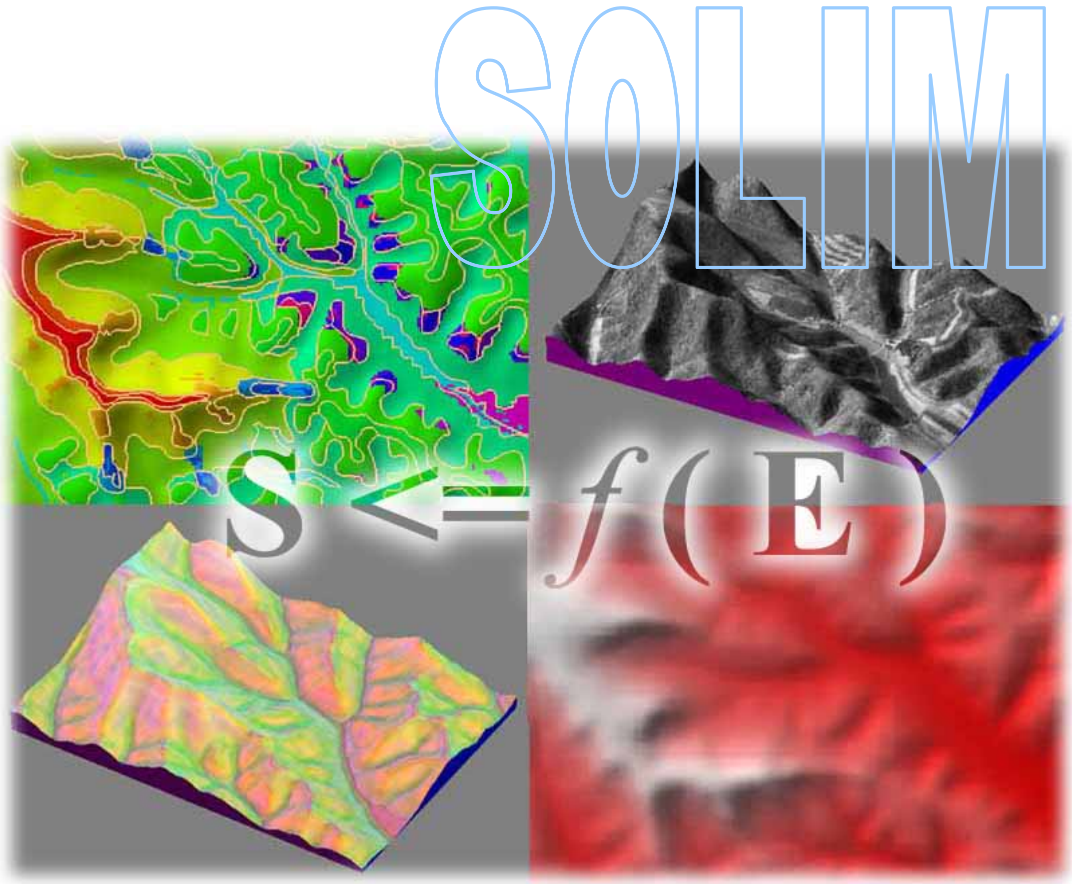
SoLIM: A New Technology For Soil Mapping Using GIS, Expert Knowledge & Fuzzy Logic

---

Overview



Prepared by

A-Xing Zhu[1], James E. Burt[1], Amanda C. Moore[2], Michael P. Smith[1], Jian Liu[1], Feng Qi[1]

[1]Department of Geography
University of Wisconsin-Madison

[2]Natural Resources Conservation Service
U.S. Department of Agriculture

Table of Contents

1. Introduction

SoLIM, a soil land inference model, was developed to address the limitations of conventional soil survey (Zhu, 1997; Zhu and Band, 1994; Zhu et al., 1996; Zhu et al., 1997; and Zhu et al,. 2001). 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 (Glinka 1927) and Hilgard (Jenny 1961) and the soil-landscape model described by Hudson (1992) 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.

This paper provides an overview of the SoLIM technology. Subsections present the traditional soil survey process and associated challenges, the similarity model designed to address these limitations, the pedological basis and methods for populating the similarity model, and the derivation of soil information products from SoLIM.

2. Traditional Soil Mapping Process and Its Challenges

2.1 Traditional Soil Mapping Process

The primary repository of soil spatial information is a currently traditional, polygon-based soil survey (Zhu et al., 2001). The traditional soil survey is largely a manual based process and is conducted over the course of several years using a combination of field reconnaissance and airphoto interpretation techniques to determine the spatial distribution of different types of soil. The theoretical foundation of traditional soil survey is that soil forms through the interaction among environmental factors (such as climate, parent material, vegetation, topography) over time. In other words, one can predict that locations that share the same environmental conditions will likely have the same soils. This assumption obviates the need to map soils by examining soil at every location across landscape.

To conduct soil survey under this assumption, soil scientists first investigate variability of soils over an area. This information is then used to develop map units. A map unit is “a collection of areas defined and named the same in terms of their soil components or miscellaneous areas or both” (Soil Survey Division Staff, 1993). Map units can contain predominately one type of soil (a consociation), multiple soils in a complicated pattern (a complex), other soils similar to the dominant soil type, and small areas of contrasting soils called inclusions. Once map units are defined, soil scientists formulate a soil-landscape model over the area to be mapped through direct observation of soils and the environment conditions. Once the landform and landscape position on which a given map unit occurs has been established, map units are delineated on aerial photographs through airphoto interpretation. Additional field observations are often conducted to check the accuracy of the delineations. Soil polygons from multiple photographs are compiled to orthorectified imagery to create a base map. Base maps are scanned, vectorized, and edge-matched for validation, archiving, and later use in a GIS. Each step is time consuming and can introduce errors into soil maps.

2.2 Issues and Limitations of Traditional Soil Surveys

Several issues affect the reliability and usefulness of traditional soil survey process and its products. First, traditional soil maps employ the polygon data model which simplifies the complex continuous distribution of soil types across a landscape to discrete polygons with definite boundaries. Spatial variation of soils within polygons is not captured and small bodies of soil are ignored. This results in generalization in both the spatial and parameter domains. Second, manual delineation of soil polygons is a very tedious, time-consuming and error prone process. Due to the due to the inherent nature of gradual variation of soils over space, it is not only difficult to place soil lines over areas of gradual variation, but also it is very easy for different soil scientists to place the soil line differently over the area. This inconsistency in manual soil mapping is very common. It is also quite possible that soil scientists can misplace a soil line owing to fatigue associated with stereoscope use. The misplacement of lines could also be due to the limitation of visual perception of landscape changes. For example, it is difficult for a human being to discern areas of 2% gradient from areas of 1% gradient. Thus, it would be quite easy to misplace a line which separates the two. Additional spatial and attribute errors can be introduced into soil maps during the lengthy compilation process. Each time soil boundaries are redrawn, either manually or digitally, there is a chance lines could be moved or polygons incorrectly labeled. Third, soil survey maps are the only documentation of the knowledge acquired during the traditional soil mapping process. Knowledge of soil-landscape relationships for a given area is not explicitly documented and is lost when soil scientists leave the survey area. New soil scientists are left to rediscover the soil-landscape relationships each time the survey area is revisited. Finally, updating traditional soil surveys as new data becomes available is time consuming since the data must be reinterpreted and maps must be redrawn manually.

The following sections elaborate the problems associated with the polygon model used in conventional soil survey, namely, the generalization of soils in the parameter (attribute) domain and the generalization of soil in the spatial domain. The polygon model problem is the core issue that the SoLIM approach was designed to overcome. The other two issues (the second and the third issues above) are addressed by the SoLIM through the process of populating a new model of soil information representation which SoLIM employs. The later two issues will be addressed in section 8.1.

2.2.1 Generalization in the Spatial Domain:

Spatial generalization occurs when detailed geographic features are simplified due to cartographic and map scale limitations (Zhu, 1997). Standard mapping and publication scales for soil surveys are 1:12,000 and 1:24,000. Minimum delineation (polygon) size is 0.4 – 4 hectares at 1:12,000 and 1.6 – 16 hectares at 1:24,000 according to accepted mapping standards (USDA-NRCS, 2001). Delineations smaller than the minimum polygon size are simply “included” in a larger polygon (Figure 2.1) and their actual spatial locations are lost. Subsequently, the spatial resolution of traditional polygon maps is simply the minimum delineation size (Zhu, 1997).

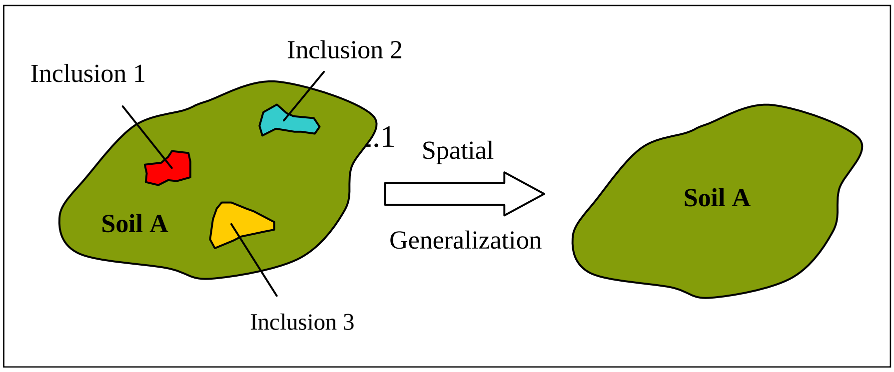


Figure 2.1: Spatial generalization of soils with smaller spatial extents

Generalization also occurs when two or more dissimilar components occur in a regularly repeating pattern and cannot be delineated separately at the mapping scale (Soil Survey Manual, USDA-NRCS). In this case, all of the components are grouped into a single map unit called a complex (Figure 2.2). Each component in a complex has a standard percent composition associated with it, though the actual percent composition of a component in a complex varies from polygon to polygon. The spatial distribution of each major soil type within a complex may be mentioned in the accompanying text, but is not explicitly recognized in the spatial data. Complexes can also contain inclusions, the spatial distribution of which are unknown as well.

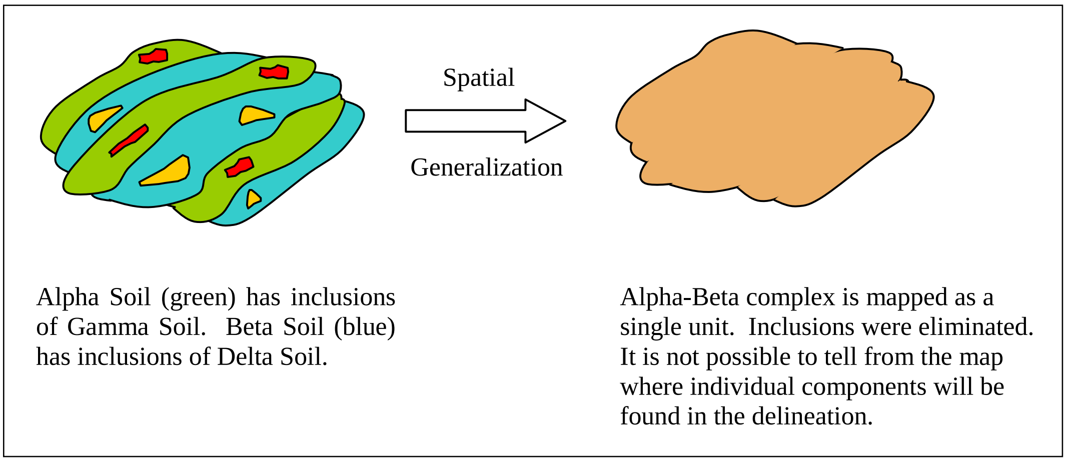


Figure 2.2: Spatial generalization of dissimilar soils which occur intermittently over space

Generalization results in soils data with significantly coarser resolutions (0.4 – 16 hectares for standard soil surveys) than most other data generated from digital terrain analysis or remotes sensing techniques (30 x 30 meters or better) (Zhu, 1997). Conventional soil maps are not capable of providing information about small but potentially important environmental niches that may be described by other higher resolution environmental data (such as digital elevation data and remotely sensed imagery) (Zhu, 1997). Soils spatial information typically has the lowest resolution of all data used for large scale (detailed) environmental modeling. This creates an incompatibility between information derived from soil map and other environmental data derived through digital terrain analysis and remote sensing techniques (Zhu, 2000). This incompatibility, subsequently, limits our ability to properly interpret results from environmental models (Zhu, 1997).

2.2.2 Generalization in the Parameter Domain

Soil characteristics vary continuously across a landscape in response to changes in environmental conditions. This variation is usually noticed during field mapping but can be difficult to quantify and nearly impossible to depict on traditional soil maps. Limitations of map scale and mapping processes have resulted in a simplification of the soil properties attributed to each soil type. Additionally, adoption of a vector data model to represent soils spatial information results in abrupt changes in soil properties at polygon boundaries.

The modal profile of each soil type, or soil series, is designated as the typical instance of that soil for the survey area. Characteristics of the typical instance as well as other examples of the same soil are used to determine the range of characteristics for the soil series. The range of characteristics for a given soil type contains all acceptable values for each soil property for the given soil type. It is used to approximate the value of a soil property at a given location in the survey area. If a single value is needed to describe a soil property, mean or median values from the range of characteristics are commonly used. Actual soil property values at a given location cannot be determined from soil survey data. Figures 2.3 is an example of the difference between the soil property values for profile depth as stated in a soil survey and soil property values as measured in the field.

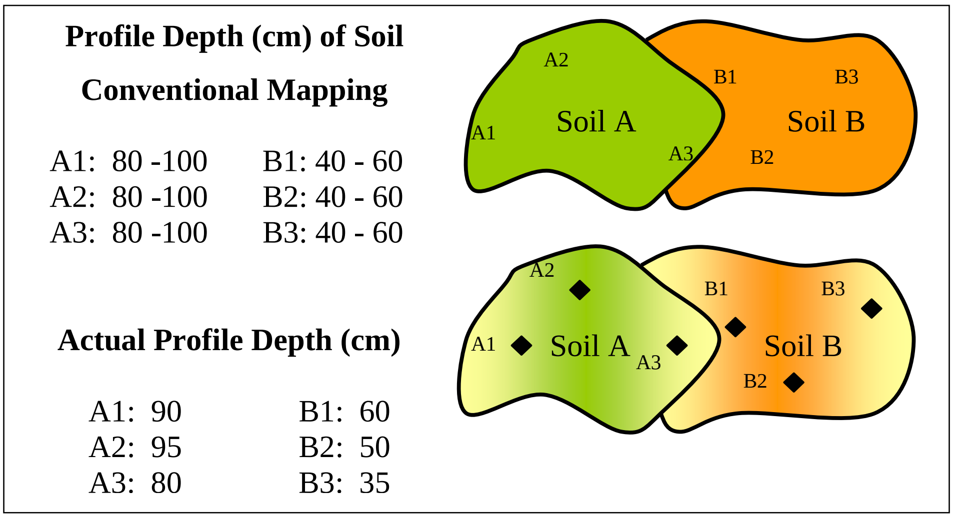


Figure 2.3: Generalization of soils through ignoring variability within soil polygons

Traditional soil maps portray spatial variation of soil as a step function (Figure 2.4) rather than showing the continuous gradation that is often observed over space. As such, soil property values are constant throughout a polygon and change abruptly at the boundary between polygons. Figure 2.4 shows an example of the difference between a measured soil property value, in this case soil depth, and the value stated in the soil survey. Note that properties measured at a specific location have a single value, while property values from the soil survey are given as a range.

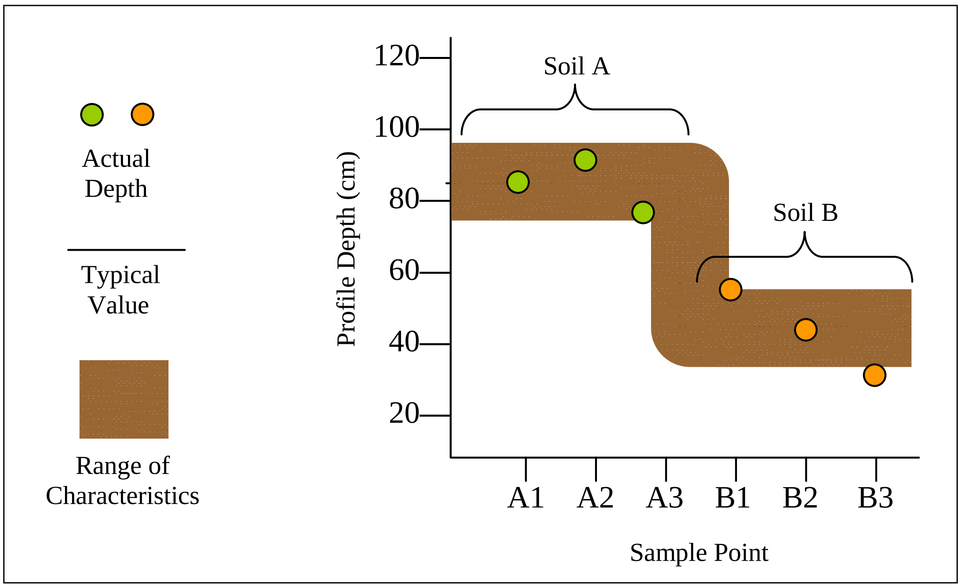


Figure 2.4: Step function of soil spatial variation as portrayed in soil maps

3. Overcoming the Soil Polygon Model: the Soil Similarity Model

3.1 Overcoming Spatial Generalization with a Raster Data Model

SoLIM uses a raster data model to address issues of spatial generalization. The raster data model is better suited to representing smooth, continuous geographic features and phenomena than the vector data model. The level of gradation captured by raster data model depends on the spatial resolution of the raster data model and in turn is limited only by the spatial resolution of the input data, rather than by arbitrary standards imposed by cartographic or mapping techniques (Zhu, 1997). Each pixel, at the spatial resolution of the input data, represents the type of soil found at that particular location and information about small pockets of unique soil types is not eliminated. This minimizes the discrepancies between the spatial resolution of soil spatial information and other environmental data layers (Zhu 1997). However, one important assumption for using the raster data model for representing soil spatial variation is that the soil within a given pixel is perceived to homogeneous, that is, the variation of soil properties within a pixel is so small that this variation can be ignored. This assumption holds if the pixel is small enough (in the limit pedon-sized). Otherwise, the assumption is violated and the spatial generalization problem associated with the polygon model will occur with this raster data model.

3.2 Overcoming Parameter Generalization with a soil similarity vector

SoLIM uses a similarity representation to address parameter generalization. The similarity representation of soils in the parameter domain is based on fuzzy logic (Zhu, 1997a). Under fuzzy logic, the soil at a given pixel can be assigned to more than one soil class with varying degrees of class assignment. These degrees of class assignment are referred to as ***fuzzy memberships***. This fuzzy representation allows a soil at each pixel to bear a partial membership in each of the prescribed soil classes. Each fuzzy membership is regarded as a ***similarity measure*** between the local soil and the typical case of the given class. All fuzzy memberships are retained in this similarity representation (Figure 3.1), which forms an n-element vector (***soil similarity vector***, or ***fuzzy membership 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 and the kth element, S_{ij}^k, in the vector represents the similarity value between the soil at pixel (i,j) and soil class k. With this similarity representation, the local soil at a given pixel is no longer necessarily approximated by the central concept (modal concept) of a particular class but can be represented as an inter-grade to the set of prescribed classes. This method of representation, which allows the local soil to take property values intermediate to the modal (typical) values of the prescribed classes, largely circumvents the problem of generalization in the parameter domain.

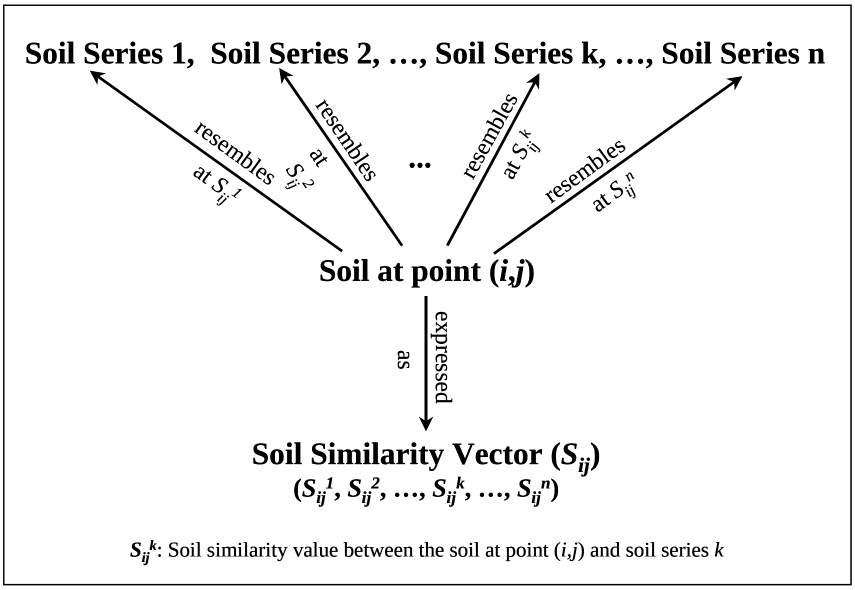


Figure 3.1: Fuzzy representation of soil information using soil similarity vector

By coupling the similarity representation with a raster GIS data model, soils in an area are represented as an array of pixels with soil at each pixel being represented as a soil similarity vector (Figure 3.2). In this way, soil spatial variation can be represented as a continuum in both the spatial and parameter domains.

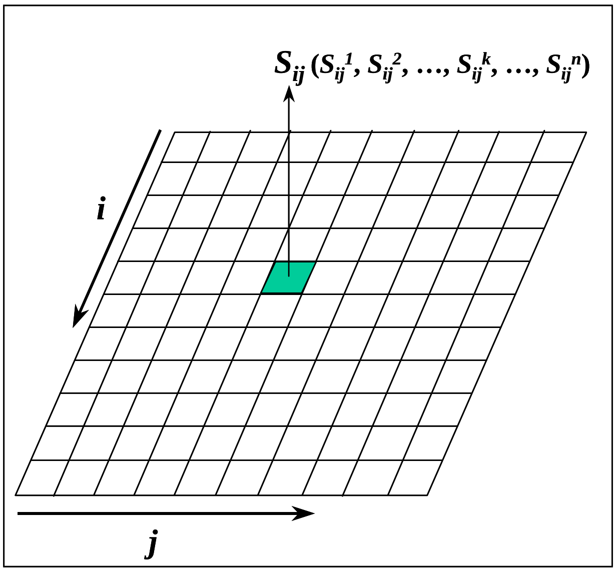


Figure 3.2: The similarity model for representing detailed soil spatial information

4. Populating the Similarity Model

The similarity model only provides added flexibility for representing soil spatial variation. The degree of success in using this model depends on how the model is populated or, equivalently, how the soil similarity values in the vector are determined at each pixel. It is impossible to use the conventional process to determine the soil similarity values at each pixel across landscape as there will be far too many pixels for manual determination. The SoLIM approach takes the advantages of recent development in geographic information science, artificial intelligence techniques and the classic concept of soil-landscape relationships to compute the soil similarity values at each pixel.

4.1 Pedological Basis

The pedological basis for populating the similarity model is the same as that used in conventional soil survey. It is the soil factor equation outlined by Dokuchaeiv (Glinka, 1927) and Hilgard (Jenny, 1961), which can be expressed as below:

$$S = \int f_1(E)dt \quad [1]$$

In Equation [1], t is time, and f_1 is the relationship of soil development to the formative environment, E, which generally includes variables describing climate, topography, parent material. S is meant to be soil which can be expressed as fuzzy membership value (soil similarity value).

In reality we cannot compute S since the exact form of f_1 is unknown to us at this moment. However, the soil landscape model as described by Hudson (1992) states that the distinctive interaction of the soil forming factors (climate, organisms, parent material, and topography over some period of time) leads to the formation of a unique soil or group of soils. From this, one can assume that locations experiencing the similar environmental conditions would have similar soils and one can further assume that the more similar the environmental conditions between two locations the more similar their soils. In other words, the similarity (S) between the soil at a given point to a given soil type can be approximated by the similarity (S') between the environmental conditions at that location and the typical environmental conditions of the prescribed soil type. The SoLIM approach employs these assumptions and uses S' to approximate S.

4.2 Implementation of the Pedological Basis

Due to the difficulty of explicitly describing integration of soil formative environmental factors over time during the course of soil formation across landscape, t is considered as part of E, thus, Equation [1] is simplified to:

$$S' = f(E) \qquad [2]$$

The implementation of Equation [2] is shown in Figure 4.1.

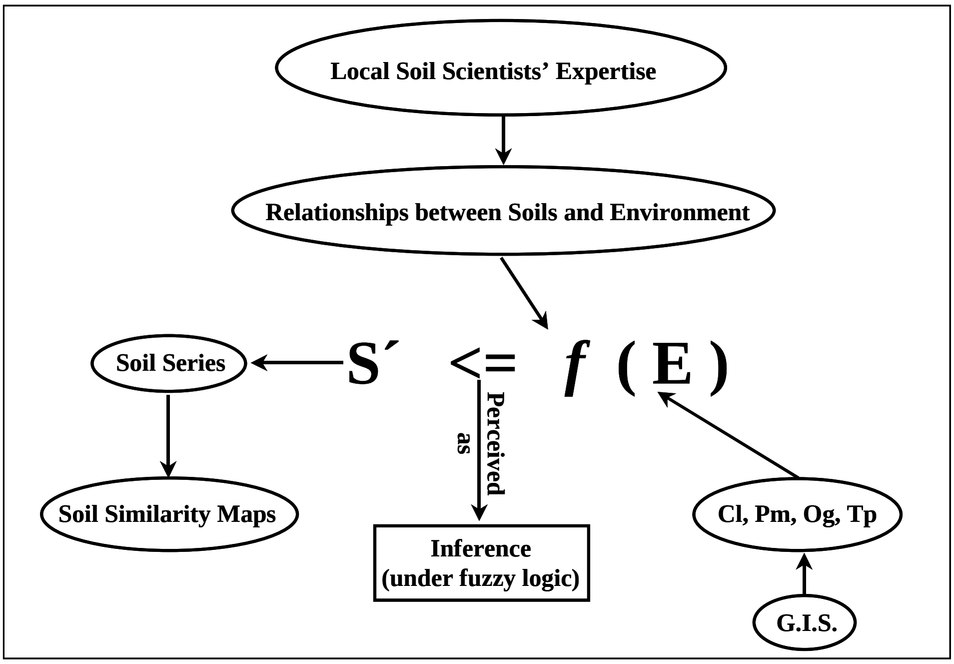


Figure 4.1: Implementation of SoLIM

Data on soil formative environmental conditions (E) can be derived using GIS techniques (Figure 4.1). 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. 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.

The soil-environmental relationships (*f*) are approximated by the expertise of local soil scientists (Zhu and Band, 1994; Zhu, 1999b). The acquired soil-environmental relationships can then be combined with data characterizing the soil formative environment conditions to infer *S'* under fuzzy logic (Zhu and Band, 1994; Zhu et al., 1996). The details of inferring *S'* is described in Section 6.

5. Knowledge Acquisition

Two basic type of knowledge regarding the relationships between soil and its formative environment conditions must be defined in order to compute the similarity (Figure 5.1). The first type of knowledge (referred to as Type I Knowledge) (Zhu, 1999) defines the typical environmental conditions under which a typical instance of a particular soil type would occur while the second type of knowledge (Type Two Knowledge) defines how similarity value will change as environmental conditions deviate from the typical environmental conditions. The process of obtaining these types of knowledge is called knowledge acquisition.

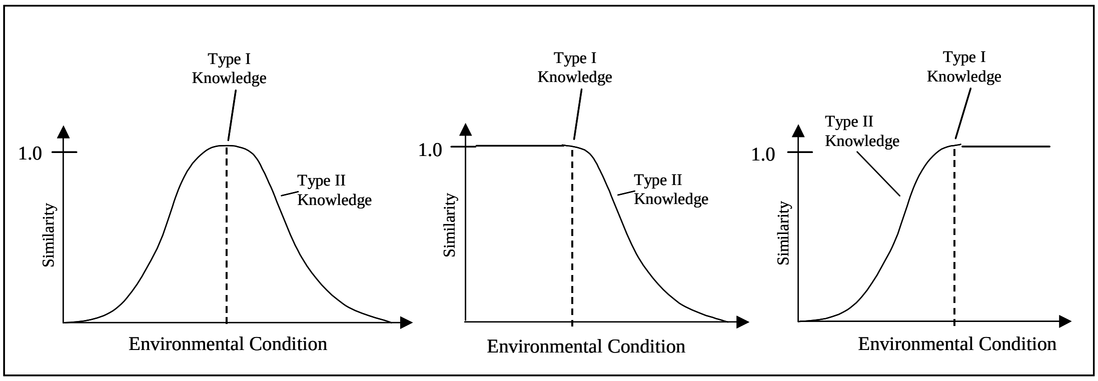


Figure 5.1: Type I and Type II Knowledge

5.1 Type I Knowledge

Type I Knowledge describes the typical combination of environmental conditions under which a typical instance of a particular soil type will form (Zhu, 1999). At this time, Type I Knowledge must be directly extracted from a local soil expert, though alternative means of developing Type I Knowledge are being explored (English 2001, Qi 2001). Type I Knowledge for a particular soil type might include the following information about formative environmental conditions:

Variable	**Value**
Aspect	South
Parent Material	Granite
Planform Curvature	Concave
Profile Curvature	Concave
Vegetation	Ponderosa pine
Slope	15-30%

Type I Knowledge is used to locate tacit points in the case-based reasoning inference process which will be discussed in section 6.

5.2 Type II Knowledge

Type II Knowledge describes the behavior of a soil type in response to deviation of environmental condition from its optimal (typical) setting (Zhu, 1999). Type II Knowledge is represented by a membership function with similarity values from zero to one on the Y-axis and the range in values for an environmental variable on the X-axis. The shape of the membership function approximates the manner in which the similarity changes as environmental value varies. There are three basic forms of membership functions: bell-shaped, S-shaped, and reverse S-shaped (Figure 5.2). The bell-shaped function describes that similarity value decreases when the environmental condition deviates from its optimal value (which is the Type I Knowledge); the S-shaped function describes that similarity decreases when the value of the environmental condition is below a given value (the Type I Knowledge) and continues to decrease. The reverse S-shaped function shows that similarity value increases as the value of environmental condition decreases and it reaches unity after the value of environmental value reaches a certain value (the Type I Knowledge). Various membership curves can be derived from each basic form by changing the steepness of the curves. These membership functions are used during the inference process to compute the similarity values. The accuracy of the inference is dependent upon the quality of the membership functions. Type II Knowledge can be extracted explicitly from a local soil expert or can be approximated by pre-defined basic functions

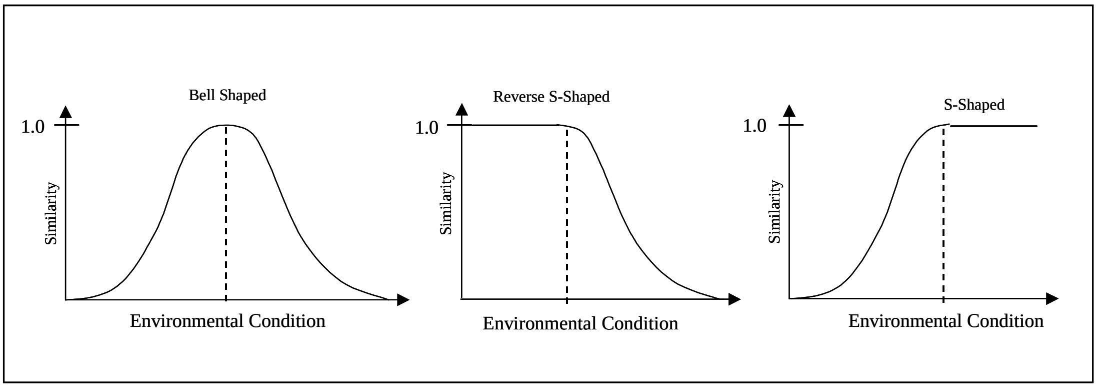


Figure 5.2: The three basic forms of membership functions

6. The Inference Process

The actual process of inferring *S'* is automated (Zhu and Band, 1994). The acquired soil-environmental relationships are stored in a database (referred to as a ***knowledgebase***). Data characterizing soil formative environments are stored in a GIS database. A set of inference techniques constructed under fuzzy logic (collectively called the fuzzy inference engine) is used to link the knowledgebase with the GIS database to derive soil similarity vectors (Figure 6.1). In general, for pixel (i,j), the inference engine takes the data on soil formative environment conditions for that pixel from the GIS database and combines the GIS data with the soil-environment relationships for soil category k from the knowledgebase to calculate the similarity value of the local environment to the typical environment of soil category k, S'^{k}_{ij}, which is then used as a surrogate for S^{k}_{ij}. Once all of the soil categories are exhausted by the inference engine the soil similarity vector ($\boldsymbol{S}_{ij}$) for this pixel is determined (fully populated with values). The inference engine then moves onto the next pixel in the GIS database and repeats the process of deriving the soil similarity vector for that pixel. When all pixels in the GIS database are exhausted, a similarity representation of soils (a raster soil database) for the entire area has been derived. The SoLIM approach relates the formative environmental conditions that are characteristic of a given soil type to the conditions that are present at a specific location through the equation presented above to determine the similarity of the soil at a specific location to other known soil types (Zhu et al. 2001). The specifics of the inference process are described in details in Zhu and Band (1994).

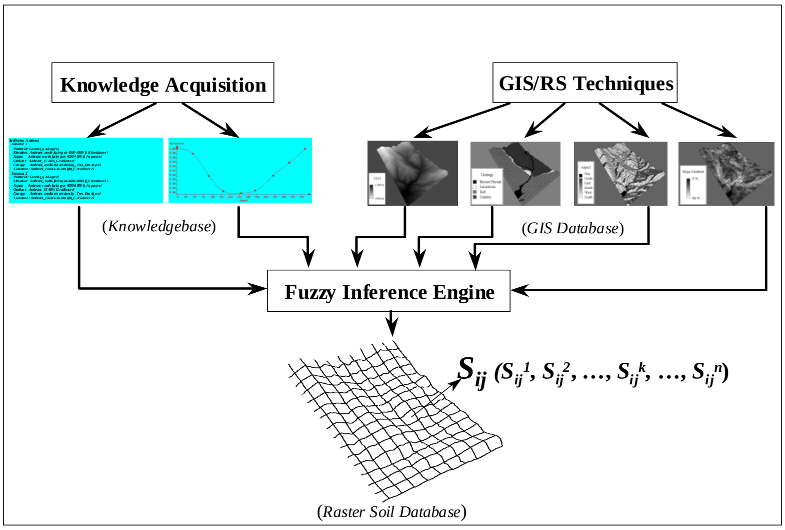


Figure 6.1: The inference process

7. Deriving Soil Information Products

Two kinds of soil information products can be generated from SoLIM: soil map products and soil knowledge products. Soil map products include fuzzy membership maps, categorical raster maps, conventional polygon maps, and other products. These products are derived from soil similarity vectors calculated during the inference process. Soil knowledge products include catenary sequences, dichotomous keys, soil-environment descriptions, and fuzzy membership functions and are developed during the knowledge acquisition and inference processes. Soil map products will be discussed in this section.

7.1 Fuzzy Membership Maps

Fuzzy membership maps are the initial product generated by the inference process. A fuzzy membership map is generated for each soil type in a survey area. Figure 7.1 shows a membership map for a single soil type; the inference process generates as many maps like Fig 7.1 as there are soil types in the area. Each map contains the similarity value of every pixel to a single soil type, thus showing the continuous gradation of membership for that soil type across the landscape. To find the similarity vector of a particular pixel, the similarity values of that pixel are extracted from the fuzzy membership maps for all soils type in a survey area.

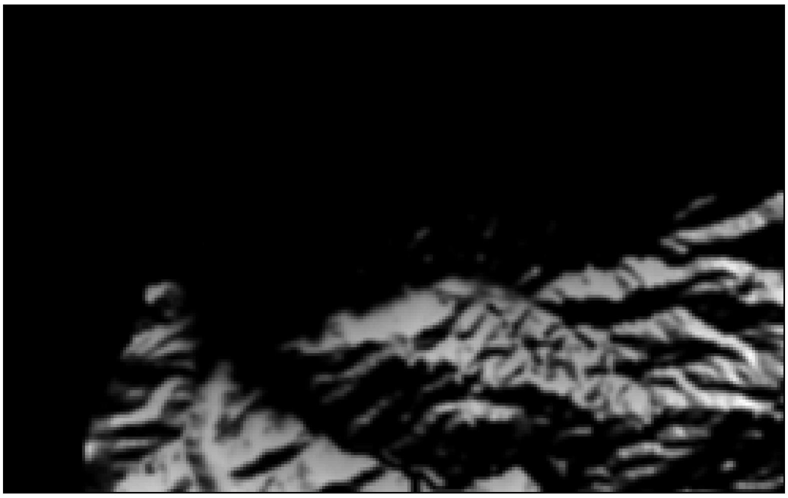


Figure 7.1: An example of fuzzy membership map

7.2 Raster Soil Categorical Maps

Raster soil categorical maps are essentially soil type maps which are created by hardening the fuzzy membership maps. The hardening is done by assigning the soil type with the highest membership value at each pixel. Individual soil bodies can be as small as a single pixel, or can contain numerous adjacent pixels of the same soil type. Raster soil categorical maps contain more spatial detail than conventional polygon maps since cartographic constraints are not an issue. Raster soil categorical maps could potentially be substituted for conventional polygon maps.

7.3 Conventional Polygon Maps

Conventional polygon maps can also be generated from a SoLIM-derived raster categorical map. The process starts by finding bounding lines for raster areas with a common soil type. Any resulting polygons smaller than a specified minimum size are filtered out. Clearly, many small soil bodies are filtered out during the conversion from raster soil categorical maps to polygon maps. The primary benefit of deriving polygon maps from SoLIM rather than through traditional soil survey methods is that the actual composition (percent of each different soil type) of each polygon is known.

8. Assessment of the SoLIM Methodology

As with any new technology, the SoLIM has many advantages and some limitations. SoLIM has the potential to advance significantly the soil survey process (Zhu et al. 2001); however, its success can be limited by the quality of GIS and knowledge inputs.

8.1 Advantages of the SoLIM Methodology

There are several advantages of the SoLIM methodology compared to the traditional soil survey process. First, the SoLIM process results in more accurate soil maps than traditional soil survey maps. Second, the SoLIM process is more efficient and less costly than the traditional soil survey update process. Third, during the SoLIM process, soil-landscape relationships are explicitly documented and stored for future use. Fourth, 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. Finally, a suite of soil information products can be generated from fuzzy membership maps, depending on the data needs of an individual or organization.

8.1.1 Mapping Consistency and Accuracy

The automated mapping techniques used in SoLIM can apply a soil-landscape model more consistently across a landscape than human soil scientists (Zhu et al. 2001). It follows that SoLIM consistently identifies the soil type that occurs under a given set of environmental conditions and multiple soil maps produced by SoLIM for areas using the same soil-landscape model will be consistent with one another. The accuracy of the inference and the derived hardened soil maps, however, depends upon the quality of the GIS database and soil-landscape model used in the inference.

Soil maps generated by SoLIM have a significantly higher spatial resolution and better represent the actual distribution of various soil types than traditional soil survey maps. SoLIM uses a raster data model to depict soil spatial data as raster data models are better suited to representing smooth, continuously varying data surfaces than vector data models. Spatial resolution of SoLIM products is limited only by the resolution of the original input data.

Hardened soil maps produced by SoLIM can be considerably more accurate than their conventional counterparts. For one study site in western Montana (Zhu et al. 2001, Zhu 1997, Zhu 1996, Zhu and Band 1994), SoLIM maps correctly predicted the soil type at 52 of 64 sample sites (81% accuracy), while conventional soil survey maps only predicted the soil type at 39 sites correctly (61% accuracy). Another study site in southwestern Wisconsin (Zhu et al. 2001) shows similar results. SoLIM correctly predicted the soil type at 83 of 99 Wisconsin sample sites (~83% accuracy) while the conventional soil survey only predicted 66 of 99 sites correctly (~67% accuracy).

The greater accuracy of soil information products generated from SoLIM is related to a number of factors. First, GIS can capture highly detailed information on the variation of environmental conditions since digital data processing capabilities allow many variables to be considered simultaneously (Zhu et al. 2001). This makes it possible to reduce the number of soil inclusions and avoid misinterpreting soil type. Second, SoLIM allows local soil conditions to be expressed at the pixel level, thus reducing the amount of spatial generalization that typically occurs on conventional polygon maps (Zhu et al. 2001). Finally, the use of fuzzy logic to determine local soil conditions allows the soil at a pixel to be represented in terms of its similarity to multiple soil types rather than being forced into a single discrete category (Zhu et al. 2001), thus enabling a more accurate estimation of soil conditions at a pixel.

8.1.2 Time and Cost Savings for Soil Survey Updates

Use of the SoLIM methodology can result in significant time and cost savings compared to traditional soil surveys. The GIS database, the knowledgebase, and the fuzzy inference engine are all reusable and independently updatable (Zhu et al. 2001). GIS information used in the SoLIM process can be updated as more accurate or dependable information becomes available. When new information – such as a higher resolution digital elevation model – becomes available, an existing knowledgebase can be reapplied to produce an updated soil survey in a matter of days or weeks rather than months or years (Zhu et al. 2001). A SoLIM knowledgebase can be updated as a better understanding of the soil-landscape relationships for that area is gained. For example, if a soil scientist learns that a certain environmental condition has a greater influence on soil development than was previously thought, the knowledgebase can be updated to reflect the new information. The updated knowledgebase can be applied to an existing GIS database, again quickly producing an updated soil survey. Finally, updated soil surveys can also be produced by processing existing knowledgebases and GIS databases with an improved inference engine.

The time savings associated with the SoLIM can be translated into cost savings as well. Traditional soil surveys can take years, and sometimes decades, to complete. In extreme cases, soil surveys can be out-of-date before they are even published. Updating an existing soil survey essentially involves redoing the entire survey, from developing the soil-landscape model to redrawing and digitizing the maps. The time and effort involved in this process is considerable, and consequently soil surveys and soil survey updates are quite expensive. Because each component of the SoLIM process persists once it has been developed and can be easily updated as new information becomes available, soil survey updates can be performed much more quickly and therefore less expensively using the SoLIM.

8.1.3 Explicit Documentation of Soil-Landscape Relationships

A significant portion of local expertise regarding the spatial distribution and physical characteristics of soil types and the relationship between soil types and environmental conditions is lost each year as experienced local soil scientists retire (Zhu et al. 2001). Unfortunately, there has not been a procedure in place for capturing this knowledge and passing it along to new soil scientists. The SoLIM provides a means of quantifying and storing the relationships between environmental conditions and soil types. This knowledgebase is well documented and can be easily updated. This documentation can serve as an important resource for new soil scientists as they begin to learn the process of identifying soil-landscape relationships and conducting a soil survey. Additionally, explicit documentation of knowledge can increase the consistency of soil-landscape models that span multiple generations of soil scientists (Zhu et al. 2001).

8.1.4 Digital Soil Information Products

A variety of soil information products can be derived from the soil similarity vectors. Products include hardened soil maps and soil property maps, in addition to fuzzy membership maps. Soil information products generated from the SoLIM methodology are already in digital format, thus enabling to data to be used directly in a GIS without requiring the lengthy, expensive, and potentially error prone digitization process (Zhu et al. 2001). These digital products are also more compatible with existing environmental data than conventional soil maps as the spatial resolution of SoLIM data is limited only by the spatial resolution of the original input data (Zhu 1997).

8.2 Current Limitations of the SoLIM Methodology

The quality of soil information produced using the SoLIM is dependent upon the quality of the soil-landscape model and on the accuracy of environmental conditions characterized in GIS (Zhu et al. 2001). In the absence of a detailed, accurate soil-landscape model, it is not possible to produce accurate soil information. Currently, extracting information from local soil experts is the only method of developing a soil-landscape model (Zhu et al. 2001). Methods such as fuzzy *c*-means clustering and data mining need to be further explored as alternative options for knowledge acquisition when extracting knowledge from local soil experts is not feasible.

The accuracy of environmental conditions characterized using GIS is related to the availability of data layers, the quality and resolution (spatial and attribute) of existing data, and the ability to define relevant environmental conditions using GIS (Zhu et al. 2001). Where environmental gradients across a survey area are small and the relationships between soil types and environmental conditions are subtle, higher spatial resolution of the original input data may be needed to produce satisfactory results. Work is in progress to examine the performance of SoLIM in areas with very gentle environmental gradients (Zhu et al. 2001).

9. General Steps in Using SoLIM in Soil Survey

The following general steps should be followed while conducting soil survey using SoLIM:

1) Prepare the basic information: a list of soil types and a basic GIS database (DEM with orthophoto)
2) Extract the knowledge from the local soil scientists
3) Complete the GIS database compilation
4) Perform the inference
5) Validate the result with the soil scientists
6) Document the results

10. References

Brady, N.C. 1990. *The Nature and Properties of Soils*, 10th ed. New York: MacMillan Publishing Co. 621pp.

Buol, S.W., F.D. Hole and R. J. McCracken. 1989. *Soil Genesis and Classification*, 3rd ed. Ames, Iowa: Iowa State University Press. 446pp.

Burrough, P. A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40:477-492.

Burrough, P. A., R.A. MacMillan, and W. van Deursen. 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. Journal of Soil Science 43:193-210.

English, E.M. 2001. *Assisting Knowledge-Based Inferential Soil Mapping: The Application of Fuzzy c Means Clustering to Expose Environmental Niches.* Master's thesis. University of Wisconsin, Madison, Wisconsin.

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

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

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

Jenny, H. 1941. *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw-Hill. 281pp.

MacMillan, R. A., W.W. Pettapiece, S.C. Nolan, and T.W. Goddard. 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Sets and Systems 113:81-109.

Qi, F. 2001. A *Data Mining Approach to Knowledge Discovery from Soil Maps.* Master's thesis. University of Wisconsin, Madison, Wisconsin. 88pp.

Shi, X. 2002. *A Case-based Reasoning Approach to Fuzzy Soil Mapping*. Ph.D.Dissertation. University of Wisconsin, Madison, Wisconsin.

Soil Survey Division Staff. 1993. *Soil Survey Manual*. Soil Conservation Service. U.S. Department of Agriculture Handbook 18. 437pp.

U.S. Department of Agriculture, Natural Resources Conservation Service. 2001. National Soil Survey Handbook, title 430-VI. [Online] Available: http://soils.usda.gov/procedures/handbook/main.htm.

Zhu, A.X, B. Hudson, J. Burt, K. Lubich, and D. Simonson. 2001. Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. Soil Science Society of America Journal 65:1463-1472.

A.X. Zhu. 2000. "Mapping soil landscape as spatial continua: the neural network approach". *Water Resources Research*, Vol. 36, No. 3, pp. 663-677.

Zhu, A.X. 1999. A personal constructed-based knowledge acquisition process for natural resource mapping using GIS. International Journal of Geographic Information Systems 13:93-118.

Zhu, A.X. 1997. A similarity model for representing soil spatial information. Geoderma 77:217-242.

Zhu, A.X., L.E. 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.

Zhu, A.X., L.E. Band, B. Dutton, and T.J. Nimlos. 1996. Automated Soil Inference Under Fuzzy Logic. Ecological Modeling 90:123-145.

Zhu, A.X. and L.E. Band. 1994. A Knowledge-based Approach to Data Integration For Soil Mapping. Canadian Journal of Remote Sensing 20:408-418.

Zhu, A.X. 1994. *Soil Pattern Inference Using GIS Under Fuzzy Logic*. Ph.D. Dissertation. University of Toronto, Toronto, Ontario. 177pp.