equation of Dokuchaeiv [20] and Hilgard [21] and the soil-landscape model described by Hudson [22] which contend that if one knows the relationship between a soil and its environment, one can predict the occurrence of that soil in other areas having the same environment. SoLIM uses a suite of GIS and remote sensing techniques to characterize environmental conditions and knowledge acquisition techniques to extract and document soil landscape relationships from local soil experts. Environmental conditions are integrated with the extracted soil-landscape relationships to infer the spatial distribution of soil types under fuzzy logic. Implementation of SoLIM is shown in fig. 3.

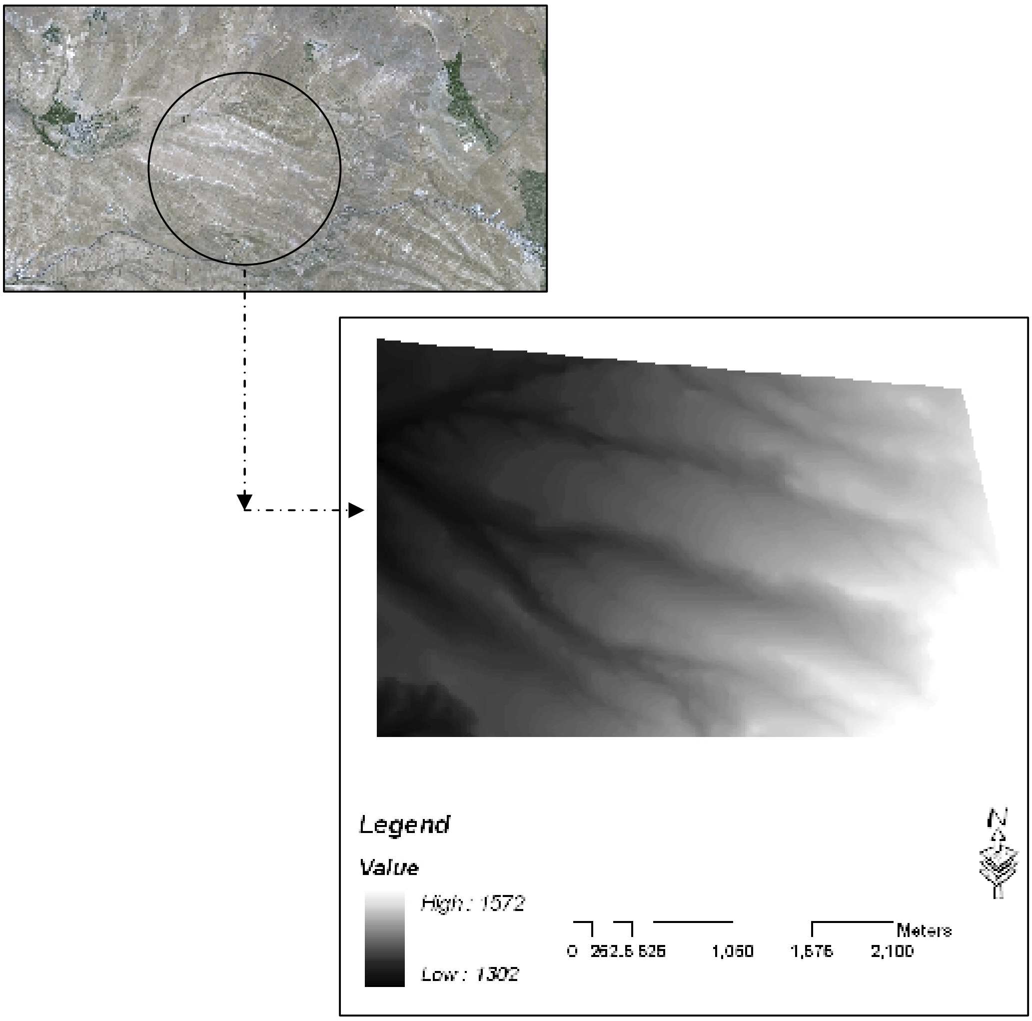

Figure 2. Digital elevation model of Kouhin region, central Iran (Developed by authors)

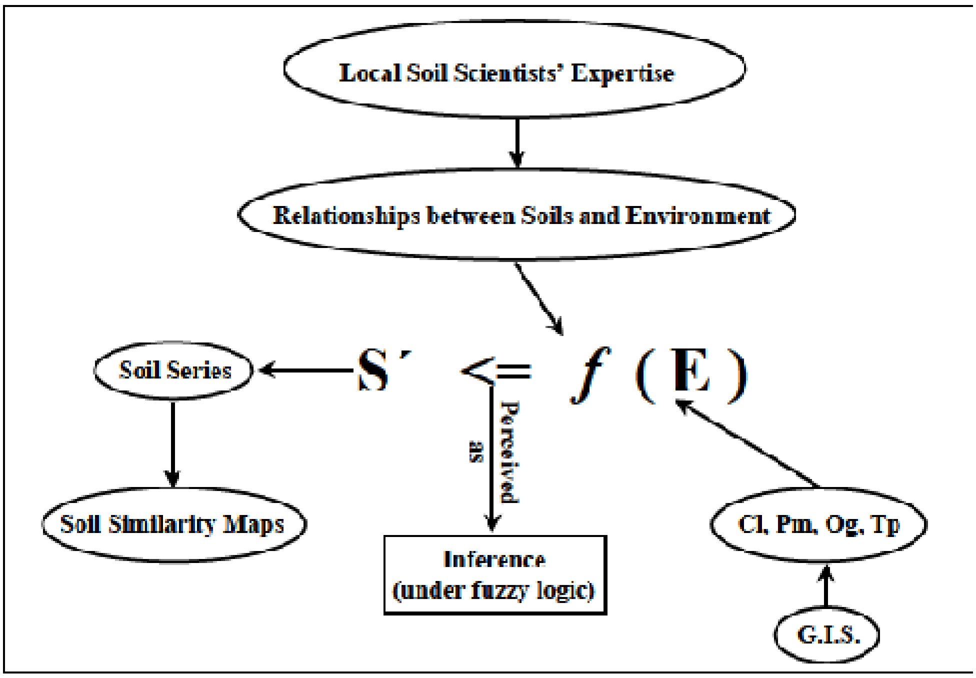

Figure 3. Implementation of SoLIM [18]

Data on soil formative environmental conditions (E) can be derived using GIS techniques (fig. 3). The variables used to characterize the soil-formative environmental conditions are decided based on the discussion between the person who conducts the knowledge acquisition (knowledge engineer) and the local soil expert(s). For a given area the local soil expert would provide an initial list of environmental variables to be considered. This list is modified by the knowledge engineer based on the data availability and the importance of the variables impacting the pedogenesis in the study area. Due to the data availability and difference in pedogenesis over different areas, there is no fixed list of environmental variables to be included. The list varies from area to area. Common data layers used to describe topography include elevation, slope aspect, slope gradient, profile and planform curvatures, upstream drainage area and wetness index, distance to streams, and distance to ridges. Bedrock and/or surficial geology data are necessary, but often not available at the level of details (fig. 4). The deficiency of geological data poses a major problem (it is a problem for manual mapping, too). Other data layers could include vegetation information derived from remotely sensed data such as LAI, tree canopy coverage, etc. It must be pointed out that the sufficiency and quality of environmental data layers will directly impact the quality of computed similarity values.

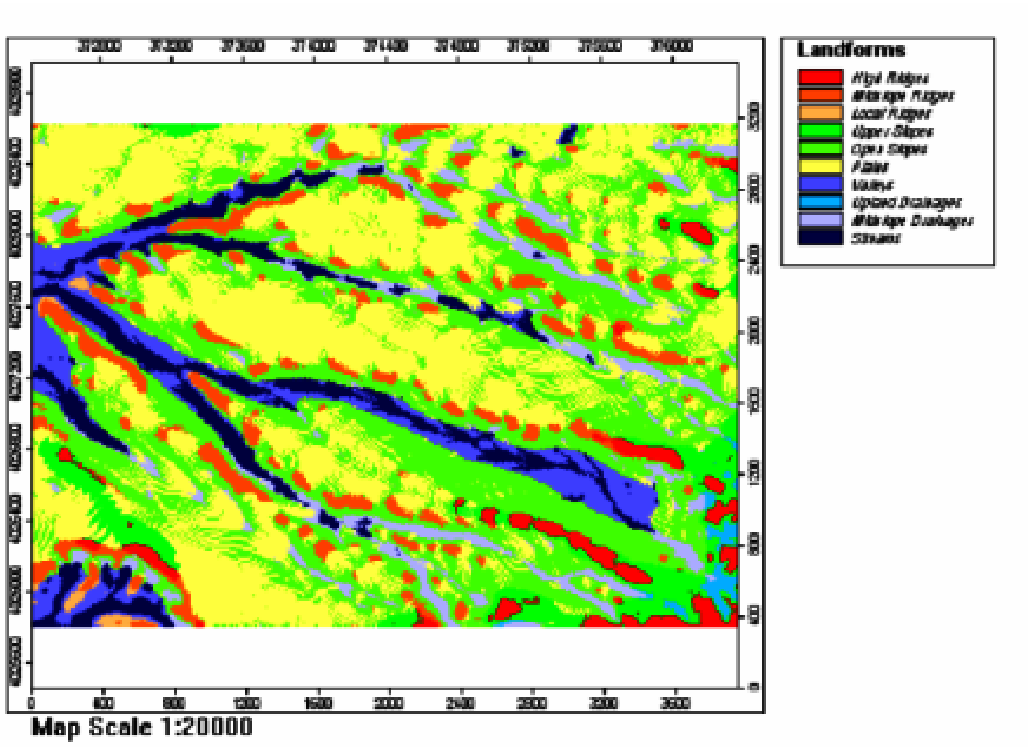

Figure 4. Topographic Position Index of Kouhin region, central Iran (Developed by authors)

The soil-environmental relationships (f) are approximated by the expertise of local soil scientists [23]. The acquired soil-environmental relationships can then be combined with data characterizing the soil formative environment conditions to infer S' under fuzzy logic [23, 24].

2. Conclusions

The aim of this study is to review the application of the SoLIM methodology with respect to environmental covariates (soil and terrain attributes) for sampling design optimization. Some of conclusions can be summarized as follows:

- Sampling design optimization based on Soil-Land Inference Model (SoLIM) is more efficient and less costly than the traditional soil sampling.

- During the SoLIM process, soil-landscape relationships are explicitly documented and stored for future use.

- The initial digital nature of SoLIM saves time and money that would otherwise be spent converting analog products to digital products and the high resolution raster dataset is more compatible with other sources of environmental data.

Acknowledgments

This study was carried out at the Laboratory of Remote Sensing and GIS, Department of Soil Science Engineering, University of Tehran, Karaj, Iran. The authors are grateful to anonymous reviewers for their valuable comments and suggestions.

References

[1] Brus, D.J. and Heuvelink, G.B.M. 2007. Towards a Soil Information System with quantified accuracy; Three approaches for stochastic simulation of soil maps. Wageningen, Statutory Research Tasks Unit for Nature and the Environment. 96 pp.

[2] Lagacherie P, McBratney AB, Voltz M. 2007. Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science 31 (Elsevier: Amsterdam).

[3] Minansy B, McBratney AB, Malone B, Sulaeman Y. 2010. Digital mapping of soil carbon. 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1 – 6 August 2010, Brisbane, Australia.

[4] McBratney AB, Mendonça Santos ML, Minasny B. 2003. On digital soil mapping. Geoderma, 117: 3-52.

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

[6] Park SJ, McSweeney K, Lowery B. 2001. Identification of the spatial distribution of soils using a process based terrain characterization. Geoderma 103: 249–272.

[7] Zhu AX, Mackay DS. 2001. Effects of spatial detail of soil information on watershed modeling. Journal of Hydrology 248: 54–77.

[8] Smith MP, Zhu AX, Burt JE, Stiles C. 2006. The effects of DEM resolution and neighborhood size on digital soil survey. Geoderma, 137: 58-69.

[9] Zhu AX. 1997. A similarity model for representing soil spatial information. Geoderma, 77: 217–242.

[10] McKenzie NJ, Gessler PE, Ryan PJ, O'Connell DA. 2000. The role of terrain analysis in soil mapping. In: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. John Wiley & Sons, New York, 245–265.

[11] Moore ID, Gessler PE, Nielsen GA, Peterson GA. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57:443-452.

[12] Gessler PE, Chadwick OA, Chamran F, Althouse L, Holmes K. 2000. Modeling soil-landscape and ecosystems properties using terrain attributes. Soil Sci. Soc. Am. J. 64:2046-2056.

[13] Lark RM. 1999. Soil-landform relationships at within-field scales: An investigation using continuous classification. Geoderma, 92: 141-165.

[14] Yang L, Zhu AX, Qi F, Qin C, Li B, Pei T. 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.

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

[16] MacMillan RA, Torregrosa A, Moon D, Coupé R, Philips N. 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.

[17] Drăguț L, Dornik A. 2013. Land-Surface Segmentation as sampling framework for soil mapping. Geomorphometry.org/2013.

[18] Zhu AX, Qi F, Moore A, Burt JE. 2010. Prediction of soil properties using fuzzy membership values. Geoderma, 158: 199-206.

[19] Zhu AX, 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.

[20] Glinka KD. 1927. The Great Soil Groups of the World and Their Development. (Translated from German by C.F. Marbut.) Edwards Brothers, Ann Arbor, Michigan.

[21] Jenny H. 1961. E.W. Hilgard and the Birth of Modern Soil Science. Farallo Publication, Berkley, California.

[22] Hudson BD. 1992. The soil survey as paradigm-based science. Soil Science Society of America Journal, 56: 836-841.

[23] Zhu AX, Band LE. 1994. A Knowledge-based Approach to Data Integration For Soil Mapping. Canadian Journal of Remote Sensing, 20: 408-418.

[24] Zhu AX., L.E. Band, B. Dutton, T.J., Nimlos 1996. Automated Soil Inference Under Fuzzy Logic. Ecological Modeling, 90: 123-145.