Reprinted from

GEODERMA

Geoderma 77 (1997) 217–242

A similarity model for representing soil spatial information

A-Xing Zhu *

Department of Geography, University of Wisconsin, 550 N. Park St. Madison, WI 53706, USA

Received 29 February 1996; accepted 27 February 1997

ELSEVIER

GEODERMA

AN INTERNATIONAL JOURNAL OF SOIL SCIENCE

EDITORS-IN-CHIEF

H. Insam, Innsbruck
A.B. McBratney, Sydney, N.S.W.
K. McSweeney, Madison, Wisc.
D.L. Sparks, Newark, Del.

HONORARY EDITOR

R.W. Simonson, Oberlin, Ohio

EDITORIAL BOARD

J. Bouma, Wageningen
R.J. Coventry, Townsville, Qld.
K. Dalsgaard, Arhus
R.B. Daniels, Raleigh, N.C.
F. de Coninck, Ghent
J. de Gruijter, Wageningen
J.M. de Villiers, Pietermaritzburg
D. Dent, Norwich
D.P. Franzmeier, West Lafayette, Ind.
K. Haider, Deisenhofen
R. Horn, Kiel
P.M. Huang, Saskatoon, Sask.
R.F. Isbell, Townsville, Qld.
H.E. Jensen, Copenhagen
E.F. Kelly, Fort Collins, Colo.
K. Kyuma, Hikone
W.L. Lindsay, Fort Collins, Colo.
A.R. Mosier, Fort Collins, Colo.
M. Oades, Glen Osmond, S.A.
Ya.A. Pachepsky, Moscow
R.L. Parfitt, Palmerston North
M. Robert, Versailles
M. Schnitzer, Ottawa, Ont.
U. Schwertmann, Freising
N. Senesi, Bari
S. Shoji, Sendai
G. Sposito, Berkeley, Calif.
K. Stahr, Stuttgart
G. Stoops, Ghent
L. Stroosnijder, Wageningen
R.S. Swift, Glen Osmond, S.A.
P.B. Tinker, Oxford
G.C. Topp, Ottawa, Ont.
N. van Breemen, Wageningen
E. van Ranst, Ghent
A. van Wambeke, Ithaca, N.Y.
Gy. Várallyay, Budapest
M.J. Vepraskas, Raleigh, N.C.
H. Vereecken, Jülich
R.J. Wagenet, Ithaca, N.Y.
G.J. Wall, Guelph
A.W. Warrick, Tucson, Ariz.
D.H. Yaalon, Jerusalem

Scope of the journal

The primary intention of the journal is to stimulate wide interdisciplinary cooperation and understanding among workers in the different fields of pedology. Therefore, the journal tries to bring together papers from the entire field of soil research, rather than to emphasize any one subdiscipline. Interdisciplinary work should preferably be focused on occurrence and dynamic characterization in space and time of soils in the field.

Publication information

Geoderma (ISSN 0016-7061). For 1997 volumes 75–80 are scheduled for publication. Subscription prices are available upon request from the publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries where air delivery via SAL is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, PR China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries airmail rates are available upon request. Claims for missing issues must be made within six months of our publication (mailing) date. For orders, claims, product enquiries (no manuscript enquiries) please contact Customer Support Department at the Regional Sales Office nearest to you:

New York, Elsevier Science, P.O. Box 945, New York, NY 10159-0945, USA. Tel: (+1) 212-633-3730, [Toll Free number for North American Customers:1-888-4ES-INFO (437-4636)], Fax: (+1) 212-633-3680, E-mail: usinfo-f@elsevier.com
Amsterdam, Elsevier Science, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. Tel: (+31) 20-485-3757, Fax: (+31) 20-485-3432, E-mail: nlinfo-f@elsevier.nl
Tokyo, Elsevier Science, 1-9-15, Higashi-Azabu 1-chome, Minato-ku, Tokyo 106, Japan. Tel: (+81) 3-5561-5033, Fax: (+81) 3-5561-5047, E-mail: kyf04035@niftyserve.or.jp
Singapore, Elsevier Science, No. 1 Temasek Avenue, #17-01 Millenia Tower, Singapore 039192. Tel: (+65) 434-3727, Fax: (+65) 337-2230, E-mail: asiainfo@elsevier.com.sg

US mailing notice – *Geoderma* (ISSN 0016-7061) is published montly by Elsevier Science B.V. (Molenwerf 1, Postbus 211, 1000 AE Amsterdam). Annual subscription price in the USA US$ 1667 (US$ valid in North, Central and South America), including air speed delivery. Application to mail at periodicals postage rate is pending at Jamaica, NY 11431.
USA POSTMASTERS: Send address changes to *Geoderma*, Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003.
Airfreight and mailing in the USA by Publications Expediting.

Advertising information:

Advertising orders and enquiries may be sent to: *International:* Elsevier Science, Advertising Department, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK; Tel.: (+44) (0) 1865 843565; Fax: (+44) (0) 1865 843976. *USA and Canada*: Weston Media Associates, Dan Lipner, P.O. Box 1110, Greens Farms, CT 06436-1110, USA; Tel.: +1 (203) 261 2500; Fax: +1 (203) 261 0101. *Japan*: Elsevier Science Japan, Marketing Services, 1-9-15 Higashi-Azabu, Minato-ku, Tokyo 106, Japan; Tel.: +81 3 5561 5033; Fax: +81 3 5561 5047.

ELSEVIER

Geoderma 77 (1997) 217–242

GEODERMA

A similarity model for representing soil spatial information

A-Xing Zhu *

Department of Geography, University of Wisconsin, 550 N. Park St. Madison, WI 53706, USA

Received 29 February 1996; accepted 27 February 1997

Abstract

A fuzzy logic based model (called a similarity model) was developed to represent soil spatial information so that soil landscape is perceived as a continuum in both the parameter space and the geographic space. The similarity model consists of two components: the similarity representation component and a raster representation scheme. The similarity representation component uses a set of prescribed soil taxonomic categories as the central concepts of the fuzzy soil classes and represents a soil at a given location as a set of similarity values to these central concepts. The collection of these similarity values forms an *n*-element vector called a soil similarity vector. With the use of a raster representation scheme, soil spatial information over an area can be represented as an array of soil similarity vectors. This similarity model has two main advantages for representing spatial soil information over conventional polygon-based soil maps. Firstly, the details of soil spatial information can be represented at the resolution of a raster data model rather than at the minimal mapping sizes as in conventional polygon-based soil maps. Secondly, under the similarity representation, the deviation of a soil at a given location from typical soil classes can be preserved and its properties can then take values intermediate to the typical values of the prescribed soil types. A case study conducted in the Lubrecht Experiment Forest of western Montana demonstrated that soil spatial information represented under the similarity model has a higher resolution at both the attribute level and the spatial level than that in the conventional soil map of the area.

Keywords: classification; environment; fuzzy logic; geographic information systems; models; soils

* Tel.: +1 (608) 262-2138; Fax: +1 (608) 265-3991; E-mail: azhu@geography.wisc.edu

0016-7061/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0016-7061(97)00023-2

1. Introduction

Information about soil spatial distribution is an essential part of land surface attributes required for a variety of environmental modeling activities since soil mediates the energy and material fluxes at the earth's surface. Soil maps produced from conventional soil surveys are a major source of soil spatial information. There are two major inter-related limitations for using the soil spatial information derived from conventional soil maps with other environmental data such as digital terrain and remotely sensed data for environmental modeling at large scales (Band et al., 1993; Moore et al., 1993; Band and Moore, 1995). These two limitations are, firstly, attribute resolution incompatibility and spatial resolution incompatibility between the soil information derived from conventional soil maps, and secondly, environmental data derived from digital terrain analyses and remote sensing techniques.

This paper presents a model (a similarity model) for representing soil spatial information at the level of detail compatible with data from digital terrain analyses and remote sensing techniques. The model is based on fuzzy logic and a raster scheme for representing spatial data. It can be considered as a continuous spatial model (Bregt, 1992), or a layer model (field model) (Goodchild, 1989, 1992, 1993). The soil at a given point is represented by a vector of membership values which describe the degrees of similarity of the local soil to a prescribed set of soil taxonomic units (classes). Each element in the vector represents the similarity (membership) of the local soil to a prescribed soil taxonomic unit. In this way, the local soil does not need to be assigned to one and only one soil category. The deviation of the local soil from typical soil categories can be preserved by the varying membership values in the vector. Local soil properties can then take the values intermediate to the typical values of the prescribed soil classes. Using a raster representation scheme, soil spatial information over an area can be represented as an array of these vectors with each of these vectors corresponding to each location in the area. The spatial detail (spatial resolution) of soil spatial information would then be compatible with other environmental data.

The paper first discusses the two incompatibilities between the soil information from conventional soil maps and the other detailed environmental data. This is followed by a brief overview of current research efforts to overcome the limitations of the existing scheme for representing soil spatial information. In Section 4, the proposed similarity model is then presented and discussed, which is followed by a case study to further illustrate the concepts of this model and to demonstrate the potential use of this model for deriving detailed soil information. Section 6 provides a summary of this paper.

2. Spatial and attribute incompatibilities

2.1. The soil map production process

In order to understand the two incompatibilities and to provide a background for the similarity model presented here, a brief overview of the soil map production process is necessary. The mapping process can be divided into two conceptual parts (or sub-processes), although in real practice the two sub-processes may very well intertwine or overlap with each other. For the convenience of this presentation, we will discuss them as if they were completely separate processes. The first sub-process in soil map production is classification. During the classification process, field observations on soils are grouped into types (classes) according to their diagnostic properties (SCS, 1975; CSSC, 1978). Each of these soil classes is then assigned typical soil property values and their ranges within that class. In other words, soil classification is a process of identifying patterns (classes) of soil property values in the soil property domain (used interchangeably with parameter domain, parameter space) (Fig. 1a). It is important to identify these patterns so that the major pedogenic processes which control the development of soils, and the relationships between these processes and their pedogenic environments can be studied and understood. However, it should also be realized that each of these classes (patterns of soil property values) is characterized by typical property values (the means or the modes of property values) and their ranges in this class. In many environmental modeling applications, only the typical soil property

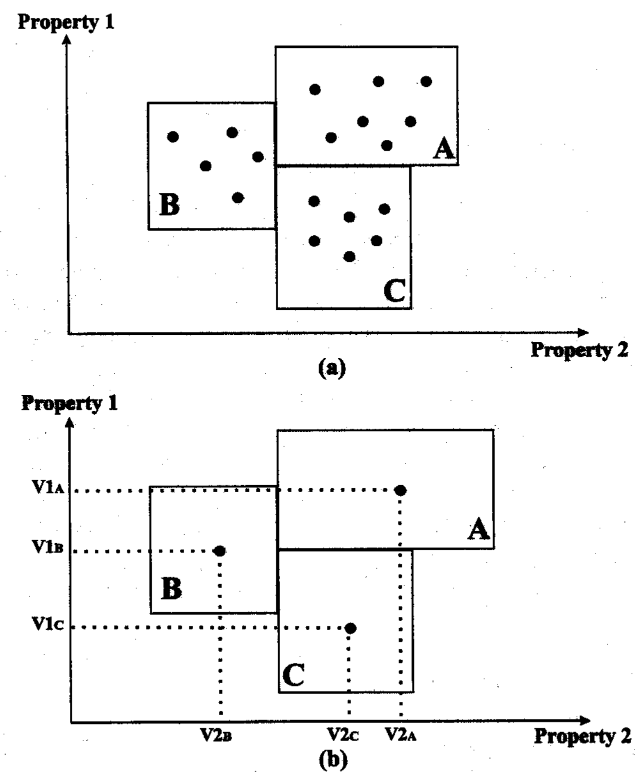


Fig. 1. Discretization of soils in the parameter domain. (a) Dots represent the locations of soils in the parameter domain, rectangles represent the boundaries of soil classes in the parameter domain. (b) Dots represent the centers of soil classes, the intervals between the projected centers on their respective axes represent the attribute resolution on these property axes.

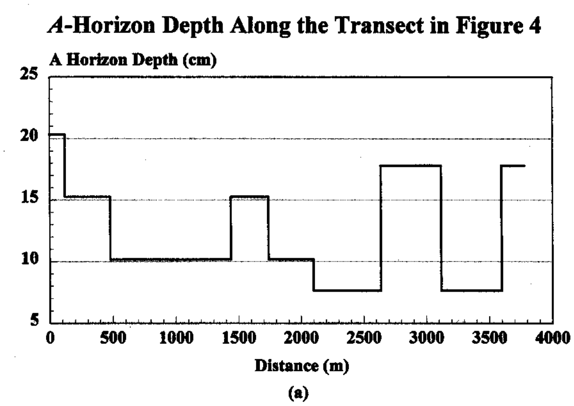
values are used. The more specific property values cannot be derived using only the range of soil property values within a soil class since the joint distribution of soil variation with the other environmental data within a soil polygon is often unknown. The typical soil property values are therefore the only means of characterizing these soil classes in the soil parameter domain. The power of describing the changes of soil property values (attribute resolution) in the parameter domain is limited to the intervals of the typical values of two adjacent soil classes (Fig. 1b). Intermediate soil property values between two typical values of two adjacent classes cannot be obtained. This reduction of soil attribute resolution is further manifested in the second sub-process of soil map production, the mapping process.

During the mapping stage, areas are delineated and assigned to mapping classes (mapping units). These units can be single-class units (made of one soil class) or mixed units (made of more than one soil class). For areas mapped as a single-class unit, all soils within the delineated polygon are considered the same as the typical soil of that soil class. For areas mapped as a mixed unit, each of the delineated polygons is said to be made of several soil classes with each occupy a certain percentage of the polygon area. The specific location of each of these soil classes within the polygon is unknown. Soil mapping can also be considered as the process of realizing the soil classification in geographic space (geographic domain). During this mapping process, two types of generalization take place: class assignment generalization, and spatial generalization.

2.2. *Incompatibilities as results of generalization*

Class assignment generalization is the process of assigning similar soils wholly to a single mapping unit. What is being said here is that all of these similar soils are to have the properties of the prescribed mapping unit to which these soils are assigned. It is because conventional soil maps are conducted under crisp logic. Under crisp logic, soil at a given point can belong to one and only one soil class and that soil is to have the soil properties of the soil class to which the soil is assigned. Under this notion, the difference in soil properties between two neighboring soil objects can either be perceived as the difference of two different mapping units (when these two soil objects are assigned to two different mapping units) or be completely ignored (when the two soil objects are assigned to a single mapping unit). On the other hand, data derived from digital terrain analyses and remote sensing techniques normally retain the subtle differences in the attribute values between the neighboring objects due to the higher attribute resolutions of these data. For example, the attribute resolution of slope gradient can be smaller than 1%. The spatial manifestation of this attribute resolution incompatibility between soil spatial information from soil maps and other environmental data derived from digital terrain analyses and remote sensing techniques is shown in Fig. 2, which depicts the changes of environmental conditions along a transect in Lubrecht, Montana (Figs. 3 and 4). Due to the higher attribute resolutions of digital terrain data and data from remote sensing techniques, the detailed gradation of environmental properties over space can be preserved (Fig. 2b,d). However, the change of soil A-horizon depth is perceived as a step function in the conventional soil map (Fig. 2a). This incompatibility can have serious implications on interpretation of results from large-scale environmental modeling applications.

Spatial generalization is related to the map scale and the cartographic techniques employed for producing soil maps. At a certain scale, only soil objects larger than a certain size (scale-dependent, called minimum mapping size) can be represented on soil maps. Soil objects smaller than the minimum mapping size are either omitted completely or merged into the surrounding soil objects (Fig. 5). Soil units **C** and **D** in (a) of Fig. 5 are too small to be represented in (b) of Fig. 5. Soil unit **A** in (b) of Fig. 5 is a mixed unit consisting of soil units **A** and **D**. This inclusion may be noted in the mapping legend as a percentage of inclusion but the spatial location of these included units are often completely lost on small-scale maps. Therefore, the spatial resolution of a soil map is the minimum mapping size, which can be a few hectares on large-scale maps to hundreds of hectares or more on small-scale maps. On the other hand, most data generated from digital terrain analyses and remote sensing techniques have the spatial resolution of 30 m by 30 m or higher and are capable of describing small (spacially) but



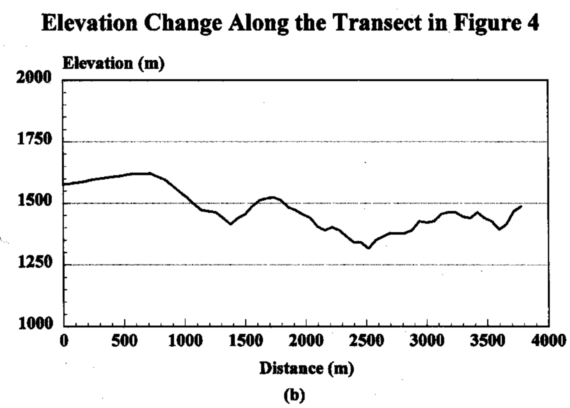
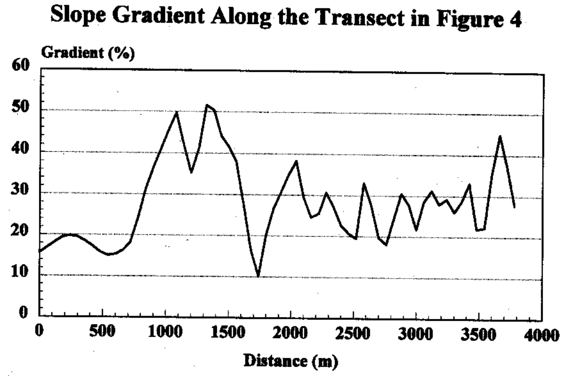
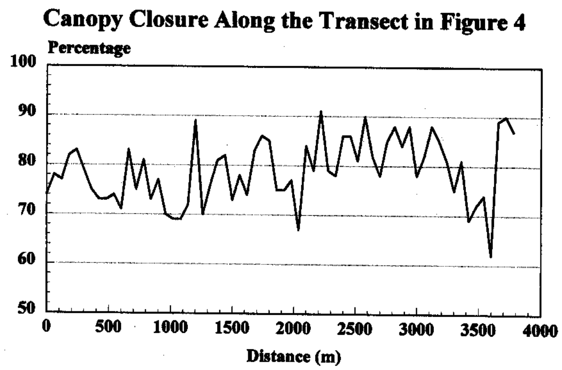


Fig. 2. Data incompatibility between soil A-horizon depth (a) and other environmental data: (b) elevation; (c) slope gradient; (d) canopy closure (remotely sensed data).



(c)



(d)

Fig. 2 (continued).

important environmental niches. Conventional soil maps are often not capable of providing soil spatial information about these small but important environmental niches. This spatial resolution incompatibility between soil spatial information from conventional soil maps and other environmental data limits the interpretation of results from a variety of large-scale environmental models.

The two incompatibilities between the soil spatial information from soil maps and other environmental data are due to the crisp logic under which soil maps are produced and due to the limited representation capacity of cartographic techniques at certain scales. In reality, soil often varies gradually and the boundaries between different types of soils are often diffuse rather than sharp (Mark and Csillag, 1990; McBratney, 1992). It may be true that soil experts know the existence of the gradual gradation of soil properties over space and the inclusion of different soil objects in soil mapping units but these cannot be mapped on soil maps due to the crisp logic employed and limitations of map scale and cartographic techniques. Therefore, the knowledge of soil scientists about soil variation cannot be fully expressed by soil maps constructed under crisp logic with the conventional cartographic techniques.

3. Recent efforts at representing soil landscape as a continuum

With the understanding of the limitations of conventional soil map making techniques for representing detail soil spatial variations, many researchers started exploring other means of quantifying and representing soil spatial variations. These efforts have initiated the development of two major types of approaches for deriving and representing soil spatial information: the statistical–geostatistical approaches and the fuzzy logic based approaches.

3.1. Statistical–geostatistical efforts

The statistical approaches first extract statistical relationships between soil properties and other landscape factors from point samples and then use the relationships together

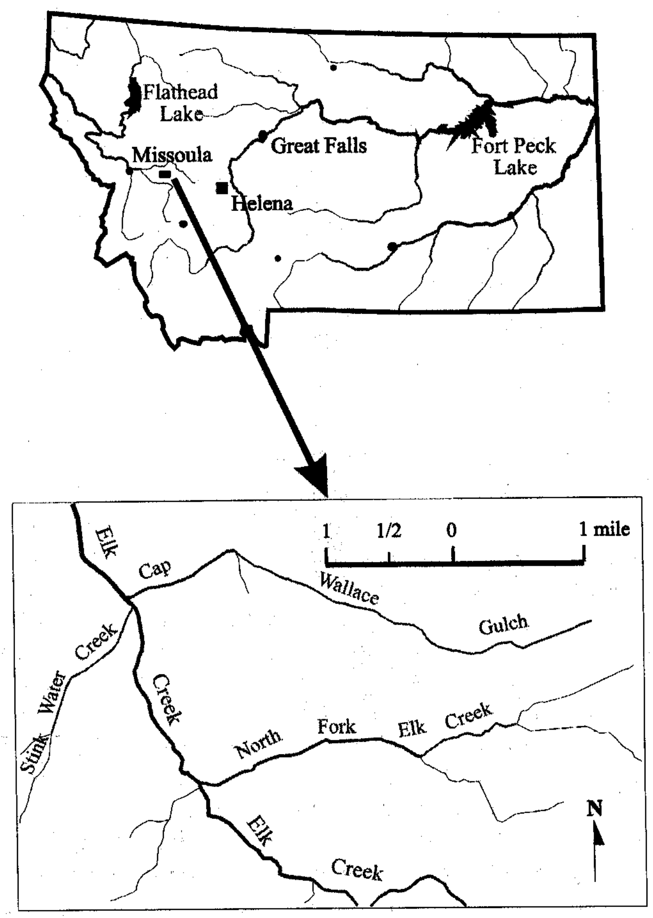


Fig. 3. Location of the Lubrecht study area, Montana.

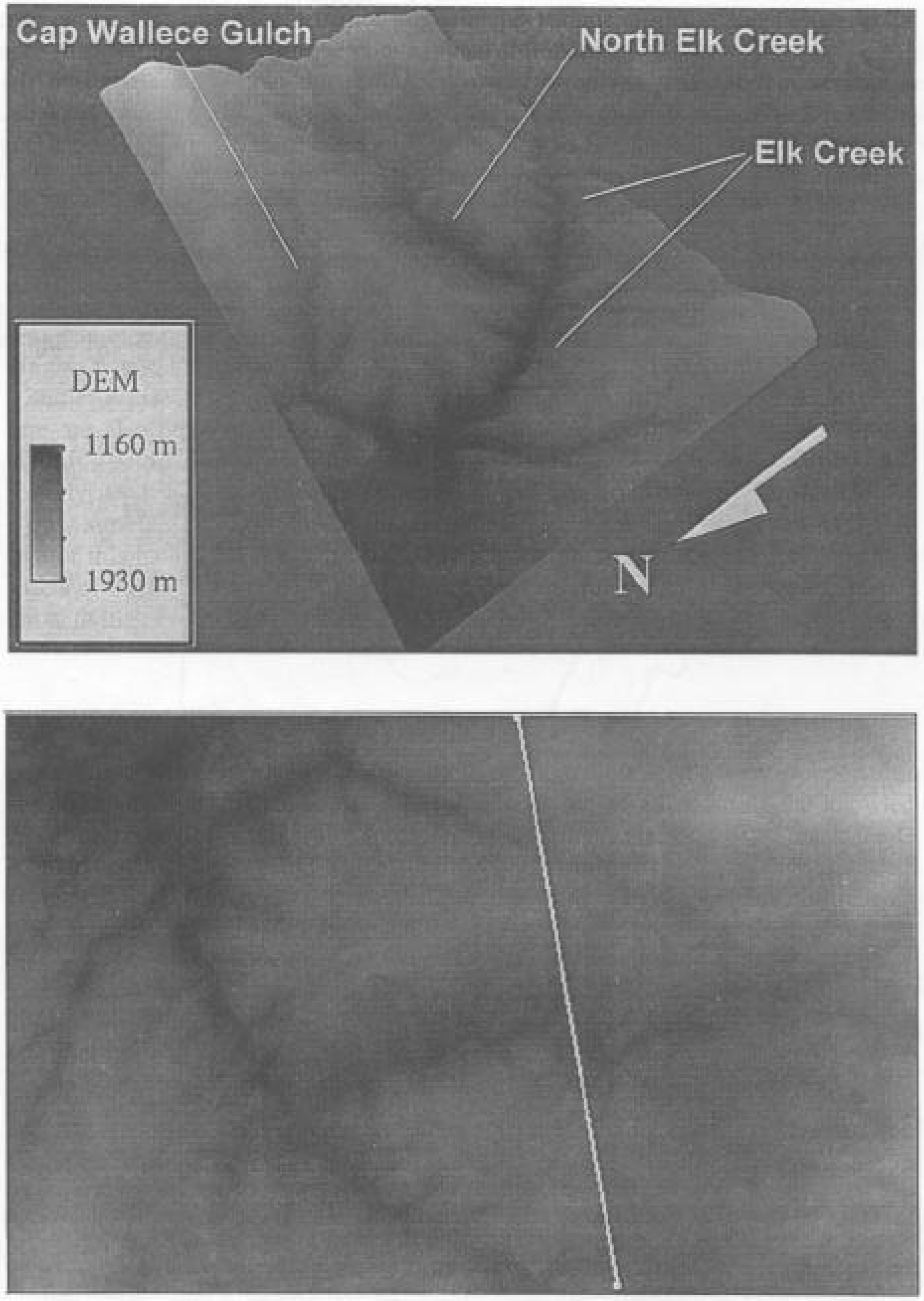


Fig. 4. The topography of Lubrecht (top) and the location of the transect (superimposed on a DEM) for revealing data incompatibility between soil A-horizon depth and other environmental data shown in Fig. 2.

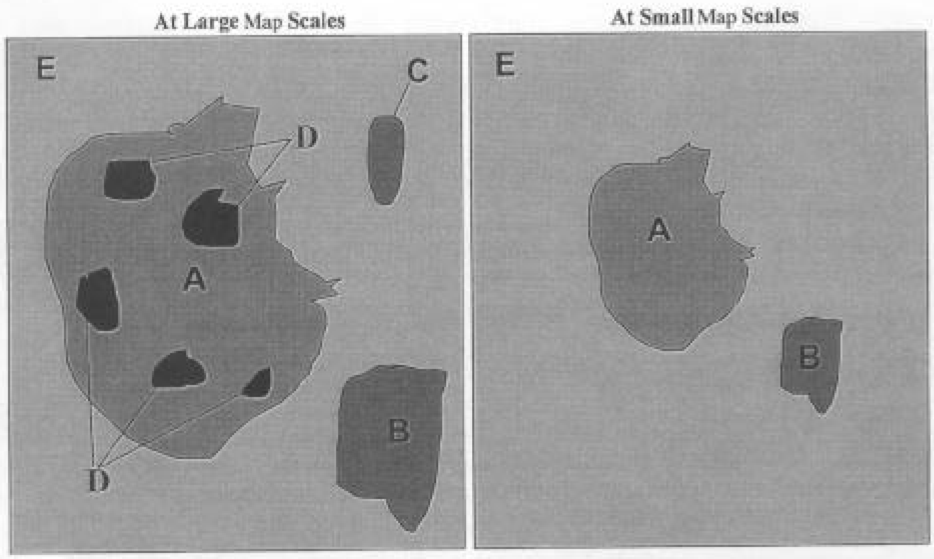


Fig. 5. Representation of soil bodies at different scales. Soil unit **D** and **C** on large-scale soil maps will disappear or be merged into soil unit **A** on small-scale soil maps.

with the landscape data in a GIS to predict the soil properties over an area (Moore et al., 1993; Gessler et al., 1995). These techniques assume that the relationship between soil properties and other landscape variables are static over space and they also require a great deal of field data to extract the relationships. Because of its assumption and the data requirement, the usefulness of the techniques has been limited for areas where the relationships between the soils and other landscape variables are very complex and where there is little field data available for deriving the relationships.

The geostatistical approaches explore the use of spatial autocorrelation of soil properties in field-sampled data sets for interpolating the soil properties at the unknown sites (for example, Burgess and Webster, 1980; McBratney and Webster, 1986; Webster and McBratney, 1989; Webster and Oliver, 1989; Webster, 1991; Loague, 1992; Bierkens and Burrough, 1993; Zhang et al., 1995). These quantitative interpolation techniques are based on the stationarity assumptions of geostatistics and also require a large amount of field data to define the spatial autocorrelation. These techniques may have limited usage for complex terrain where pedogenesis arises in a complex manner and the stationarity assumptions of geostatistics may not be met.

3.2. Fuzzy logic based approaches

The fuzzy approaches employ fuzzy logic in the classification process (for example, Burrough, 1989; Burrough et al., 1992; McBratney and De Gruijter, 1992; Odeh et al., 1992a; De Gruijter et al., 1997; Lagacherie et al., 1997). Under fuzzy logic, soils can be assigned to more than one soil class with varying degrees (fuzzy memberships) of class assignment so that the soil gradation in the parameter space can be described using these varying fuzzy membership values. Some current attempts using fuzzy classification of soils employ the unsupervised classification strategy, which means that each of the fuzzy classes generated from a fuzzy classification does not necessarily relate to any existing taxonomic class. It may be the intention of these efforts to form natural clusters of soils and to examine the objectiveness of the existing soil classes. However, the good deal of knowledge pertaining to the existing taxonomic classes are not utilized in these fuzzy classification exercises.

Odeh et al. (1992b) employed kriging techniques to create isarithmic maps of membership values of some fuzzy classes which were derived from a fuzzy-c-means (FCM) classification procedure. This might be one of the earliest attempts to represent soil landscape as a continuum in geographic space in terms of a set of isarithmic maps. However, the usefulness of these fuzzy membership maps (whether in the form of isarithmic maps or raster layers) is yet to be explored in terms of providing more accurate soil spatial information for a variety of management and modeling activities.

This paper presents a soil similarity model which uses the conventional soil taxonomic classes as the centroids of fuzzy classes. The model employs the fuzzy set theory for the assignment of local soils to these classes and uses a raster representation scheme for the representation of these fuzzy membership values over space. The soil spatial information over an area is then represented as a set of maps with each representing the spatial distribution of membership values to a particular soil class.

4. A soil similarity model for soil spatial information

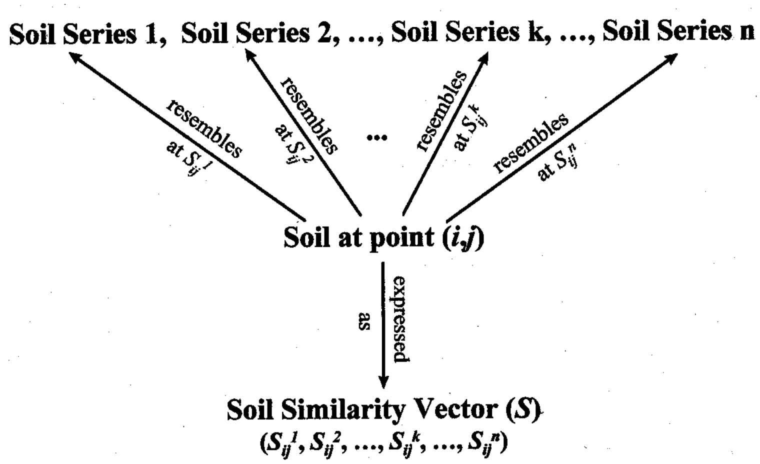
4.1. Soil similarity representation

The limitations of soil spatial information from soil maps for environmental modeling applications are related to the class assignment generalization and the spatial generalization outlined above. The class assignment generalization can be reduced or eliminated if the assignment of soil to soil classes is done under fuzzy logic. Under fuzzy assignment, a soil object can be labeled as more than one soil type (class, or taxonomic unit) with different degrees of assignment depending on the similarities between the soil and a set of prescribed soil classes. The more similar the soil to a prescribed soil class, the stronger is the assignment. Therefore, the degree of assignment is called the similarity value between the soil object and the prescribed soil class since it measures how close the soil object is to the centroid of the prescribed soil class in the parameter space. Similarity value of 1 means that the soil object is exactly located at the center of the prescribed class (typical instance) while similarity value of 0 means that the soil object does not belong to the prescribed soil class at all. Given that there are n prescribed soil classes, a soil at a given location (i,j), under fuzzy assignment, will have n similarity values with each similarity value corresponding to one of the n prescribed soil classes. The collection of these n similarity values forms an n-element vector for the soil at that location (Fig. 6). This vector is called Soil Similarity Vector at location (i,j), $S_{ij} = (S_{ij}^1, S_{ij}^2, \ldots, S_{ij}^k, \ldots, S_{ij}^n)$, where S_{ij}^k is the similarity value of the soil at point (i,j) to the prescribed soil class k, and n is the number of prescribed soil classes (taxonomic units).

It must be pointed out that S_{ij}^{k} is not a probability of whether a certain soil class occurs at a location or not. It is an index which measures the similarity between the local soil at (i,j) to soil class k.

It may have occurred that the concept of soil similarity representation used in the similarity model is similar to the fuzzy classification used by other authors (e.g., Burrough et al., 1992; McBratney and De Gruijter, 1992; Odeh et al., 1992a). There exists a major difference between the soil similarity representation and the fuzzy classification. With the former, the centers of fuzzy classes used are predefined and set to be the central concepts of conventional soil classes. In other fuzzy classifications, the centers of fuzzy classes are determined by a fuzzy clustering algorithm, often in the form of fuzzy-c-means algorithm. The reason for using existing soil classes as the central concepts of fuzzy classes is that the existing soil classes are based on certain classification schemes and are well studied by generations of soil scientists. There is a good deal of knowledge on these soil classes, particularly at the soil series level which has been used for soil mapping for many decades in the US. The knowledge about soil classes includes the relationships between soil classes and their respective environments, the characteristics of these soils and the information about management practice on these soils. By using these well-defined soil classes as fuzzy classes, we can utilize this knowledge during the interpretation and application of the results based on fuzzy logic.

It is worth pointing out that with the use of existing soil classes, the sum of all similarity values for a local soil to a set of prescribed soil classes (the sum of the elements in the soil similarity vector) does not need to be unity. This is because soil classes can be very similar to each other and it is possible for a local soil to have high similarity values to many similar soil classes and the sum can then be over unity. It is also possible that a local soil may be very unique and it may not bear much similarity to any of the prescribed soil classes, and the sum of the membership values in the vector can therefore be less than unity.



S_{ij}^{k}: Soil similarity value between the soil at point (i,j) and soil series k

Fig. 6. Fuzzy representation of soil information.

The fuzzy representation of soil information using similarity vectors is very much different from the conventional crisp representation. The soil information at a location is represented as a vector of similarity values with each of these values capable of ranging from 0 through 1. Soil information is no longer represented by the information of just one single class as it was done under crisp logic. Under this fuzzy representation, the class assignment generalization is minimized and the soil at a given point can be represented as what it is, not approximated by a typical instance of a certain prescribed soil class. In other words, S_{ij} has a greater flexibility of representing the deviation of soil at a point from a set of prescribed soil classes.

One use of this representation of soil information is the derivation of intermediate soil attribute values. On a conventional soil map, the value of a given soil property at a given location can only be the value prescribed to the soil class as which the location is mapped even though the soil property value at the point is very different from the prescribed value. Since S_{ij} has a greater flexibility of representing the deviation of soil at a point from a set of prescribed classes and it is also a vector of similarity measures to the set of prescribed soil classes, it is possible that S_{ij} can be used to derive a soil property value intermediate to the typical values of the prescribed soil classes (Zhu et al., 1997). In other words, S_{ij} can be used to provide users with a finer attribute resolution than that provided in conventional soil maps. The subtle difference between two neighboring soil objects can now be accommodated by the subtle difference between the two soil similarity vectors and the spatial gradation of soil information can be preserved under this similarity representation of soil.

One other use of the similarity representation of soil information is the generation of uncertainty information for assigning a particular soil class to the soil at a given location. When it comes to the point that we must give a class label to the soil at a location, we can also provide information about the uncertainty associated with this assignment so that management decisions can be made not just based on the soil class label but also the uncertainty involved. For example, the soils at two points (say Point 1 and Point 2) are represented by the following soil similarity vectors: (0.23, 0.25, 0.27, 0.25) and (0.1, 0.05, 0.7, 0.15), with elements representing the membership values for soil class **A**, **B**, **C**, and **D**, respectively. These vectors tell us that the soil at Point 2 is highly similar to soil class **C** but the soil at Point 1 is not so much different from the other soil classes. If we are to assign a soil class label to each of the soils, we would label the soils at both points as soil class **C** based on the highest membership value in each of the vectors. We know that this labeling is associated with different levels of uncertainty for each of the two points. Instead of ignoring this difference in uncertainty, we can measure (Goodchild et al., 1994) and report the level of uncertainty along with the class label. Under this notion, when we present a soil map to resource managers, we also present an uncertainty map associated with that soil map so that they are better informed in their decision-making processes.

4.2. The raster representation scheme

In order to reduce the impact of spatial generalization, a raster data model is used to represent the spatial distribution of soil similarity vectors since the raster model is more suitable for representing continuous spatial variation of geographic features and phenomena. The spatial resolution of a raster data model is limited only by the spatial resolution of the original input data, not by the minimum mapping size imposed by the mapping techniques. Therefore, each pixel (at the spatial resolution of the input environmental data) will have its own soil similarity vector and the soil information of small yet important environmental niches can then be provided under the similarity model. The spatial resolution incompatibility between other environmental data and soil information represented under this model can then be minimized.

If a soil similarity vector contains n elements, the collection of soil similarity vectors for all pixels in a raster database of an area forms an n-element image vector, $\boldsymbol{S}$. The kth element, S^k, in this image vector is an image representing the similarity distribution of prescribed soil class k over the area. This spatial distribution of similarity for soil class k is called fuzzy membership distribution (fuzzy membership map) of soil class k over the area. It should be emphasized that this fuzzy membership map is a representation of similarity of the soils in the area to the prescribed soil class k, and is not a probability distribution of occurrence of the soil class k. This membership map is different from the conventional representation of a soil class using soil polygons since it also shows the varying degrees of belonging of the soils in the area to the prescribed soil class k (Fig. 7).

It must be made clear that the similarity model only provides a possibility for accurately representing soil spatial information. Whether the information represented under the model is accurate depends on how this information is derived. The strategy and methods for populating the similarity model have been discussed in other papers (Zhu, 1994; Zhu and Band, 1994; Zhu et al., 1996) and are beyond the scope of this discussion. The next section presents an example to demonstrate how this model can be used in soil spatial information representation and to show the representation power of this new model.

5. An example of using the similarity model

5.1. Study area

A case study was conducted using the similarity model to represent the soil spatial information in the southeast part of the Lubrecht Experimental Forest located about 50 km northeast of Missoula, Montana, USA (Fig. 3). The study area is centered around North Fork of Elk Creek with a north–south dimension about 5 km and an east–west dimension of 7.5 km. The elevation in the study area ranges from 1130 m to 1950 m with high elevations in the east and southwest and low elevations in the northwest (Fig. 4). The study area is considered as a semi-humid and semi-arid region (Nimlos, 1986).

Most of the mountain slopes in the study area are forested, dominated by Douglas-fir (*Pseudotsuga mensiesii*) although lesser amounts of western larch (*Larix occidentalis*) and Ponderosa pine (*Pinus ponderosa*) are present. Much of the timber is second growth. There have been no large wild fires in the study area since 1937 (Nimlos, 1986). Ponderosa pine forests occupy the low elevations, particularly the south-facing slopes

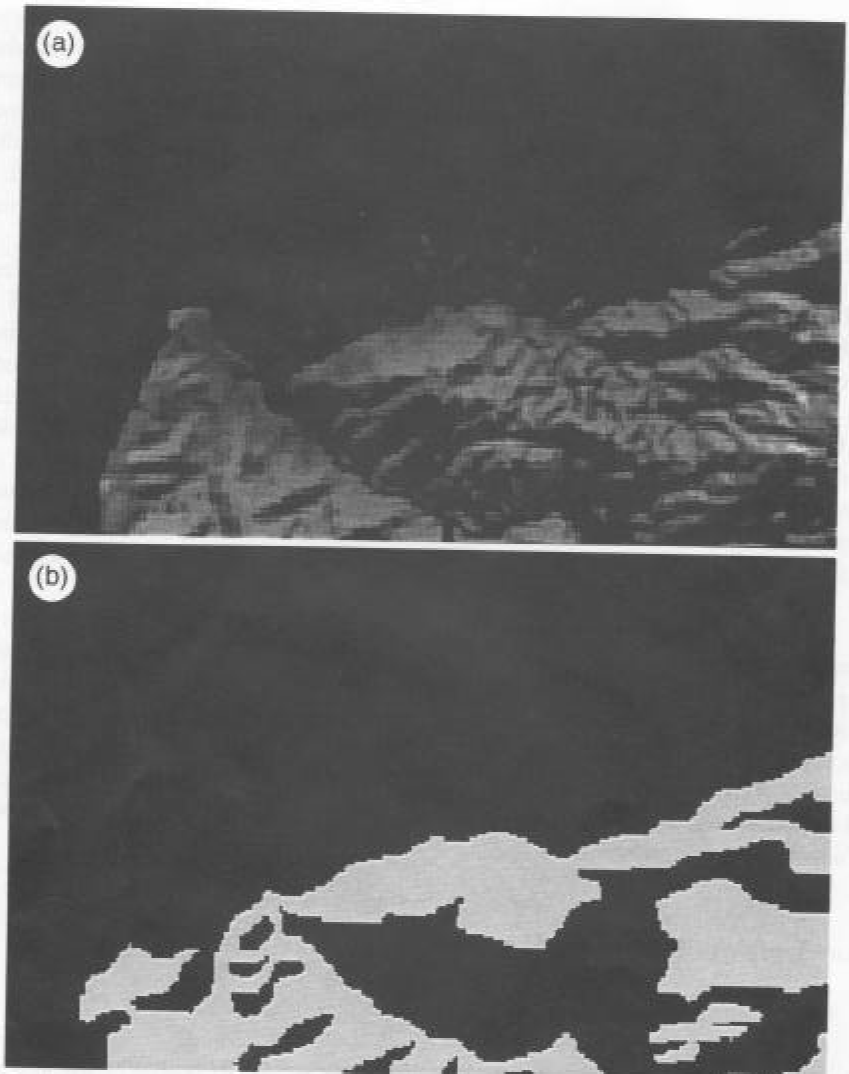


Fig. 7. Distribution of Soil Complex Elkner–Ovando: (a) under the fuzzy representation; (b) under the crisp representation employed in conventional soil maps. Figure from Zhu and Band (1994), reproduced by permission, Canadian Journal of Remote Sensing.

due to high temperatures and lower-moisture conditions on these slopes. As elevation increases, Ponderosa pine forests give way to Douglas-fir forests with a lesser amount of western larch and lodgepole pine (*Pinus contorta*). At high elevations (over 1700 m), subalpine fir (*Abies lasiocarpa*) and Engelmann-spruce (*Picea engelmannii*) replace Douglas-fir and become the dominant species.

There are four major types of parent materials in the area: Belt rocks, granite, limestone, and alluvium (Brenner, 1968). The alluvium materials occur only in limited

Table 1
Soil series in the study area

Soil series	Parent material	Soil order	Soil subgroup
Ambrant	Granite	Inceptisol	Udic Ustochrepts
Elkner	Granite	Inceptisol	Typic Cryochrepts
Evaro	Belt	Inceptisol	Typic Cryochrepts
Ovando	Granite	Entisol	Typic Cryorthents
Repp	Limestone	Inceptisol	Typic Cryochrepts
Rochester	Granite	Entisol	Typic Cryorthents
Sharrott	Belt	Inceptisol	Lithic Ustochrepts
Tevis	Belt	Inceptisol	Dystric Eutrochrepts
Trapps	Limestone	Alfisol	Typic Eutroboralfs
Whitore	Limestone	Inceptisol	Typic Cryochrepts
Winkler	Belt	Inceptisol	Udic Ustochrepts
Winkler Cool	Belt	Inceptisol	Udic Ustochrepts

areas along the North Fork and South Fork of Elk Creek. The other three parent materials make up the majority of the area with Belt rocks in the north, granite in the south, and limestone through the center part of the area. Soils on these three materials are formed from a mantle of colluvium. Belt rocks are the oldest rocks in the region and were formed from sediments deposited during the Precambrian period in a shallow sea, subsequently buried and then metamorphosed into quartzites, argillites and siltites (Nimlos, 1986). Soils formed from these materials in the study area are similar, so Belt is considered as one group of parent materials here.

Three soil orders were found to be in the study area: Alfisol, Entisol and Inceptisol (Nimlos, 1986). Alfisols are soils with leached, gray surface horizons and subsurface horizons with accumulations of illuvial clay. Entisols are weakly developed soils with very little organic-matter accumulation and no illuvial clay or sesquioxides and they are usually found on ridge crests in the study area. Inceptisols are young soils with little or no illuviated clays but brown subsoil horizons that indicate some translocations of sesquioxides. About 90% of the soils (in terms of areal extent) in the study area are Inceptisols.

The soil taxonomic unit (soil classes) used in this study is soil series, listed in Table 1. The reason for using soil series as the basic taxonomic unit for the soil similarity vectors is that soil series is the taxonomic unit which has been extensively used in soil surveys. There has accumulated a good deal of knowledge on the relationships between these soil series and their environments. This knowledge together with the environmental conditions contained in a GIS database can be used to infer the spatial distribution of these soil series and to populate the similarity model (Zhu and Band, 1994; Zhu et al., 1996).

5.2. Deriving soil similarity vectors over the study area

Zhu et al. (1996) and Zhu and Band (1994) developed a strategy for deriving soil similarity vectors based on the classic concept of Jenny (1980, 1994) that there exist

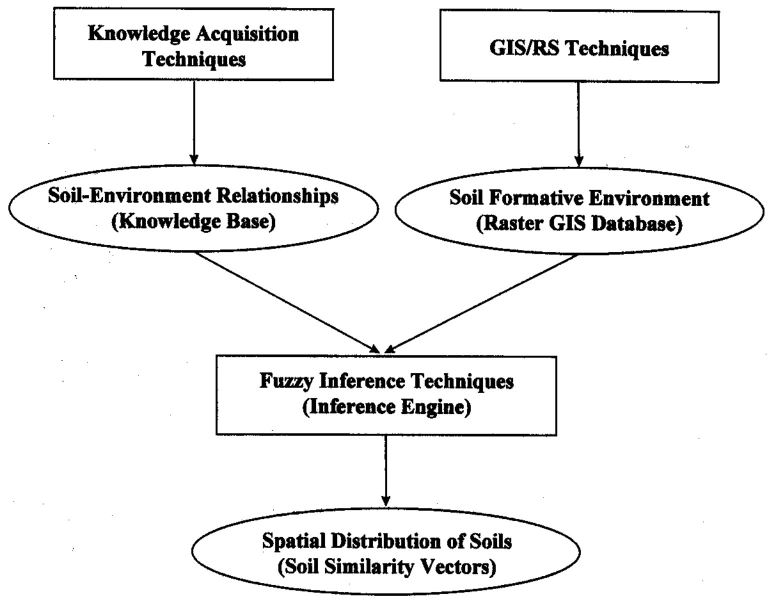


Fig. 8. The general process of deriving soil similarity vectors.

relationships between soils and their formative environments. The details of this strategy and the techniques are beyond the scope of this discussion. However, a brief overview of the soil similarity derivation for this study is given in the following paragraphs. In general, they used geographic information processing techniques to characterize the soil formative environments and developed a set of knowledge acquisition techniques to capture the knowledge on the relationships between soils and their formative environments from a soil scientist (soil expert). A set of fuzzy inference techniques (collectively called the fuzzy inference engine) were used to integrate the soil formative environments in a geographic information system (GIS) with the soil–environment relationships to derive soil similarity vectors over an area (Fig. 8).

In this illustration, six environmental variables were used to characterize the soil formative environments. These are: elevation, slope gradient, slope aspect, profile curvature, canopy coverage, and the soil parent material. Data on the first four environmental variables were derived from a USGS 30 m by 30 m digital elevation model (DEM). The canopy coverage was approximated with an index derived from the Landsat Thematic Mapper imagery (Nemani et al., 1993). The soil parent material was derived from a geological map of the study area (Brenner, 1968).

The knowledge on the relationships between the soil series and these six environmental variables was obtained from a soil expert for the study area through the use of an iterative and structured interview process (Zhu, 1995). For each soil series, the knowledge consists of six functions with each describing the relationship between the given soil series and a specific environmental variable. The relationship between an environmental variable and a given soil series was defined as the degree to which the conditions of this environmental variable favor the development of the given soil series in the area (Zhu et al., 1996). In this context, the function describing a relationship between an environmental variable and a soil series gives the various degrees of favorableness for the development of this soil series along the gradient of this environmental variable. The similarity value of the soil at a location to a given soil series can be thought of as the degree to which the local environmental conditions favor the development of the soil series.

Once the environmental data on the variables were prepared and the knowledge on the relationships was captured, the fuzzy inference engine was used to compute each of the similarity values for the soil at a location to each of the prescribed soil series (Zhu et al., 1996). In general, for each location in a raster database of the study area, the fuzzy inference engine takes the set of environmental conditions for the location and combines them with the relationship functions for a soil series to compute the similarity value of the local soil to the soil series. This process is repeated for the second soil series, and so on. When all soil series are exhausted, the soil similarity vector for the location is derived. The inference process continues onto the next location in the database and so on. When all locations are visited by the inference engine, soil similarity vectors over the entire area are derived.

Table 2 shows the soil similarity vectors for a few field sites in the study area. Although each element (the fuzzy membership value) in the vectors can have a value between 0 and 100 (unity), the sum of these values within each vector can be more or less than 100 due to the reason given in Section 4. It should be noticed that soils at the sites on Belt materials (lub03_02, lub04_01, T1_16, T1_18, and T2_08) have zero degree of similarity to soil series developed on granite and limestone materials (Ambrant, Rochester, Elkner, Ovando, Repp, Trapps, and Whitore). This means that soils on Belt materials do not belong to the soil series designated for the other two parent materials at all. Within Belt materials, soils bear different similarity values to the soil series on the Belt materials (Evaro, Tevis, Winkler Cool, Winkler, and Sharrott). It is worth pointing out that a soil at a given point is represented by the entire vector and every value in the vector is important since they together define the uniqueness of the soil at a given point. For example, the soil at Site 91_03 is similar to Ambrant, Rochester, Elkner, and Ovando at varying degrees with the strongest similarity to Elkner. Although the soil at Site lub06_02 also has the highest similarity to soil series Elkner, the distribution (combination) of similarity values in the vector is different from that of Site 91_03. Under conventional mapping techniques, soils at both sites will be assigned to soil series Elkner and will have the properties of soil series Elkner since they both have the highest similarity values to soil series Elkner. Under the similarity representation model, the entire similarity vector is retained for each individual soil. In other words, the combination of the similarity values in the vector is important, not just the highest value. It is this combination of similarity values which provides users with information about the gradation of soil in both parameter space and geographic space.

5.3. *The use of the similarity vectors*

To illustrate the advantages of this soil similarity representation, two specific uses of the soil spatial information represented under the similarity model are given and

Table 2
Soil similarity vectors for a few field sample sites

Site ID	Ambrant	Rochester	Elkner	Ovando	Repp	Trapps	Whitore	Evaro	Tevis	Winkler Cool	Winkler	Sharrott
lub03_02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*	*1.01*	*16.79*	*1.22*	*0.00*
lub04_01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*45.07*	*79.18*	*34.01*	*7.98*	*1.88*
T1_16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*	*17.78*	*25.06*	*4.57*	*0.46*
T1_18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*	*10.84*	*32.67*	*11.63*	*1.33*
T2_08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*80.89*	*38.02*	*29.09*	*4.01*	*8.80*
lub04_02	0.00	0.00	0.00	0.00	*0.00*	*16.29*	*51.16*	0.00	0.00	0.00	0.00	0.00
lub04_03	0.00	0.00	0.00	0.00	*2.01*	*18.79*	*55.50*	0.00	0.00	0.00	0.00	0.00
lub07_01	0.00	0.00	0.00	0.00	*13.96*	*65.03*	*48.91*	0.00	0.00	0.00	0.00	0.00
lub10_02	0.00	0.00	0.00	0.00	*55.02*	*30.88*	*0.00*	0.00	0.00	0.00	0.00	0.00
T1_15	0.00	0.00	0.00	0.00	*11.70*	*34.96*	*0.00*	0.00	0.00	0.00	0.00	0.00
lub06_02	*9.41*	*0.00*	*53.79*	*52.80*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
lub06_03	*29.37*	*29.23*	*16.25*	*45.81*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
lub07_02	*26.62*	*30.36*	*0.19*	*1.85*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
lub10_03	*7.63*	*1.55*	*0.00*	*47.49*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
91_03	*16.54*	*14.47*	*69.23*	*39.79*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

discussed in the following sections. The first use is the derivation of a spatially detailed soil map from the soil similarity vectors of the area and is for the illustration of the high spatial resolution which pertains to the resulting soil map. The second use is the derivation of a spatially continuous soil property map (A-horizon depth) for the study area and is for the demonstration of spatial gradation of soil information preserved through the use of this similarity model.

5.3.1. The derivation of the detailed soil map

The soil similarity vectors can be hardened to produce a soil map. The hardening is done by assigning each location the label of the soil class which has the highest membership value in the similarity vector for that point. The newly created soil map and the conventional soil map over the Lubrecht study area are shown in Fig. 9. It can be observed from the two maps that the newly created soil map contains more spatial details than the conventional soil map of the area. The different soil series occurring along the small draws (shallow but very wide gullies, ravines or valleys) on large slopes are shown on the newly created map but not on the conventional soil map.

In order to further illustrate the higher quality of the newly created soil map, field observations were made to determine the soil series at 64 sites. Special attention was paid to sample the small draws when the 64 sites were distributed over the study area. The locations (in terms of map coordinates) of these sites were determined with a GPS (global positioning systems) receiver, and topographic maps. With these map coordinates soil series mapped at these sites on the two soil maps were also determined. The comparison between the soil series observed in the field and the two sets of soil series obtained from the two soil maps reveals that the soil series from the newly created soil map match the field observations at 81% (52 out of 64), while the soil series from the conventional soil map match the field observations at 61% (39 out of 64). The increase in quality for the newly created soil map over the conventional one is mainly attributed to the high spatial resolution underlying the similarity model which minimizes the inclusion of small soil objects into larger objects.

5.3.2. The derivation of the continuous soil property map

Zhu et al. (1997) assumed that the more similar two soils, the closer are their soil property values and derived a soil A-horizon depth image over the study area using the following equation:

$$D_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} \cdot D^{k}}{\sum_{k=1}^{n} S_{ij}^{k}} \qquad (1)$$

where D_{ij} is the soil A-horizon depth at site (i,j), D^k is the prescribed soil A-horizon depth of soil series k, and n is the total number of prescribed soil series in the area. The so derived soil A-horizon depth image and the soil A-horizon depth image derived from the conventional soil map are shown in Fig. 10.

The contrast between the two images is strong and clear. The image of depth based on the similarity vectors (Fig. 10, top) shows a spatially continuous pattern of A-horizon

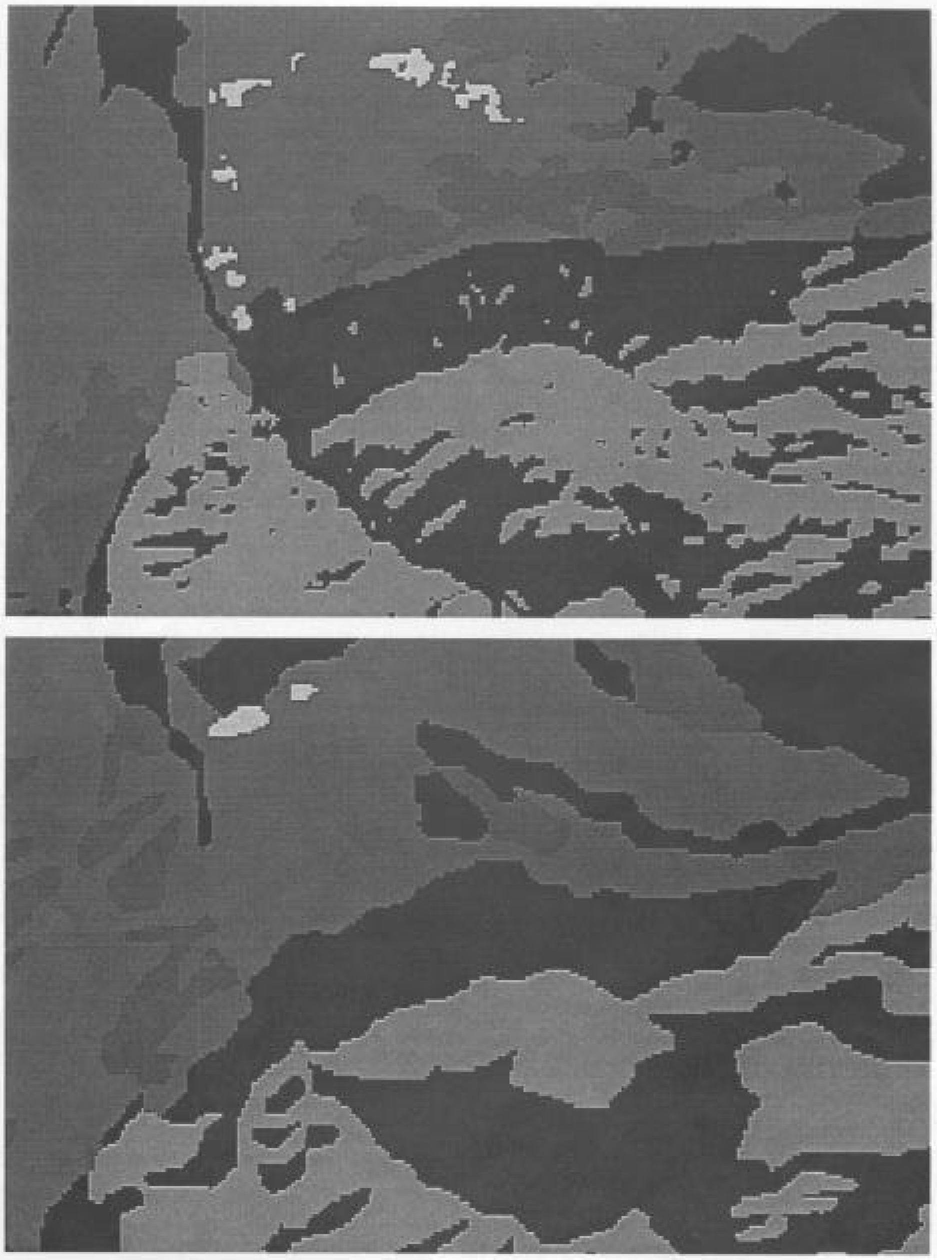


Fig. 9. The soil map derived from the similarity vectors (top) and the conventional soil map of the Lubrecht study area (bottom).

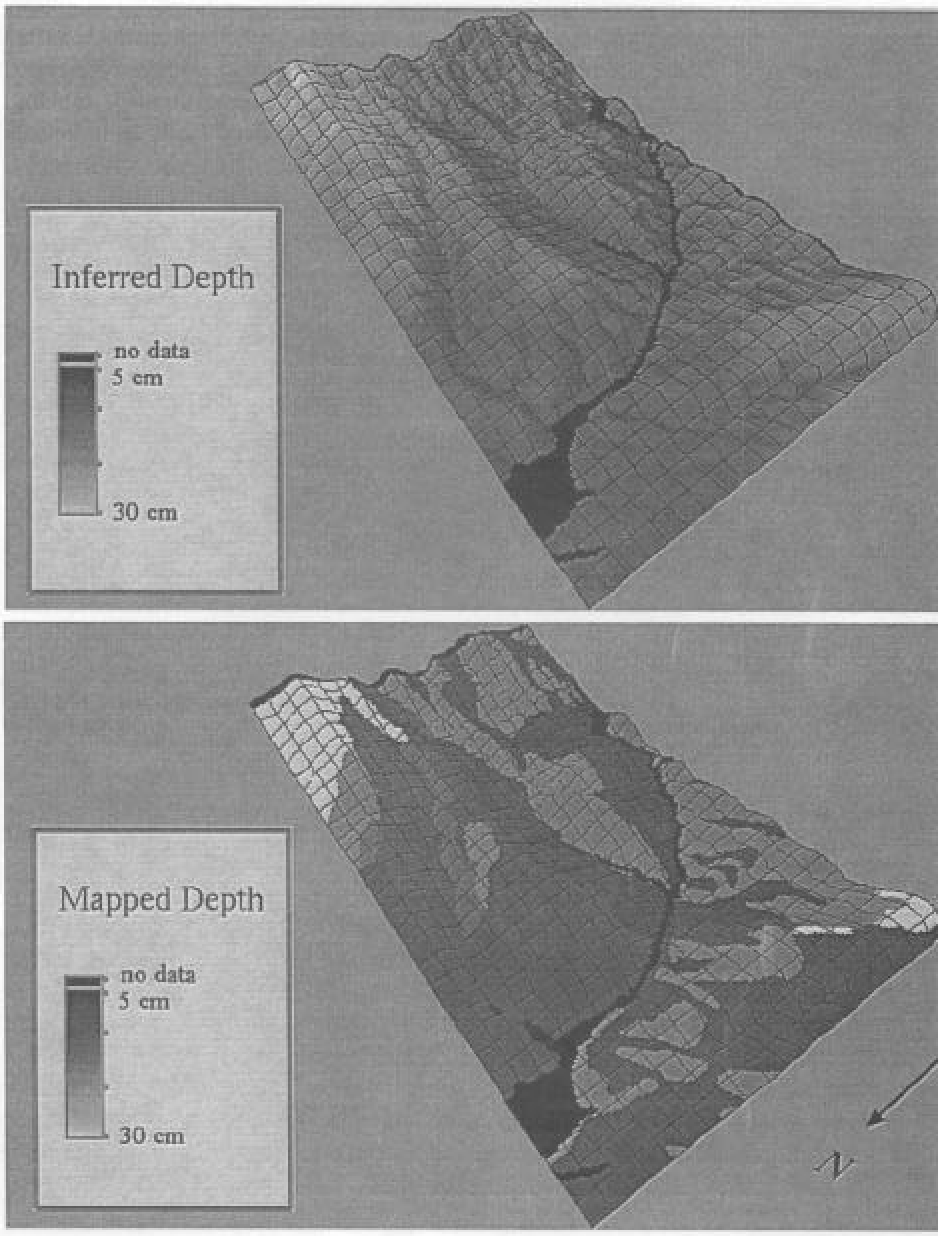


Fig. 10. The A-horizon depth image inferred from the soil similarity vectors (top) and the A-horizon depth image derived from the conventional soil map (bottom) of the Lubrecht area. Figure from Zhu and Band (1994), reproduced by permission, Canadian Journal of Remote Sensing.

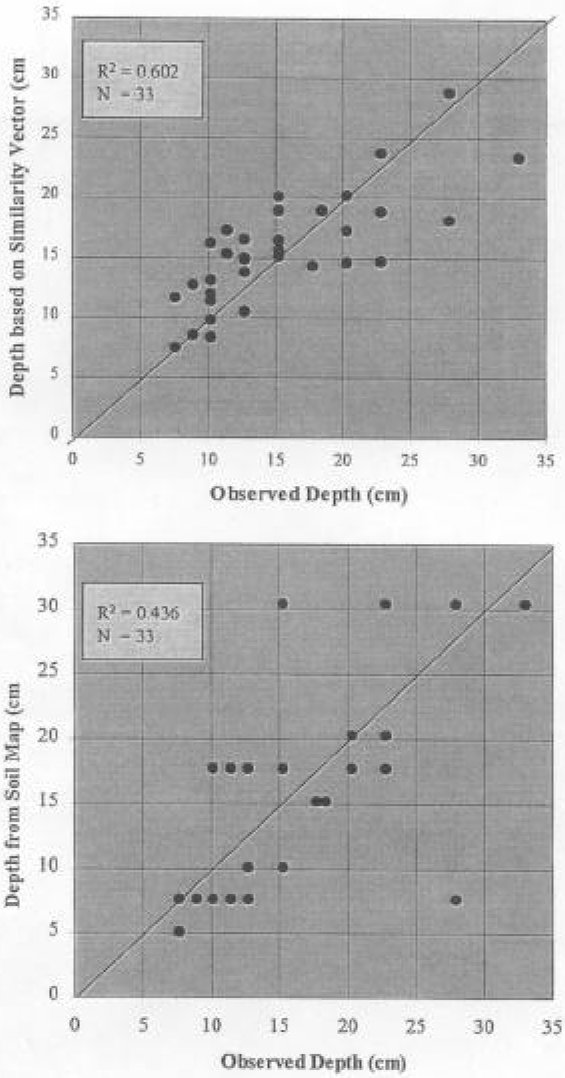


Fig. 11. Scatter plot of the depths based on the similarity vectors vs. observed depths (top) and that of the depths from the soil map vs. observed depths (bottom) for the 33 field sites.

depth over the area while the image of soil A-horizon depth derived from the conventional soil map (Fig. 10, bottom) inherits the exact spatial pattern of the conventional soil map, on which the soil landscape is discretized into distinct and discreet spatial units and the soil property variation is perceived as a step function. Since the study area is in a semi-humid to semi-arid region of western Montana, the soils on north-facing slopes and at high elevations were better developed than soils on south-facing slopes and at lower elevations due to the limited moisture conditions at the low elevations and on south-facing slopes. The A-horizons are deep for these better developed soils. Although both images show this trend, the depth image based on the similarity vectors shows a gradual change of depth over space while the depth image from the soil map shows the changes only occurring at the boundaries of neighboring soil polygons. This supports the argument that soil spatial information represented using the similarity model can preserve higher levels of detail at the attribute level (higher attribute resolution and better spatial gradation) than that in conventional soil maps.

From the comparison of two A-horizon depth images, it can also be observed that the depth image based on the similarity model shows soil A-horizon depth at a greater spatial detail. The variation of soil A-horizon depth on different facets of small draws can be very well identified in the A-horizon depth image based on the similarity model. The depth image derived from the conventional soil map shows the depth by polygons and there is no variation of soil A-horizon depth within each of these polygons. This uniform distribution of soil A-horizon depth within a mapped soil polygon cannot be realistic for this mountainous area. This comparison suggests that the soil spatial information represented under the similarity model has a much higher spatial resolution (better spatial details, fewer spatial inclusions) than that represented in conventional soil maps.

In order to verify the accuracy of the soil A-horizon depth based on the similarity model, observations of soil A-horizon depth were made at 33 field sites during the summer season of 1993. The scatter plot of the depths based on the soil similarity vectors versus the field observed depths at these sites and that of the depths from the conventional soil map versus the field depths are shown in Fig. 11. From these two plots, it can be observed that the depths based on the similarity vectors at these field sites are more closely associated with the field-observed depths than the corresponding depths obtained from the soil map. Although both correlation coefficients are highly significant (numbers in the white boxes in the two plots), the correlation between the depths from the similarity vectors and the observed depths is much stronger than that between the depths from the soil map and the observed depths. It further illustrates that the similarity representation model has less attribute generalization than the model used in conventional soil maps.

6. Discussion and conclusions

The two incompatibilities (attribute resolution and spatial resolution) between the soil information from conventional soil maps and other environmental data from digital terrain and remote sensing techniques are results of the class assignment generalization under crisp logic and spatial generalization due to the scale and cartographic techniques used in the soil-map-making process. This paper presents a similarity model to overcome the two generalizations in soil spatial information representation. The model uses the soil similarity vector based on fuzzy logic for representing a soil in its parameter domain so that it overcomes the class assignment generalization problem. In the spatial domain, a raster data model is used to provide the capability of representing soil spatial information at a very fine spatial resolution. The employment of the raster data model helps to minimize the spatial generalization which often occurs in the production of conventional soil maps. Through a case study in the Lubrecht Experimental Forest, Montana, it has been demonstrated that the similarity model has a greater capacity for representing soil spatial information in both the parameter space and the geographic space than the model used in the conventional soil map.

It must be realized that this similarity model is only a way for representing soil spatial information at the level of details compatible with other detailed environmental data. It provides soil scientists a more flexible method for depicting their understanding of soil landscape than the model underlying conventional soil maps. It should also be pointed out that although the similarity model is capable of representing soil spatial information at higher spatial and attribute resolutions, the accuracy of the soil information represented under this model entirely depends on the accuracy of the similarity vectors which in turn relies on the process of generating these vectors (Zhu and Band, 1994; Zhu et al., 1996).

In this paper, the knowledge-based approach for deriving soil similarity vectors taken by Zhu et al. (1996) was used as an example on how to derive soil similarity vectors. There are also other means (such as supervised fuzzy classification) for deriving soil similarity vectors. Research in developing methods for deriving these similarity vectors is clearly needed. The two usage examples of the similarity model given in this paper are only simple demonstrations of the usefulness of this model. Further research on the usage of this model for deriving detailed soil spatial information is also needed and the impacts of this detailed soil spatial information on environmental modeling and management activities need to be thoroughly examined.

Acknowledgements

The author wishes to thank Dr. L.E. Band for his encouragement and support in the research reported in this paper. Mr. Barry Dutton's co-operation during the knowledge acquisition processes is gratefully appreciated. The author also wishes to thank the GIS laboratory, School of Forestry, University of Montana for providing the GIS data used in this paper. The startup funds provided to the author by the Graduate School at University of Wisconsin-Madison made it possible to complete the work reported in this paper.

References

Band, L.E., Moore, I.D., 1995. Scale: Landscape attributes and geographical information systems. Hydrol. Process. 9, 401–422.

Band, L.E., Patterson, J.P., Nemani, R., Running, S.W., 1993. Ecosystem processes at the watershed scale: Incorporating hill-slope hydrology. Agric. For. Meteorol. 63, 93–126.

Bierkens, M.F., Burrough, P.A., 1993. The indicator approach to categorical soil data, II. Application to mapping and land use suitability analysis. J. Soil Sci. 44, 369–381.

Bregt, A.K., 1992. Processing of Soil Survey Data. Doctoral Thesis, Wageningen Agricultural University, Wageningen, 167 pp.

Brenner, R.L., 1968. The geology of Lubrecht Experimental Forest. Mont. For. Conserv. Exp. Stn. Lubrecht Ser. 1, 71 pp.

Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties, I. The semi-variogram and punctual kriging. J. Soil Sci. 31, 315–331.

Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. J. Soil Sci. 40, 477–492.

Burrough, P.A., MacMillan, R.A., Van Deursen, W., 1992. Fuzzy classification methods for determining land suitability from soil profile observations. J. Soil Sci. 43, 193–210.

CSSC (Canada Soil Survey Committee), 1978. The Canadian System of Soil Classification. Can. Dep. Agric. Publ., 1646, Supply and Services Canada, Ottawa, Ont., 164 pp.

De Gruijter, J.J., Walvoort, D.J.J., Van Gaans, P.F.M., 1997. Continuous soil maps: a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. In: De Gruijter, J.J., McBratney, A.B., McSweeney, K. (Editors), Fuzzy Sets in Soil Science. Geoderma 77, 169–195 (this issue).

Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Syst. 9 (4), 421–432.

Goodchild, M.F., 1989. Modeling error in objects and fields. In: Goodchild, M.F., Gopal, S. (Eds.), Accuracy of Spatial Databases. Taylor and Francis, New York, pp. 107–113.

Goodchild, M.F., 1992. Geographical data modeling. Comput. Geosci. 18, 401–408.

Goodchild, M.F., 1993. The state of GIS for environmental problem-solving. In: Goodchild, M.F., Parks, B.O., Steyaert, T. (Eds.), Environmental Modelling With GIS. Oxford Univ. Press, New York, pp. 8–15.

Goodchild, M.F., Chih-Chang, L., Leung, Y., 1994. Visualizing fuzzy maps. In: Hearnshaw, H.M., Unwin, D.J. (Editors), Visualization in Geographical Information Systems. John Wiley, New York, pp. 158–167.

Jenny, H., 1980. The Soil Resource: Origin and Behavior. Springer, New York, 377 pp.

Jenny, H., 1994. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, New York, 281 pp.

Lagacherie, P., Cazemier, D.R., Van Gaans, P. and Burrough, P.A., 1997. Fuzzy *k*-means clustering of fields in an elementary catchment and extrapolation to a larger area. In: De Gruijter, J.J., McBratney, A.B., McSweeney, K. (Eds.), Fuzzy Sets in Soil Science. Geoderma 77, 197–216 (this issue).

Loague, K., 1992. Soil water content at R-5, Part 1. Spatial and temporal variability. J. Hydrol. 139, 233–251.

Mark, D.M., Csillag, F., 1990. The nature of boundaries on 'area-class' maps. Cartographica 27 (1), 65–78.

McBratney, A.B., 1992. On variation, uncertainty and informatics in environmental soil management. Aust. J. Soil Resour. 30, 913–935.

McBratney, A.B., De Gruijter, J.J., 1992. A continuum approach to soil classification by modified fuzzy *k*-means with extragrades. J. Soil Sci. 43, 159–175.

McBratney, A.B., Webster, R., 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J. Soil Sci. 37, 617–639.

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

Nemani, R., Pierce, L., Band, L.E., Running, S., 1993. Forest ecosystem processes at the watershed scale: sensitivity to remotely sensed leaf area index estimate. Int. J. Remote Sensing 14, 2519–2534.

Nimlos, J.T., 1986. Soils of Lubrecht Experimental Forest. Mont. For. Conserv. Exp. Stn., Misc. Publ. 44, 36 pp.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992a. Soil pattern recognition with fuzzy-*c*-means: application to classification and soil–landform interrelationships. Soil Sci. Soc. Am. J. 56, 505–516.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992b. Fuzzy-*c*-means and kriging for mapping soil as a continuous system. Soil Sci. Soc. Am. J. 56, 1848–1854.

SCS (Soil Conservation Service), 1975. Soil Taxonomy. US Department of Agriculture, Agriculture Handbook No. 436, Superintendent of Documents. U.S. Government Printing Office, Washington, D.C., 754 pp.

Webster, R., 1991. Local disjunctive kriging of soil properties with change of support. J. Soil Sci. 42, 301–318.

Webster, R., McBratney, A.B., 1989. On the Akaike information criterion for choosing models for variograms of soil properties. J. Soil Sci. 40, 493–496.

Webster, R., Oliver, M.A., 1989. Optimal interpolation and isarithmic mapping of soil properties, VI. Disjunctive kriging and mapping the conditional probability. J. Soil Sci. 40, 497–512.

Zhang, R., Rahman, S., Vance, G.F., Munn, L.C., 1995. Geostatistical analyses of trace elements in soils and plants. Soil Sci. 159 (6), 383–390.

Zhu, A.X., 1994. Soil Pattern Inference using GIS under Fuzzy Logic. Ph.D. Diss., Department of Geography, University of Toronto, Toronto, 177 pp.

Zhu, A.X., 1995. A knowledge-based approach in environment management. Proc. Geoinformatics '95, Hong Kong, May 25–28, pp. 100–110.

Zhu, A.X., Band, L.E., 1994. A knowledge-based approach to data integration for soil mapping. Can. J. Remote Sensing 20 (4), 408–418.

Zhu, A.X., Band, L.E., Dutton, B., Nimlos, T.J., 1996. Automated soil inference under fuzzy logic. Ecol. Modeling 90, 123–145.

Zhu, A.X., Band, L.E., Vertessy, R., Dutton, B., 1997. Derivation of soil property using a soil–land inference model (SoLIM). Soil Sci. Soc. Am. J. 61(2), 523–533.