Reprinted from the *Soil Science Society of America Journal*
Volume 61, no. 2, March–April 1997
677 South Segoe Rd., Madison, WI 53711 USA

Derivation of Soil Properties Using a Soil Land Inference Model (SoLIM)

A-Xing Zhu,* Lawrence Band, Robert Vertessy, and Barry Dutton

ABSTRACT

SoLIM (Soil Land Inference Model) is a fuzzy inference scheme for estimating and representing the spatial distribution of soil types in a landscape. This study developed the inference method a step further to derive continuous soil property maps through two case studies. The first case illustrates the derivation of soil A horizon depth in a mountainous area in western Montana. It was found that the inferred depths are a closer fit to observed depths than those derived from the conventional soil map at both spatial and attribute levels. The second case shows the derivation of soil transmissivity values across a small catchment with a gentle environmental variation in Tumut, NSW, Australia. This case shows that the derived soil transmissivity map is comparable to the results from systematic field survey over a small area. SoLIM works well in an area where there is a good understanding of the relationships between soils and their formative environment and where the soil formative environment can be characterised using current geographical information system techniques. However, we experienced difficulty with the methodology when it was applied in an area where the environmental gradient is gentle and the soil formative environment cannot be very well described using the primitive environmental indices currently employed in SoLIM.

WE DEVELOPED A METHODOLOGY for deriving continuous soil property maps. The methodology builds on the approach given by Zhu et al. (1996), Zhu (1996), and Zhu and Band (1994), a model (Soil Land Inference Model, referred to as SoLIM) for acquiring and representing soil spatial information. SoLIM consists of three basic components: a knowledge acquisition process, a set of GIS techniques, and a fuzzy inference engine. The knowledge acquisition process is used to extract the relationships between soils and their environmental conditions from a soil expert (soil scientist). The GIS techniques are used to characterize soil formative environmental conditions. The fuzzy inference engine combines the extracted relationships with the soil environmental conditions to produce soil spatial information. The inference engine was constructed under fuzzy logic so that the resultant soil information is not represented as conventional soil maps but as fuzzy membership maps. We show how these fuzzy membership maps can be used to produce continuous soil property maps.

Many environmental modeling and land management applications require detailed soil spatial and attribute information in order to match other detailed environmental data obtained from remote sensing and digital terrain analysis (Band et al., 1991, 1993). Currently, soil maps produced from standard soil surveys are the major source of soil information for a variety of land analyses, ecological modeling, and management applications. However, standard soil surveys were not designed to provide the detailed (high-resolution) soil information required by some environmental modeling applications (Moore et al., 1993) and crop management (Petersen, 1991). The major limitations of soil information derived from conventional soil maps for environmental modeling and land management applications are: (i) low spatial resolution; and (ii) uniform attribute value within the unit delineated (low attribute resolution).

These two limitations are due to the fixed map scale and the cartographic model used in the map production processes (Zhu, 1996). Only soil bodies larger than a certain size (minimum mapping size, scale dependent) can be shown on a soil map. Soil bodies smaller than this minimum mapping size are merged or lumped into the surrounding or neighboring soil bodies. The cartographic model used in conventional soil mapping further degrades the quality of the information contained in soil maps. Under the cartographic model, the soil continuum is discretized into discrete spatial objects (Fig. 1a) and the spatial variation of soil information is implicitly considered as a step function (the solid line of Fig. 1b). In other words, all spatial variation of soil information occurs at the boundaries of delineated soil polygons, and soil properties have uniform values within each soil polygon. In reality, soil often varies gradually and the boundaries between different soils are often diffuse rather than sharp (Mark and Csillag, 1990), and a soil property within a soil polygon is often not uniform (the dashed line in Fig. 1b).

The SoLIM approach (Zhu et al., 1996; Zhu and Band, 1994) is different from conventional soil mapping in the way the soil landscape is perceived and how soil information is represented. Under the SoLIM approach, soil is considered a result of the interaction among its formative environmental factors with time (Jenny, 1941, 1980). In other words, there exists a relation between soil and the formative environment, which can be qualitatively expressed as

$$S = f_1(\mathbf{E},t)$$

where S is soil, $\mathbf{E}$ represents a vector of environmental conditions, t is time, and f_1 represents the interaction between soil and the formative environment. It is very difficult at this stage to determine the length of time of interaction among the soil-forming factors. Sometimes, the time information is embedded in the environmental conditions, such as topographic position, vegetation cover, and substrate. Therefore, the relation is simplified to be

$$S = f(\mathbf{E})$$

A.X. Zhu, Dep. of Geography, Univ. of Wisconsin, 550 North Park Street, Madison, WI 53706-1491; L.E. Band, Dep. of Geography, Univ. of Toronto, 100 St. George Street, Toronto, ON, Canada M5S 1A1; R. Vertessy, CRC for Catchment Hydrology, CSIRO Division of Water Resources, Canberra, ACT 2601, Australia; and B. Dutton, Land & Water Consulting Inc., P.O. Box 8254, Missoula, MT 59807. Received 31 July 1995. *Corresponding author (azhu@geography.wisc.edu).

Published in Soil Sci. Soc. Am. J. 61:523–533 (1997).

Abbreviations: GIS, geographical information system; DEM, digital elevation model; MAE, mean absolute error; RMSE, root mean square error; AC, agreement coefficient.

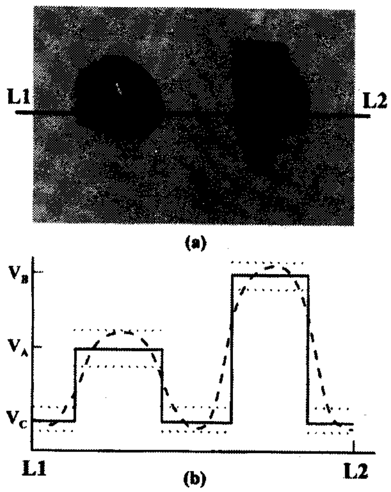


Fig. 1. (a) Soil map showing location of Transect L1–L2; (b) property values along Transect L1–L2.

The soil formation environment (**E**) of an area can be characterized using GIS techniques and local soil scientists' expertise on soil–environment relationships can be used to approximate the interaction (f) between soil and its environment in an area (Fig. 2). Soil spatial information across an area can then be inferred by combining information on the soil formative environment (**E**) with soil scientists' expertise (f).

Soil landscape is considered a continuum under SoLIM and change of soil properties across space is often gradual. Soil classes are ideal concepts of soils and soils in fields are usually intermediate to these concepts. Under this perception, the soil (S) at any point or location is assumed, in SoLIM, to be similar in varying degrees to a prescribed set of soil taxonomic units (such as soil series) or central concepts (Zhu, 1996). Soil at any point or location (i,j) can then be expressed by an n-element vector (soil similarity vector), $\mathbf{S}_{ij} = (S_{ij}^1, S_{ij}^2 \ldots S_{ij}^k \ldots S_{ij}^n)$, where S_{ij}^k is a membership value that measures the similarity between the soil at point (i,j) and prescribed

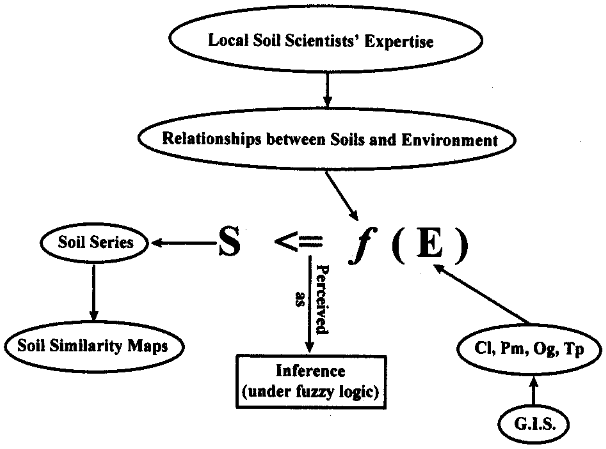


Fig. 2. The theoretical basis of SoLIM: Soil (S) is a function of its environmental factors. Cl: climate, Pm: parent materials, Og: organism, Tp: topography. Reproduced from Zhu and Band (1994) with permission from the publisher.

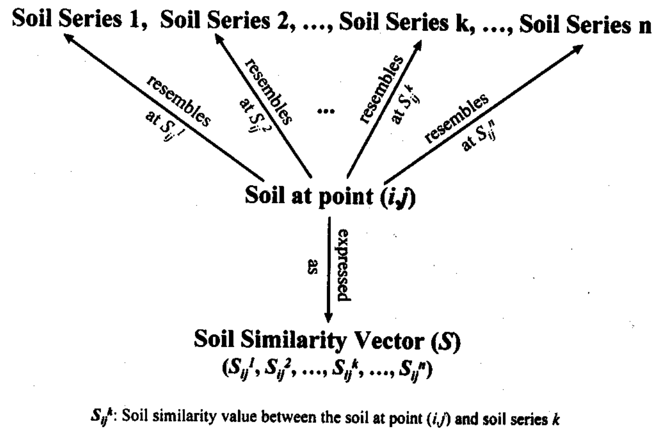


Fig. 3. Fuzzy representation of soil information using soil similarity vector.

soil taxonomic unit k (Fig. 3) and n is the number of prescribed soil taxonomic units in the area (Zhu, 1996).

This representation of soil information is different from the conventional representation. The similarity of the soil at a location to a soil taxonomic unit is expressed in terms of a membership value between 0.0 and 1.0 (that is, $0.0 \leq S_{ij}^k \leq 1.0$). It should be clarified that S_{ij}^k is not a probability but a fuzzy membership expressing the similarity of the soil at point (i,j) to the prescribed taxonomic unit k. It is this membership value that can be used to provide soil property values intermediate to the typical values of the prescribed soil units.

We use two case studies to illustrate the usefulness and limitations of SoLIM for deriving continuous soil property maps. The first case is the derivation of soil A horizon depth of the Lubrecht Experimental Forest, Montana. The second case is the derivation of soil water transmissivity (T) of the Redhill Catchment, Tumut, NSW, Australia.

METHODS AND STUDY AREAS

Deriving the Soil Similarity Vector

As discussed above, under SoLIM, soil at location (i,j) is represented by membership (similarity) vector $\mathbf{S}_{ij}$. The collection of $\mathbf{S}_{ij}$ across a region forms $\mathbf{S}$, which represents a set of fuzzy membership maps. Each of these maps represents the spatial distribution of similarity (membership) to a particular soil category. In other words, $\mathbf{S}$ consists of the following elements: $\mathbf{S}^1$, $\mathbf{S}^2$,.... $\mathbf{S}^k$, ... $\mathbf{S}^n$, where $\mathbf{S}^k$ represents the spatial distribution of membership to soil category k. For example, Fig. 4 shows the membership map for the Elkner soil series ($\mathbf{S}^{\mathrm{Elkner}}$) in the Lubrecht study area.

In order to obtain these fuzzy membership maps, the basic element of the soil similarity vector, S_{ij}^k, must be determined first. The details of deriving each S_{ij}^k, as well as the process of knowledge acquisition, are beyond the scope of this study and are extensively discussed in Zhu et al. (1996) and Zhu (1994). In general, the knowledge of the relationships between the environment and soil are stored in a knowledge base in the form of a knowledge frame (Fig. 5). Each soil category would have its own knowledge frame. Each slot in the knowledge frame contains an optimality function (curve), which quantitatively describes the relationship between a given soil and a particular environmental factor. This optimality function

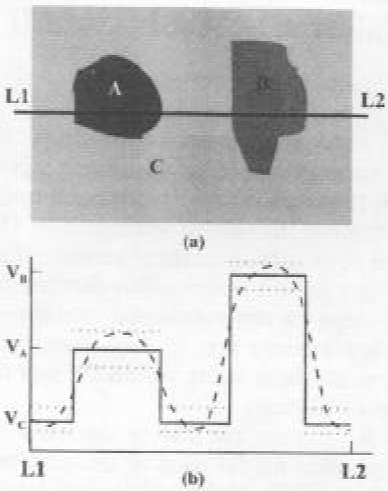


Fig. 1. (a) Soil map showing location of Transect L1–L2; (b) property values along Transect L1–L2.

The soil formation environment (**E**) of an area can be characterized using GIS techniques and local soil scientists' expertise on soil–environment relationships can be used to approximate the interaction (f) between soil and its environment in an area (Fig. 2). Soil spatial information across an area can then be inferred by combining information on the soil formative environment (**E**) with soil scientists' expertise (f).

Soil landscape is considered a continuum under SoLIM and change of soil properties across space is often gradual. Soil classes are ideal concepts of soils and soils in fields are usually intermediate to these concepts. Under this perception, the soil (S) at any point or location is assumed, in SoLIM, to be similar in varying degrees to a prescribed set of soil taxonomic units (such as soil series) or central concepts (Zhu, 1996). Soil at any point or location (i,j) can then be expressed by an n-element vector (soil similarity vector), $\mathbf{S}_{ij} = (S^1_{ij}, S^2_{ij} \dots S^k_{ij} \dots S^n_{ij})$, where S^k_{ij} is a membership value that measures the similarity between the soil at point (i,j) and prescribed

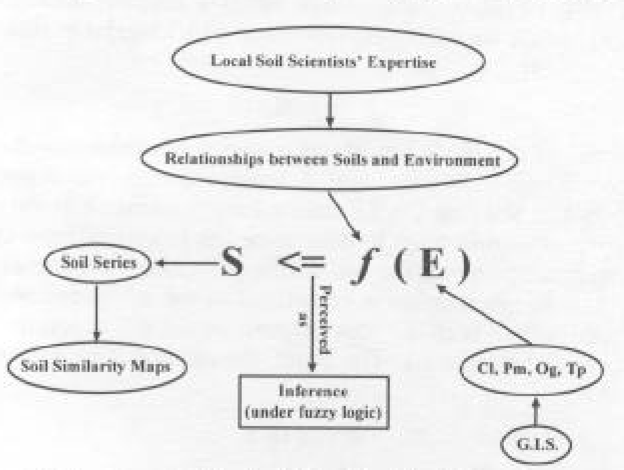


Fig. 2. The theoretical basis of SoLIM: Soil (S) is a function of its environmental factors. Cl: climate, Pm: parent materials, Og: organism, Tp: topography. Reproduced from Zhu and Band (1994) with permission from the publisher.

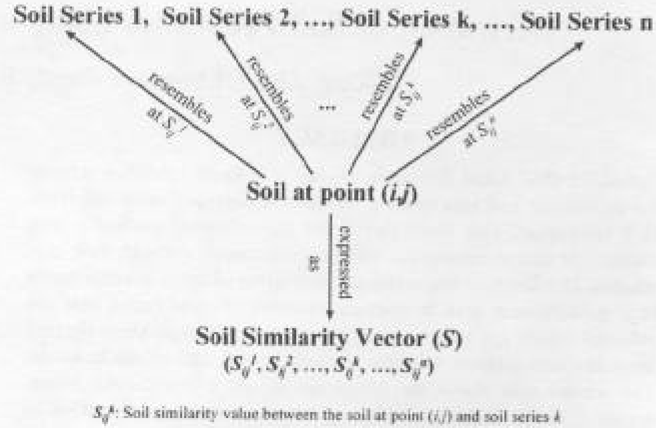


Fig. 3. Fuzzy representation of soil information using soil similarity vector.

soil taxonomic unit k (Fig. 3) and n is the number of prescribed soil taxonomic units in the area (Zhu, 1996).

This representation of soil information is different from the conventional representation. The similarity of the soil at a location to a soil taxonomic unit is expressed in terms of a membership value between 0.0 and 1.0 (that is, $0.0 \le S^k_{ij} \le 1.0$). It should be clarified that S^k_{ij} is not a probability but a fuzzy membership expressing the similarity of the soil at point (i,j) to the prescribed taxonomic unit k. It is this membership value that can be used to provide soil property values intermediate to the typical values of the prescribed soil units.

We use two case studies to illustrate the usefulness and limitations of SoLIM for deriving continuous soil property maps. The first case is the derivation of soil A horizon depth of the Lubrecht Experimental Forest, Montana. The second case is the derivation of soil water transmissivity (T) of the Redhill Catchment, Tumut, NSW, Australia.

METHODS AND STUDY AREAS

Deriving the Soil Similarity Vector

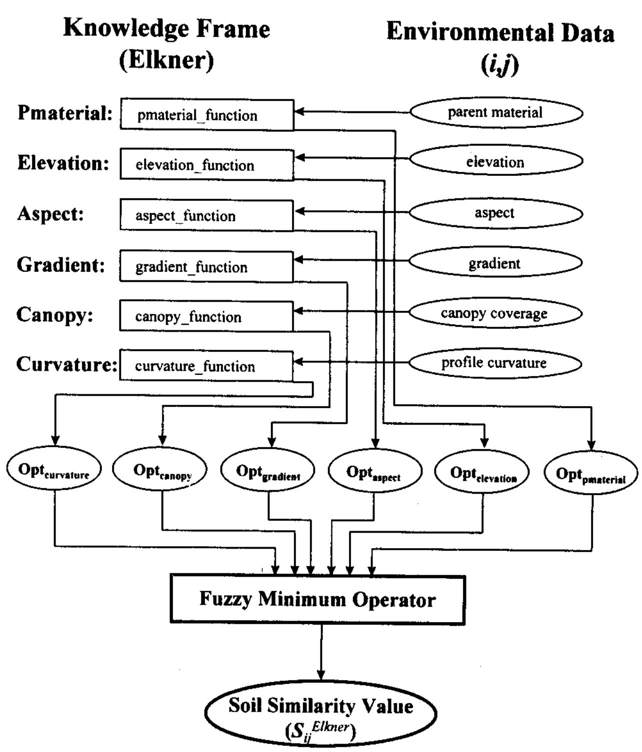
As discussed above, under SoLIM, soil at location (i,j) is represented by membership (similarity) vector $\mathbf{S}_{ij}$. The collection of $\mathbf{S}_{ij}$ across a region forms $\mathbf{S}$, which represents a set of fuzzy membership maps. Each of these maps represents the spatial distribution of similarity (membership) to a particular soil category. In other words, $\mathbf{S}$ consists of the following elements: $\mathbf{S}^1$, $\mathbf{S}^2$,.... $\mathbf{S}^k$, ... $\mathbf{S}^n$, where $\mathbf{S}^k$ represents the spatial distribution of membership to soil category k. For example, Fig. 4 shows the membership map for the Elkner soil series ($\mathbf{S}^{\text{Elkner}}$) in the Lubrecht study area.

In order to obtain these fuzzy membership maps, the basic element of the soil similarity vector, S^k_{ij}, must be determined first. The details of deriving each S^k_{ij}, as well as the process of knowledge acquisition, are beyond the scope of this study and are extensively discussed in Zhu et al. (1996) and Zhu (1994). In general, the knowledge of the relationships between the environment and soil are stored in a knowledge base in the form of a knowledge frame (Fig. 5). Each soil category would have its own knowledge frame. Each slot in the knowledge frame contains an optimality function (curve), which quantitatively describes the relationship between a given soil and a particular environmental factor. This optimality function

Fig. 4. Membership map (similarity map) of the Elkner soil series (light tone indicates a high membership value).

was derived from a soil expert through the knowledge acquisition process (Zhu et al., 1996). For example, the slot labeled *Elevation* in Fig. 5 contains the optimality function ("elevation function") describing the relationship between elevation and the soil series Elkner. To derive S_{ij}^k, the inference engine takes a set of environmental data for pixel (i,j) from the GIS database and uses the respective functions in the knowledge frame for soil category k to calculate a set of optimality values. An optimality value is defined as the propensity for a soil to develop under one given environmental condition (Zhu et al., 1996). Since each of the functions in the frame will produce its own optimality value, the soil membership value (soil similarity value), S_{ij}^k, is the minimum of these optimality values (Zhu et al., 1996). Once this process of generating soil similarity values for all locations (pixels) in an area is completed, $\mathbf{S}^k$ (the membership map for soil category k) is then derived. The entire process of deriving a soil membership map would be repeated for all other soil categories in the region to generate all other membership maps.

At the completion of the above inference process, a soil similarity vector is created for each location in the area. The soil similarity vectors for some field sites in the Lubrecht study area are listed in Table 1. For example, the soil at location Lub06−03 does not resemble any of the soil series developed on metamorphosed sedimentary (the Belt rocks) and limestone parent materials. The soil at that point, however, bears different degrees of similarities to the soil series (Ambrant, Rochester, Elkner, and Ovando) developed on the granite parent materials. These different degrees of similarities express that the soil at location Lub06−03 is intermediate to these soil series. It is these different degrees of similarities in the vectors that allow us to derive soil property values intermediate to the typical



Pmaterial: parent material
Opt: optimality value

Fig. 5. Derivation of soil similarity value for soil series Elkner.

Fig. 4. Membership map (similarity map) of the Elkner soil series (light tone indicates a high membership value).

was derived from a soil expert through the knowledge acquisition process (Zhu et al., 1996). For example, the slot labeled *Elevation* in Fig. 5 contains the optimality function ("elevation function") describing the relationship between elevation and the soil series Elkner. To derive S_{ij}^k, the inference engine takes a set of environmental data for pixel (i, j) from the GIS database and uses the respective functions in the knowledge frame for soil category k to calculate a set of optimality values. An optimality value is defined as the propensity for a soil to develop under one given environmental condition (Zhu et al., 1996). Since each of the functions in the frame will produce its own optimality value, the soil membership value (soil similarity value), S_{ij}^k, is the minimum of these optimality values (Zhu et al., 1996). Once this process of generating soil similarity values for all locations (pixels) in an area is completed, $\mathbf{S}^k$ (the membership map for soil category k) is then derived. The entire process of deriving a soil membership map would be repeated for all other soil categories in the region to generate all other membership maps.

At the completion of the above inference process, a soil similarity vector is created for each location in the area. The soil similarity vectors for some field sites in the Lubrecht study area are listed in Table 1. For example, the soil at location Lub06−03 does not resemble any of the soil series developed on metamorphosed sedimentary (the Belt rocks) and limestone parent materials. The soil at that point, however, bears different degrees of similarities to the soil series (Ambrant, Rochester, Elkner, and Ovando) developed on the granite parent materials. These different degrees of similarities express that the soil at location Lub06−03 is intermediate to these soil series. It is these different degrees of similarities in the vectors that allow us to derive soil property values intermediate to the typical

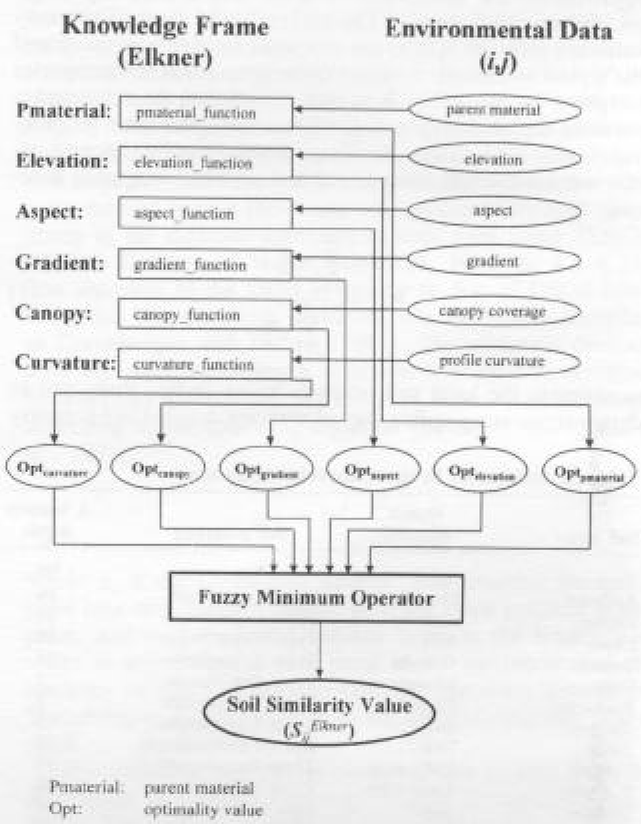


Fig. 5. Derivation of soil similarity value for soil series Elkner.

Table 1. Soil similarity values for some selective points in the Lubrecht area.

	Granite parent materials				Limestone parent materials			Belt rock parent materials				
Point	Ambrant	Rochester	Elkner	Ovando	Repp	Trapps	Whitore	Evaro	Tevis	Winkler Cool	Winkler	Sharrott
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
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
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 – 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
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

† Membership values range from 0 to 100.

property values of the prescribed soil series. The ability to produce these intermediate values enables us to estimate the spatial gradation of soil properties.

Derivation of Soil Property Maps

In this study, the soil A horizon depth and soil water transmissivity were chosen to demonstrate how the membership maps can be used to derive soil property maps. It was assumed that the more the local soil formative environment characterised by a GIS resembles the environment of a given soil category, the closer the property values of the local soil to the property values of that candidate soil category. The resemblance between the environment for soil at (i,j) and the environment for soil category k can be expressed as S_{ij}^k, which can be used as an index to measure the level of resemblance between the soil property values of the local soil and those of soil category k. In this sense, if we know $\mathbf{S}_{ij}$ [the soil similarity vector at (i,j)] and the typical value of a given soil property (such as A horizon depth) for each of the prescribed soil categories, we would be able to derive an estimation of the soil property value at (i,j). This estimated soil property value would better approximate the local value at (i,j) than that of any single prescribed soil category. The estimation of the soil property value at a given point (i,j) can take some form of combination of the typical soil property values of the prescribed soil categories weighted by $\mathbf{S}_{ij}$. While it is understood that the relationship between the similarity values of two soils and their property values may not be linear, for simplicity and as a first approximation we use the following linear and additive weighting function:

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

to estimate the local soil property value in this study and to demonstrate one possible way of deriving detailed soil property values across space. In Eq. [1], V_{ij} is the estimated soil property value at site (i,j), V^k is the typical value (often the mean value) of a given soil property of soil category k, and n is the total number of prescribed soil categories in the area. The typical A horizon depths for soil series in the Lubrecht study area (Table 2) were extracted from the soil series descriptions (Missoula County Soil Survey, 1983) and the prescribed soil transmissivity values for the soil types in the Redhill Catchment (Table 3) were obtained from a field report (Vertessy, 1990). This report defines soil transmissivity as the depth-integrated saturated hydraulic conductivity measured using the well permeameter method described in Talsma and Hallam (1980).

Table 2. Soil series in the Lubrecht study area.

Soil series	Parent material	Soil subgroup	A horizon depth
			cm
Ambrant	Granite	Udic Ustochrepts	10
Elkner	Granite	Typic Cryochrepts	20
Evaro	Belt	Typic Cryochrepts	30
Ovando	Granite	Typic Cryorthents	15
Repp	Limestone	Typic Ustochrepts	5
Rochester	Granite	Typic Ustorthents	5
Sharrott	Belt	Lithic Ustochrepts	5
Tevis	Belt	Dystric Eutrochrepts	20
Trapps	Limestone	Typic Eutroboralfs	15
Whitore	Limestone	Typic Cryochrepts	20
Winkler	Belt	Udic Ustochrepts	10
Winkler (Cool)	Belt	Udic Ustochrepts	15

Study Areas and Environmental Variables

The Lubrecht Study Area

Lubrecht Experimental Forest is located about 50 km northeast of Missoula, MT. The soil expert (Mr. Barry Dutton, a certified soil scientist) has extensive soil mapping experience in the area. The study area is centered around the North Fork of Elk Creek. The elevation in the area ranges from 1160 to 1930 m, with the highest elevations in the east and southwest and the lowest elevations in the northwest. The area is considered semiarid to semihumid (Nimlos, 1986). Most of the mountain slopes in the study area are forested, dominated by Douglas-fir [*Pseudotsuga menziesii* (Mirbel) Franco], although lesser amounts of western larch (*Larix occidentalis* Nutt.) and ponderosa pine (*Pinus ponderosa* Douglas ex P. Lawson & Lawson) are present. A small portion of the study area (in the northwest) is covered by Idaho fescue (*Festuca idahoensis* Elmer) and bluebunch wheatgrass [*Elytrigia spicata* (Pursh) D.R. Dewey]. There are four types of soil parent materials in the area: Belt rocks (metamorphosed sedimentary rocks), granite, limestone, and recent alluvium (Nimlos, 1986). The alluvial materials occur only in limited areas along the North Fork and South Fork of Elk Creek. The first three soil parent materials make up the majority of the study area with Belt rocks in the north, granite in the south, and limestone through the center part of the area.

Table 3. Variation in hydraulic conductivity (K_s) and transmissivity (T) within soil types at Redhill.

Soil type	U.S. classification equivalents, Great groups	K_s (0–50 cm) Mean	SD	K_s (50–80 cm) Mean	SD	T† (0–80 cm) Mean	SD
		cm d^{-1}				cm^2 d^{-1}	
Midslope duplex	Paleustalfs	16.31	7.87	16.49	7.76	1310.2	693.85
Sandy lower slope soils	Paleustalfs	28.67	13.54	5.34	3.15	1593.7	759.93
Shallow red earths	Paleustalfs	9.40	1.68	8.85	13.65	735.5	493.63
Upslope red earths	Agriustolls	55.18	28.31	19.69	7.51	3349.7	1559.80
Valley floor soils	Albaquults	7.25	5.41	1.01	2.62	392.8	328.91

† $T = K_{s(0-50)}(50-0) + K_{s(50-80)}(80-50)$.

There are 14 soil series in the soil map of the study area. Aquepts is the only soil suborder on the recent alluvium and it was not included in this study. According to the soil expert, soil series Mitten does not actually occur in the study area, although it was included on the soil map (this soil expert did not make the soil map). Therefore, soil series Mitten was not included in the study. The remaining 12 soil series are listed in Table 2 and the descriptions of these soil series can be found in the Missoula County soil survey report (Missoula County Soil Survey, 1983). It should be noted that the reason for using soil series as the basic soil taxonomic unit for the soil similarity vectors is that soil series is the taxonomic unit that has been extensively used in soil surveys and soil mapping in the USA. Local soil experts feel more comfortable with soil series than any other taxonomic units. There has accumulated a good deal of knowledge about the relationships between soil series and their environments. This knowledge can be used to produce soil similarity vectors (Zhu and Band, 1994; Zhu et al., 1996).

Six environmental variables (elevation, slope aspect, slope gradient, tree canopy coverage, profile curvature, and parent materials) were used to characterize the soil formative environment. The inclusion of these variables was determined through the discussion between the knowledge engineer (the person who performs the knowledge acquisition; in our study, it was the senior author) and the soil expert in the light of data availability and the importance of the variables for delineating the soil formative environment. Note that there are no data variables in the list that directly measure climatic factors. Although the study was conducted on a small drainage basin, great differences in terms of microclimate do exist within the basin. However, these differences in microclimate are well expressed by variations in elevation, slope aspect, and slope gradient.

Information on elevation, slope aspect, slope gradient, and profile curvature were derived from a DEM of the study area. The accuracy of the DEM is comparable with the accuracy of the Level 1 USGS 7.5-minute DEMs (U.S. Geological Survey, 1990). The tree canopy coverage data was approximated with an index derived from remotely sensed data (Thematic Mapper) (Nemani et al., 1993). Information on parent material was obtained from a bedrock geology map of the study area (Brenner, 1968). It is recognized that the inclusion of a bedrock geology map as the soil parent material map would affect the quality of the results because geological maps were produced in the same way as soil maps and potentially contain human errors. The other limitation of using bedrock geology as soil parent material is that the bedrock geology cannot sufficiently represent surficial geology from which soils were developed. However, the properties of soils in the region exhibit a great dependence on the properties of their respective parent materials and the bedrock geology was the only geological information available for the area at the time of this study. Therefore, it was necessary to use the bedrock geology to approximate the parent material information.

These data variables have by no means exhausted the soil formation factors and the interaction of these factors on soil development. They were used to demonstrate the potential of the new methodology of gathering and representing soil spatial information.

The Redhill Catchment

The Redhill catchment is located near Tumut, NSW, about 100 km west of Canberra (35.1°S, 148.4°E). The test area is about 2.2 km (north–south) by 2.5 km (east–west). The catchment was cleared of its eucalypt vegetation and planted to pasture >100 yr ago (Vertessy, 1990). In April 1989 catchment was planted to Monterey pine (*Pinus radiat* Don). Elevations range from 650 to 780 m with a gent moderate relief. The two major parent materials are gr in the center area and basalt around the granite.

The soil expert for the study area is Dr. Rob Vertes hydrologist and ecological modeler. The soils in the catch were classified and mapped for the specific purpose of c ment water yield research. As a result, the classification not conform to any of the existing soil classification syst The soil expert identified five major soil types in the (Table 3) (Vertessy, 1990). The soils were separated each other mainly by their distinct topographic positi Within each soil type, the saturated hydraulic conductivity and soil water transmissivity (T) values vary substan (Table 3).

For this study area, five environmental variables were u parent material, wetness index, planform curvature, slope dient, and upstream drainage area. The selection of t environmental variables were based on the discussion bet the knowledge engineer and the soil expert in the area, most of the input coming from the soil expert. The soil e felt that the slope aspect plays little role in the develop of the soils in the area because the illumination condi among the different aspects are not very different due t gentle relief and the low latitude of the area. Elevation not included in the environmental variable list becaus expert thought that elevation did not contribute much t development of soils in the region due to its gentle to mod relief. Vegetation information was not used as well sinc area was cleared of vegetation and planted to Monterey very recently and the current vegetation conditions ma have a decisive role in the formation of the soils in the s area. The other reason for not including the vegetation info tion is that such information was not available at the time study was conducted. However, the inclusion of this veget information may provide additional information for identi land units and soil conditions since the planted vegetation exhibit different growth conditions on different land units under different soil conditions.

Except for the parent material data layer, all other were derived from a DEM that was generated through s fitting of the digitized elevation contour lines using TO (CSIRO Division of Water Resources, 1992, p. 4.1–4 The accuracy of the DEM is similar to that of USGS L 1 DEMs. The planform curvature was calculated acco to Zevenbergen and Thorne (1987). The upstream dra area was calculated using a divergent flow method desc by Freeman (1991) and the wetness index was calcu according to the following equation (Beven, 1986; Qui al., 1991):

$$w_{ij} = \ln\left(\frac{a_{ij}}{\tan\beta_{ij}}\right)$$

where a_{ij} is the cumulative upslope area draining throu point (per unit contour length), β_{ij} is the slope gradient a point, and w_{ij} reflects the balance between the tendenc water to accumulate at that point in the catchment an tendency for gravitational forces to move that water downs Therefore, w_{ij} is used to express the wetness condition a point.

Information on soil parent materials was created fro digitization of a bedrock geology map of the area (Vert 1990). The soil parent material data would have the si limitations on the result as that in the Lubrecht study a

Field Sampling

In order to evaluate the results from SoLIM, field observations of soil series distributions were made. Stratified point sampling, systematic point sampling, and line sampling strategies (Griffith and Amrhein, 1991) were employed for collecting field data in the study areas.

Sampling the Lubrecht Study Area

For the Lubrecht case, the evaluation of the results from SoLIM contains two aspects: the overall performance of the system and the ability to infer spatial variation of soil information. The point sampling strategy was used to obtain field data for the assessment of the overall performance of the system and the line sampling method was employed to evaluate how well SoLIM was able to capture spatial variation of soil information. For point sampling, a stratification technique was employed. First, the soil similarity vectors were hardened to produce an inferred soil series map by assigning each pixel to the soil series that has the highest membership value in the similarity vector (Zhu et al., 1996; Zhu, 1994). The area was then divided into two subsets based on the agreement between the inferred soil series map and the conventional soil series map of the study area. The first subset contains the areas where the inferred soil map matches the conventional soil map (matched areas) and the second contains the areas where the inferred soil map disagrees with the conventional soil map (mismatched areas). Within each of these subsets, samples were drawn to cover the majority of different configurations of environmental conditions. Field observations of the spatial variation of soil information were conducted on a transect. The transect was constructed in such a way that it cuts through different types of parent materials and covers the major environmental configurations in the shortest distance.

Field surveys were carried out during the summer of 1991, 1992, and 1993. A total of 64 sites were visited during these field trips. Of the 64 sites, 18 were on the transect. Twenty-eight of the remaining 46 sites were in the matched areas and 18 were in the mismatched regions. At each of the sites, several pits were dug to reduce the influence of microvariation of soil. Three types of information about each site were recorded: the location, the environment, and the soil. The A horizon depths at 33 sites were collected during the field trip of 1993.

Samples at the Redhill Catchment

Soil surveys were carried out extensively in the catchment for studying hydrologic processes (Vertessy, 1990). Soils were inspected using the augerhole method and auger sampling was confined to the upper 140 cm of the profile, although further information was obtained from a series of trenches, dug to depths of 250 cm (Vertessy, 1990). The sampling was done using a regular grid sampling strategy (systematic sampling). In addition to the profile description at each sample site, the K_s was determined using the well permeameter method described by Talsma and Hallam (1980). Measurements from 32 sites in the catchment were available for this study. At each site, measurements of the K_s at depth intervals of 0 to 50 and 50 to 80 cm were used to calculate T (Table 3) according to Dunne and Leopold (1978, p. 205):

$$T = \sum_{i=1}^{n} K_{si} D_i \qquad [2]$$

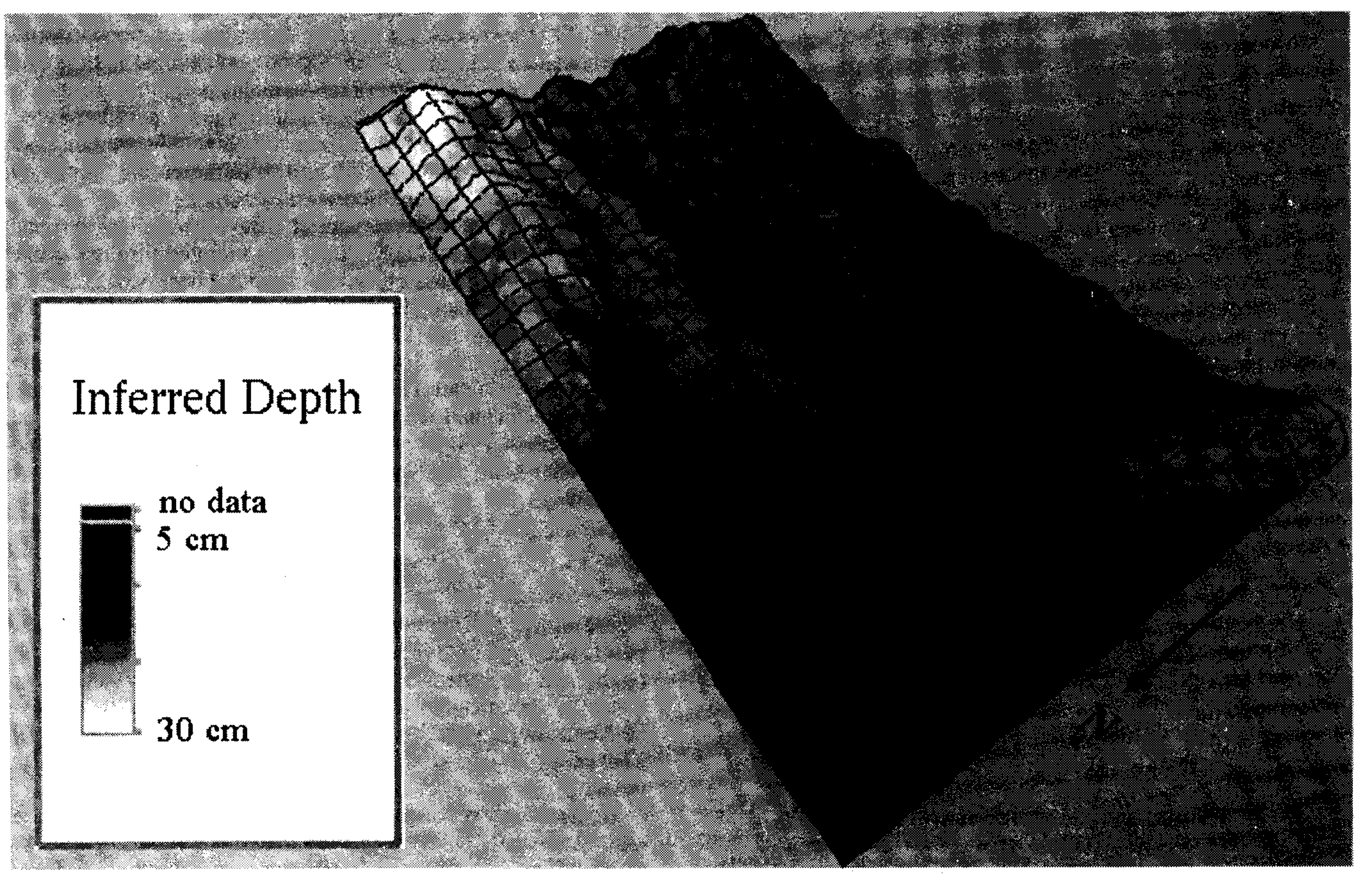


Fig. 6. Inferred soil A horizon depth of the Lubrecht area.

Field Sampling

In order to evaluate the results from SoLIM, field observations of soil series distributions were made. Stratified point sampling, systematic point sampling, and line sampling strategies (Griffith and Amrhein, 1991) were employed for collecting field data in the study areas.

Sampling the Lubrecht Study Area

For the Lubrecht case, the evaluation of the results from SoLIM contains two aspects: the overall performance of the system and the ability to infer spatial variation of soil information. The point sampling strategy was used to obtain field data for the assessment of the overall performance of the system and the line sampling method was employed to evaluate how well SoLIM was able to capture spatial variation of soil information. For point sampling, a stratification technique was employed. First, the soil similarity vectors were hardened to produce an inferred soil series map by assigning each pixel to the soil series that has the highest membership value in the similarity vector (Zhu et al., 1996; Zhu, 1994). The area was then divided into two subsets based on the agreement between the inferred soil series map and the conventional soil series map of the study area. The first subset contains the areas where the inferred soil map matches the conventional soil map (matched areas) and the second contains the areas where the inferred soil map disagrees with the conventional soil map (mismatched areas). Within each of these subsets, samples were drawn to cover the majority of different configurations of environmental conditions. Field observations of the spatial variation of soil information were conducted on a transect. The transect was constructed in such a way that it cuts through different types of parent materials and covers the major environmental configurations in the shortest distance.

Field surveys were carried out during the summer of 1991, 1992, and 1993. A total of 64 sites were visited during these field trips. Of the 64 sites, 18 were on the transect. Twenty-eight of the remaining 46 sites were in the matched areas and 18 were in the mismatched regions. At each of the sites, several pits were dug to reduce the influence of microvariation of soil. Three types of information about each site were recorded: the location, the environment, and the soil. The A horizon depths at 33 sites were collected during the field trip of 1993.

Samples at the Redhill Catchment

Soil surveys were carried out extensively in the catchment for studying hydrologic processes (Vertessy, 1990). Soils were inspected using the augerhole method and auger sampling was confined to the upper 140 cm of the profile, although further information was obtained from a series of trenches, dug to depths of 250 cm (Vertessy, 1990). The sampling was done using a regular grid sampling strategy (systematic sampling). In addition to the profile description at each sample site, the K_s was determined using the well permeameter method described by Talsma and Hallam (1980). Measurements from 32 sites in the catchment were available for this study. At each site, measurements of the K_s at depth intervals of 0 to 50 and 50 to 80 cm were used to calculate T (Table 3) according to Dunne and Leopold (1978, p. 205):

$$T = \sum_{i=1}^{n} K_{si} D_i \qquad [2]$$

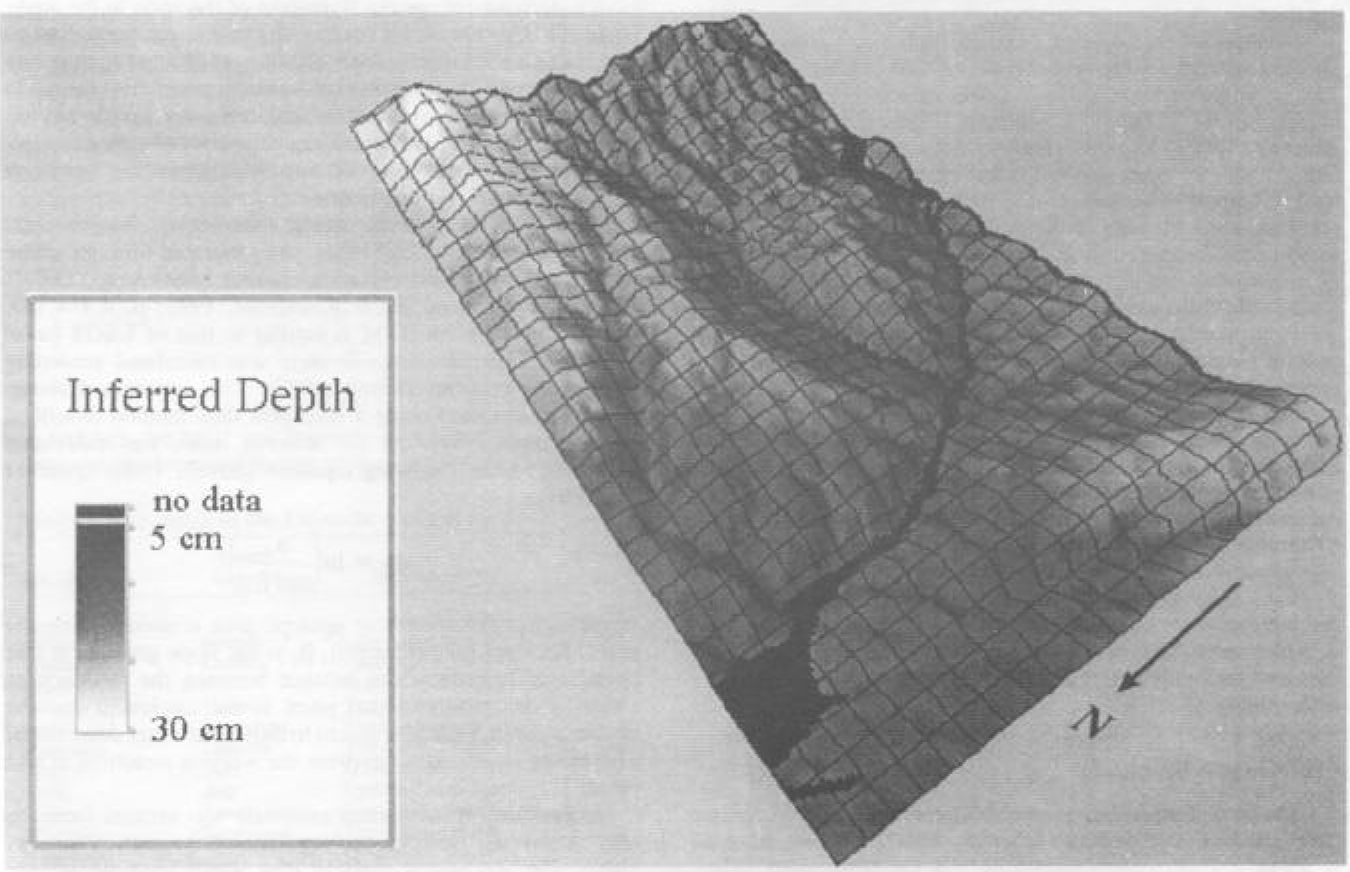


Fig. 6. Inferred soil A horizon depth of the Lubrecht area.

where K_{si} and D_i are the hydraulic conductivity and thickness of the ith horizon.

RESULTS

The Lubrecht Area

The inferred and mapped depths of the A horizons of the soils in the study area are shown in Fig. 6 and 7. The image of mapped depth was generated by assigning each pixel the typical depth of the soil series of that pixel area. The map of inferred depth was generated by applying Eq. [1] to every pixel in the area. The contrast between the two maps (Fig. 6 and 7) is strong and clear. The image of inferred depth shows a spatially continuous pattern of the A horizon depth while the image of depths from the soil map inherits the exact spatial pattern of the soil map. Because the study area is in the semihumid to semiarid area of western Montana, the soils on north-facing slopes and at higher elevations are further developed than soils on south-facing slopes and at lower elevations due to the limited moisture conditions at the lower elevations and on south-facing slopes. The A horizons are deeper for those soils on north-facing slopes and at higher elevations than those of soils on south-facing slopes and at lower elevations. The inferred depth map shows this characteristic of the depth of the A horizon in the area but the change of depth is gradual across space and is not like the change depicted in the depth image derived from the conventional soil map.

The changes in A horizon depth across space depicted by SoLIM match the changes of depth observed in the field better than the changes outlined in the soil map (Fig. 8), although neither are good reproductions of the strong local field variability. The spatial variation of soil depths is shown as a step function on the soil map. However, the field observations show that the spatial variation of soil depths vary in a form very different from a step function. The inferred soil depths along the transect follow the values of field observations better.

The two scatter plots (Fig. 9 and 10) compare the estimation of soil A horizon depths using SoLIM with that derived from the soil map at the 33 sample sites where observations on A horizon depths were made. From these two plots, it can be seen that the inferred depths at these sites are more closely associated with the observed depths in the field than are the depths derived from the soil map. Although both correlation coefficients are highly significant (at the 0.001 probability level), the correlation between the inferred depths and the observed depths is much stronger than that between the depths from the soil map and the observed depths (numbers in the white boxes in Fig. 9 and 10).

Willmott (1984) and Willmott et al. (1985) have discussed several measures (statistics) for comparing model results. They suggested that certain differences or error indices and an index of agreement should be used for quantitative evaluations of model performance (Mayer and Butler, 1993; Power, 1993; Willmott, 1984). Three

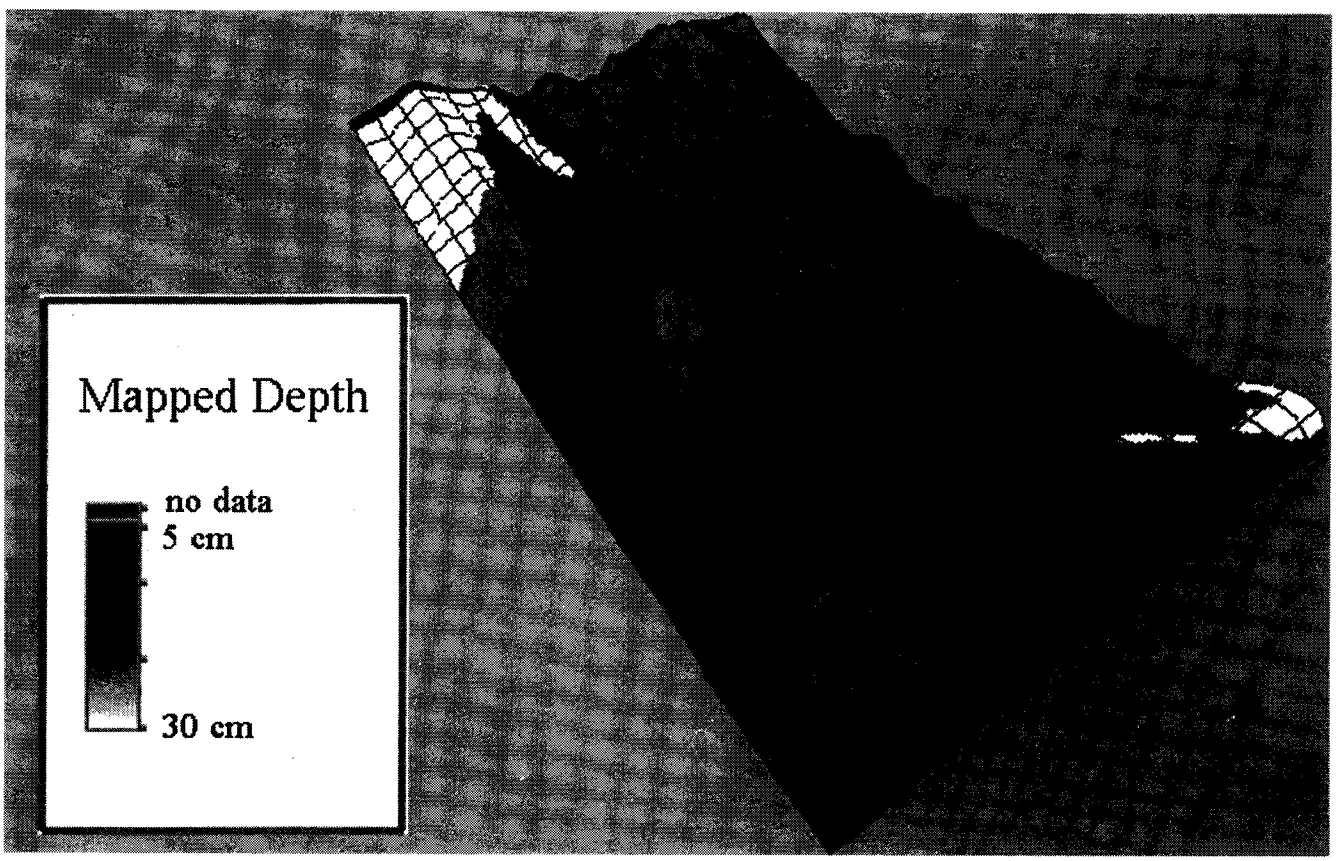


Fig. 7. A horizon depth derived from the soil map of the Lubrecht area.

where K_{si} and D_i are the hydraulic conductivity and thickness of the ith horizon.

RESULTS

The Lubrecht Area

The inferred and mapped depths of the A horizons of the soils in the study area are shown in Fig. 6 and 7. The image of mapped depth was generated by assigning each pixel the typical depth of the soil series of that pixel area. The map of inferred depth was generated by applying Eq. [1] to every pixel in the area. The contrast between the two maps (Fig. 6 and 7) is strong and clear. The image of inferred depth shows a spatially continuous pattern of the A horizon depth while the image of depths from the soil map inherits the exact spatial pattern of the soil map. Because the study area is in the semihumid to semiarid area of western Montana, the soils on north-facing slopes and at higher elevations are further developed than soils on south-facing slopes and at lower elevations due to the limited moisture conditions at the lower elevations and on south-facing slopes. The A horizons are deeper for those soils on north-facing slopes and at higher elevations than those of soils on south-facing slopes and at lower elevations. The inferred depth map shows this characteristic of the depth of the A horizon in the area but the change of depth is gradual across space and is not like the change depicted in the depth image derived from the conventional soil map.

The changes in A horizon depth across space depicted by SoLIM match the changes of depth observed in the field better than the changes outlined in the soil map (Fig. 8), although neither are good reproductions of the strong local field variability. The spatial variation of soil depths is shown as a step function on the soil map. However, the field observations show that the spatial variation of soil depths vary in a form very different from a step function. The inferred soil depths along the transect follow the values of field observations better.

The two scatter plots (Fig. 9 and 10) compare the estimation of soil A horizon depths using SoLIM with that derived from the soil map at the 33 sample sites where observations on A horizon depths were made. From these two plots, it can be seen that the inferred depths at these sites are more closely associated with the observed depths in the field than are the depths derived from the soil map. Although both correlation coefficients are highly significant (at the 0.001 probability level), the correlation between the inferred depths and the observed depths is much stronger than that between the depths from the soil map and the observed depths (numbers in the white boxes in Fig. 9 and 10).

Willmott (1984) and Willmott et al. (1985) have discussed several measures (statistics) for comparing model results. They suggested that certain differences or error indices and an index of agreement should be used for quantitative evaluations of model performance (Mayer and Butler, 1993; Power, 1993; Willmott, 1984). Three

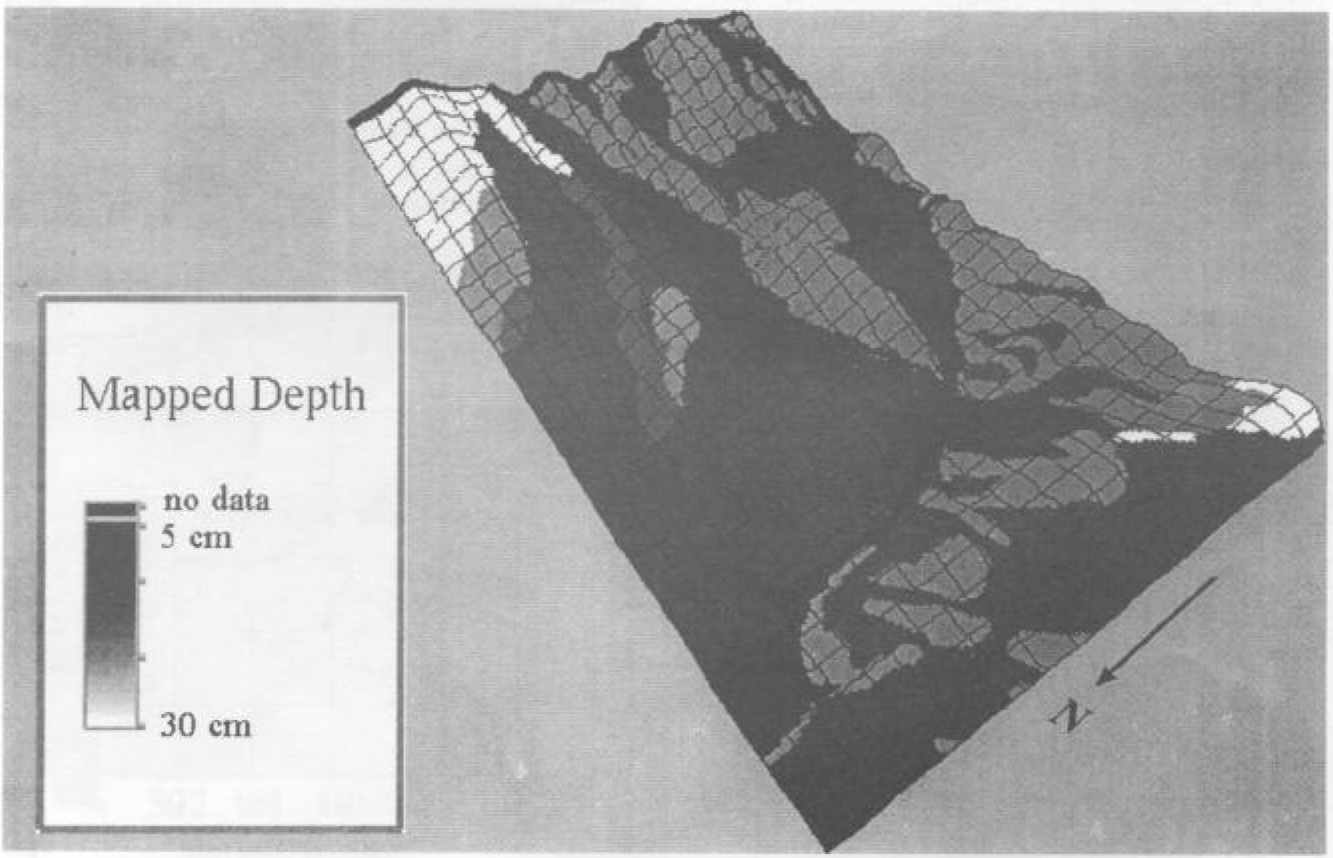


Fig. 7. A horizon depth derived from the soil map of the Lubrecht area.

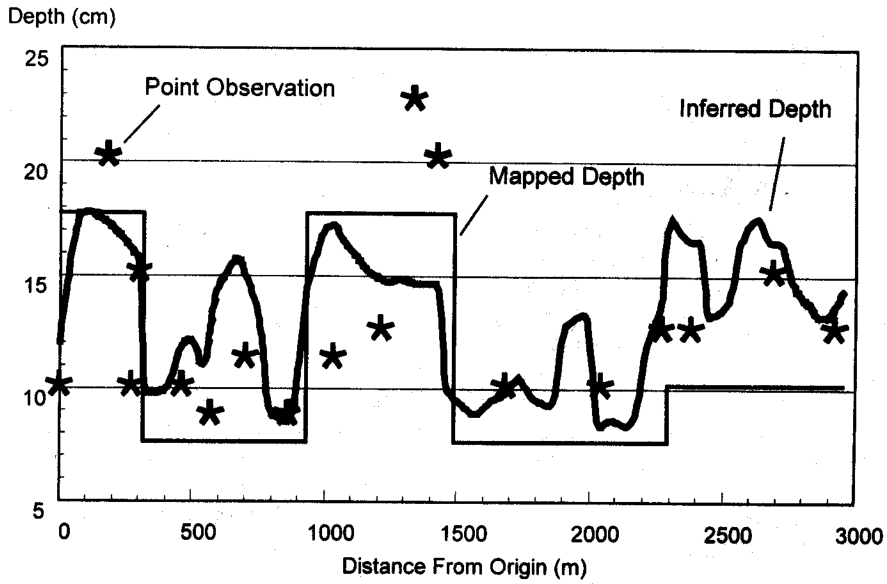


Fig. 8. Changes in soil A horizon depths along a transect in the Lubrecht area.

indices are used here to evaluate the performance of SoLIM. The first index is the MAE and the second index is the RMSE (Willmott, 1984). Both MAE and RMSE are difference measures but RMSE is more sensitive to extreme values. The smaller the values of MAE and RMSE are, the better a model performs. The third index is an AC, which is defined as (Willmott, 1984)

$$AC = 1 - \frac{N\,\mathrm{RMSE}^2}{\mathrm{PE}} \quad [3]$$

where N is the number of cases (sites or observations), PE is the potential error variance and is defined as

$$\mathrm{PE} = \sum_{j=1}^{n} (|P_i - \overline{O}| + |O_i - \overline{O}|)^2 \quad [4]$$

given that $\overline{O}$ is the observed mean, and P_i and O_i are the model predicted value and the observed (or reliable) value for the ith case, respectively. The agreement coefficient varies between 0.0 and 1.0, where a value of 1.0 expresses perfect agreement between O (observed) and P (predicted) and 0.0 describes complete disagreement (Willmott, 1984). Table 4 lists these statistics for the evaluation of the performance of SoLIM vs. that of the soil map. The MAE and RMSE statistics for SoLIM are consistently lower than those for the soil map. The AC is higher for SoLIM than that for the soil map. These statistics support the hypothesis that the estimation of the A horizon depth derived from SoLIM is more accurate than that from the soil map.

The Redhill Catchment

As in the Lubrecht case, the inferred transmissivity (Fig. 11) was calculated according to Eq. [1] and the

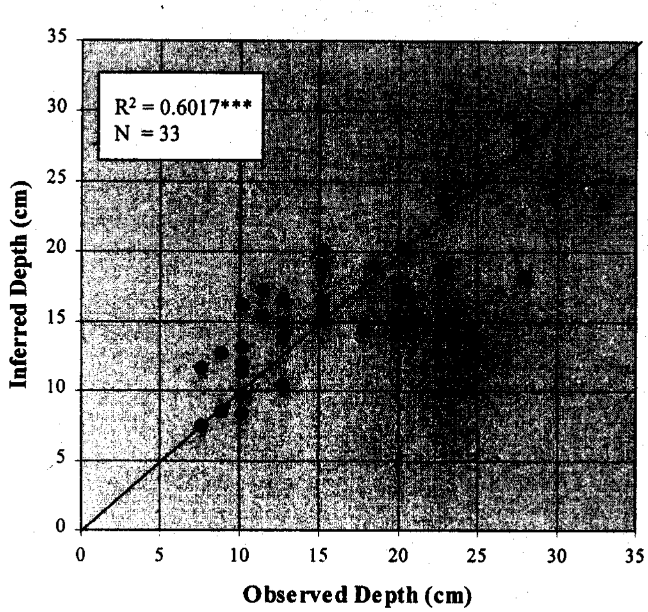


Fig. 9. Scatter plot of observed vs. inferred A horizon depths in the Lubrecht area.

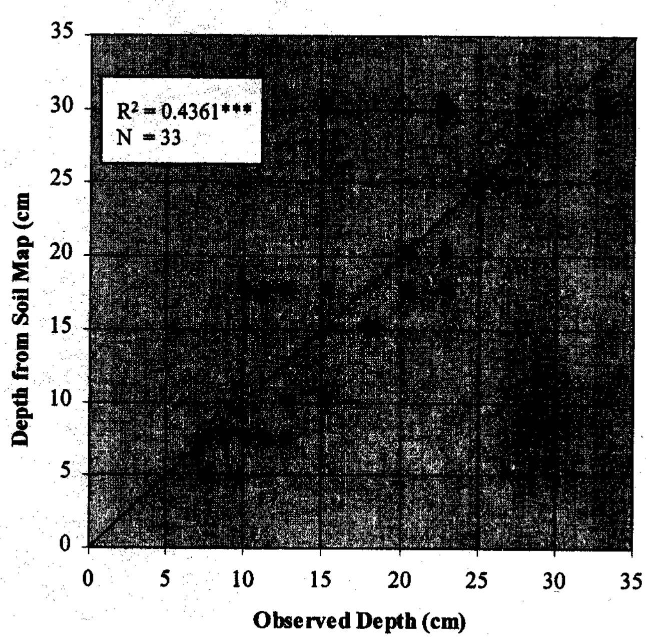


Fig. 10. Scatter plot of observed vs. mapped A horizon depths in the Lubrecht area.

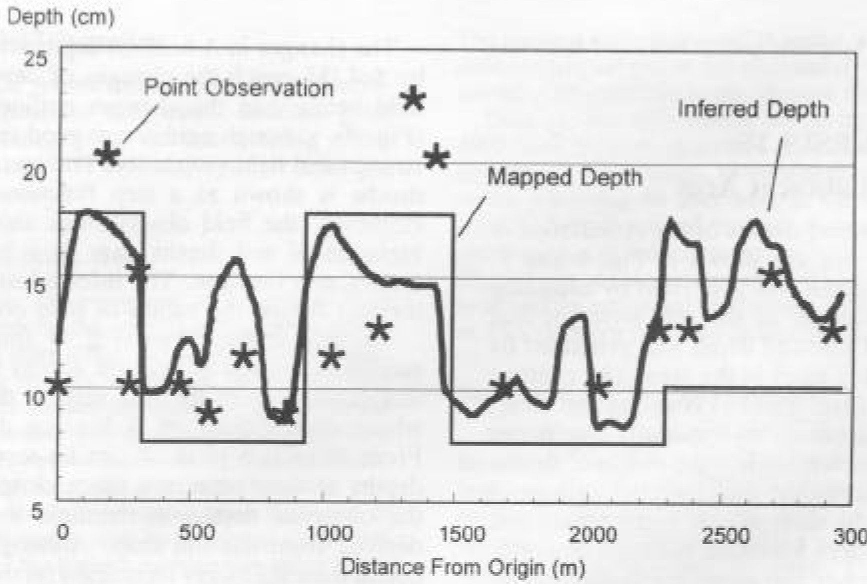


Fig. 8. Changes in soil A horizon depths along a transect in the Lubrecht area.

indices are used here to evaluate the performance of SoLIM. The first index is the MAE and the second index is the RMSE (Willmott, 1984). Both MAE and RMSE are difference measures but RMSE is more sensitive to extreme values. The smaller the values of MAE and RMSE are, the better a model performs. The third index is an AC, which is defined as (Willmott, 1984)

$$AC = 1 - \frac{N\,\mathrm{RMSE}^2}{\mathrm{PE}} \quad [3]$$

where N is the number of cases (sites or observations), PE is the potential error variance and is defined as

$$\mathrm{PE} = \sum_{j=1}^{n} (|P_i - \overline{O}| + |O_i - \overline{O}|)^2 \quad [4]$$

given that $\overline{O}$ is the observed mean, and P_i and O_i are the model predicted value and the observed (or reliable) value for the ith case, respectively. The agreement coefficient varies between 0.0 and 1.0, where a value of 1.0 expresses perfect agreement between O (observed) and P (predicted) and 0.0 describes complete disagreement (Willmott, 1984). Table 4 lists these statistics for the evaluation of the performance of SoLIM vs. that of the soil map. The MAE and RMSE statistics for SoLIM are consistently lower than those for the soil map. The AC is higher for SoLIM than that for the soil map. These statistics support the hypothesis that the estimation of the A horizon depth derived from SoLIM is more accurate than that from the soil map.

The Redhill Catchment

As in the Lubrecht case, the inferred transmissivity (Fig. 11) was calculated according to Eq. [1] and the

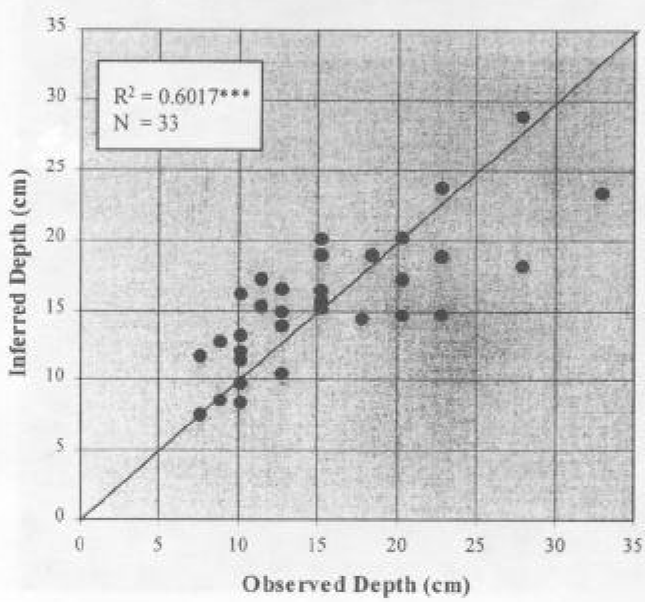


Fig. 9. Scatter plot of observed vs. inferred A horizon depths in the Lubrecht area.

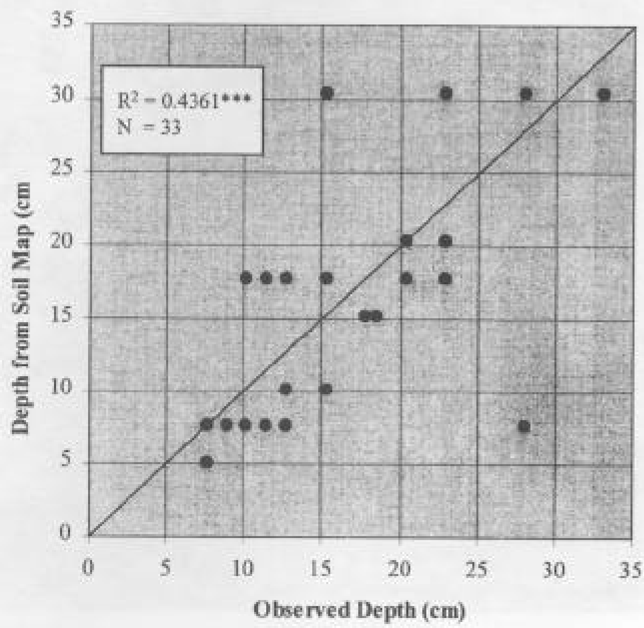


Fig. 10. Scatter plot of observed vs. mapped A horizon depths in the Lubrecht area.

Table 4. Statistics on the performance of SoLIM vs. the soil map in Lubrecht.

Quantitative measures	Observed mean	Predicted mean	MAE	RMSE	AC
SoLIM	15.49	15.58	3.04	4.00	0.85
Soil map	15.49	14.39	4.41	5.92	0.80

mapped transmissivity (Fig. 12) was derived by assigning the typical transmissivity value of each soil type to the pixels belonging to that soil type. From the comparison of these two figures (Fig. 11 and 12), the inferred transmissivity map shows more spatial gradation of transmissivity than the transmissivity map derived from the existing soil map.

The statistics on the transmissivity prediction using SoLIM and using the existing soil map at the 32 field sites in the Redhill Catchment are listed in Table 5. Of the four indices, the predicted mean value from SoLIM is closer to the observed mean than that from the soil map. The MAE produced from SoLIM is smaller than that from the soil map. The RMSE produced from SoLIM is greater than that from the soil map. The difference between the two RMSEs is relatively small. However, the AC value for the soil map is more than twice the AC value for SoLIM. The AC is an index that measures the agreement between the predicted and the observed values. It may be concluded based on these statistics that the soil water transmissivity map produced using SoLIM did not provide additional accuracy over the transmissivity map derived from the soil map of the Redhill catchment.

DISCUSSION

The SoLIM approach was able to provide more accurate soil information than the soil map with respect to both the spatial gradation and point accuracy in the Lubrecht study area. In the Lubrecht case, the soil category used was soil series. Soil series is a standard soil classification category used as the basic unit during extensive soil surveys. Therefore, a great deal of knowledge on the relationships between soil series and their formative environment was accumulated by soil experts during these soil surveys. This knowledge was captured through the knowledge acquisition process and was represented in SoLIM. In addition, the Lubrecht area is a mountainous region with a strong environmental gradient and the GIS techniques employed were able to capture this environmental gradient for characterizing the soil formative environment. Therefore, SoLIM has many advantages for soil information gathering and representation over conventional soil maps in terms of spatial details and attribute accuracy in the Lubrecht area.

In the Redhill case in Australia, SoLIM was less successful in comparison with the existing soil map. In the Redhill area, the soil units were created in an ad hoc fashion for the specific use of hydrological modeling and not for any standard soil classification. The understanding of the relationships between these soil units and their formative environments is limited to very few environmental variables such as topographic positions and landforms. On the other hand, the GIS techniques used were not able to successfully capture topographic positions and landforms. Therefore, much of the knowl-

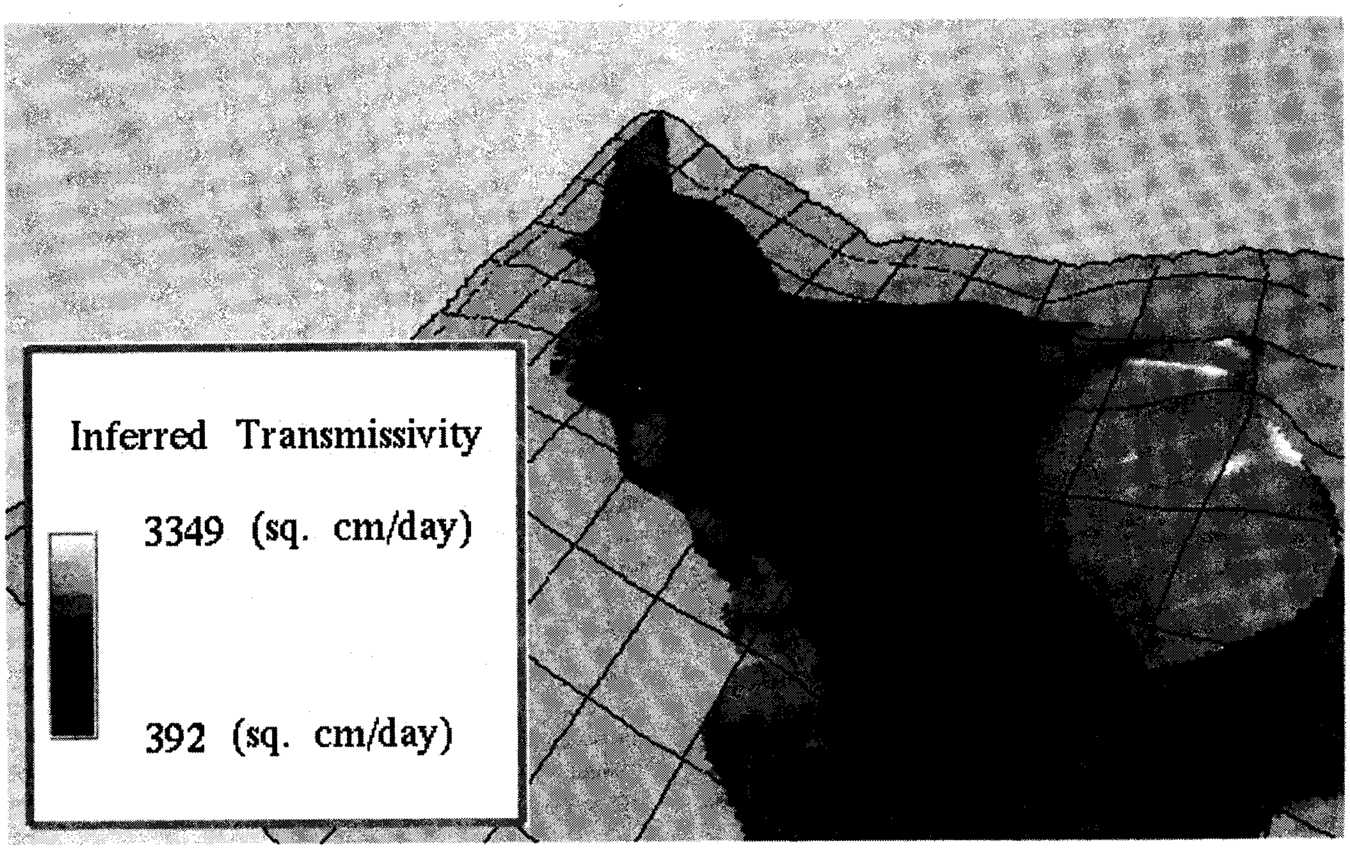


Fig. 11. Inferred transmissivity map of the Redhill study area.

Table 4. Statistics on the performance of SoLIM vs. the soil map in Lubrecht.

Quantitative measures	Observed mean	Predicted mean	MAE	RMSE	AC
SoLIM	15.49	15.58	3.04	4.00	0.85
Soil map	15.49	14.39	4.41	5.92	0.80

mapped transmissivity (Fig. 12) was derived by assigning the typical transmissivity value of each soil type to the pixels belonging to that soil type. From the comparison of these two figures (Fig. 11 and 12), the inferred transmissivity map shows more spatial gradation of transmissivity than the transmissivity map derived from the existing soil map.

The statistics on the transmissivity prediction using SoLIM and using the existing soil map at the 32 field sites in the Redhill Catchment are listed in Table 5. Of the four indices, the predicted mean value from SoLIM is closer to the observed mean than that from the soil map. The MAE produced from SoLIM is smaller than that from the soil map. The RMSE produced from SoLIM is greater than that from the soil map. The difference between the two RMSEs is relatively small. However, the AC value for the soil map is more than twice the AC value for SoLIM. The AC is an index that measures the agreement between the predicted and the observed values. It may be concluded based on these statistics that the soil water transmissivity map produced using SoLIM did not provide additional accuracy over the transmissivity map derived from the soil map of the Redhill catchment.

DISCUSSION

The SoLIM approach was able to provide more accurate soil information than the soil map with respect to both the spatial gradation and point accuracy in the Lubrecht study area. In the Lubrecht case, the soil category used was soil series. Soil series is a standard soil classification category used as the basic unit during extensive soil surveys. Therefore, a great deal of knowledge on the relationships between soil series and their formative environment was accumulated by soil experts during these soil surveys. This knowledge was captured through the knowledge acquisition process and was represented in SoLIM. In addition, the Lubrecht area is a mountainous region with a strong environmental gradient and the GIS techniques employed were able to capture this environmental gradient for characterizing the soil formative environment. Therefore, SoLIM has many advantages for soil information gathering and representation over conventional soil maps in terms of spatial details and attribute accuracy in the Lubrecht area.

In the Redhill case in Australia, SoLIM was less successful in comparison with the existing soil map. In the Redhill area, the soil units were created in an ad hoc fashion for the specific use of hydrological modeling and not for any standard soil classification. The understanding of the relationships between these soil units and their formative environments is limited to very few environmental variables such as topographic positions and landforms. On the other hand, the GIS techniques used were not able to successfully capture topographic positions and landforms. Therefore, much of the knowl-

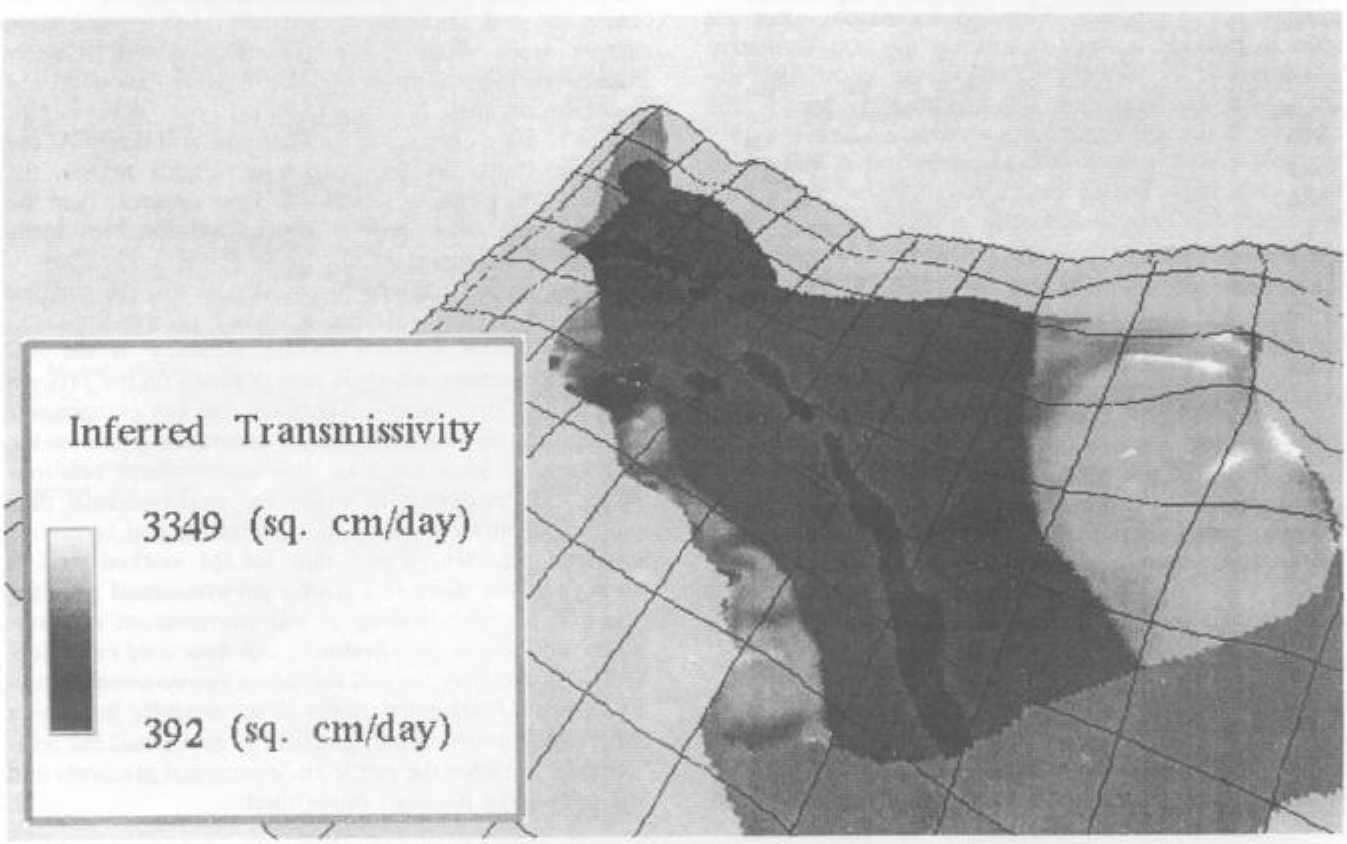


Fig. 11. Inferred transmissivity map of the Redhill study area.

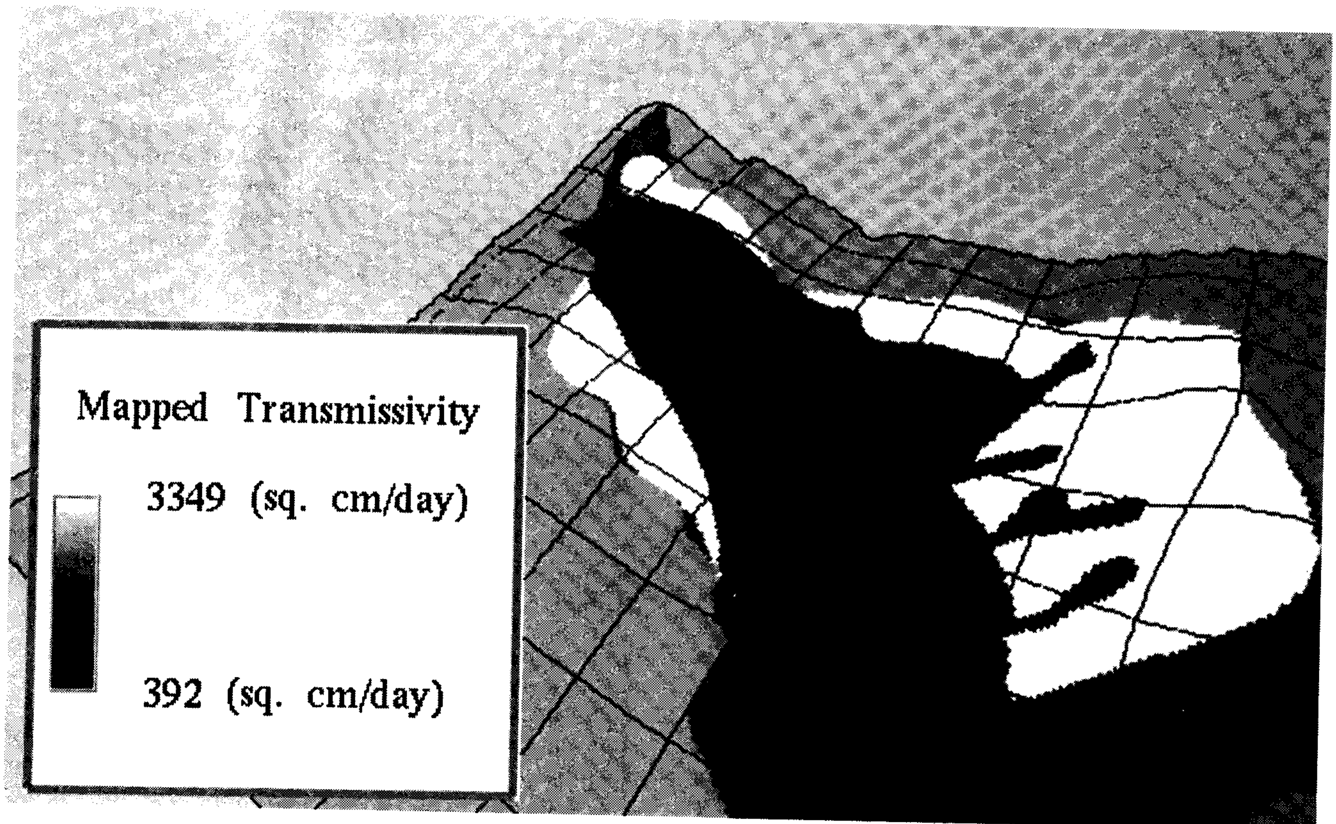


Fig. 12. Mapped transmissivity of the Redhill study area.

edge on the soil–environmental relationships was not utilized by SoLIM. In addition, the Redhill catchment is a small basin, about 173 ha, and has a very gentle environmental gradient. The GIS techniques were not able to provide enough details on the soil formative environment for SoLIM to utilize the knowledge extracted for soil inference. The soil property used in this exercise is the soil water transmissivity, which is highly variable across a very short distance and is difficult to map accurately. Under these circumstances, the SoLIM approach experienced difficulty in deriving an accurate continuous transmissivity map for the Redhill catchment. The result from SoLIM was only comparable to that from the soil map produced through many soil observations across a small area.

SUMMARY AND CONCLUSIONS

We derived soil property values across a region from the soil similarity vectors derived using SoLIM and demonstrated the usefulness of this fuzzy representation of soil information. From the above illustrations, this fuzzy representation of soil information does have some advantages over conventional representation of soil information under crisp logic (such as conventional soil survey maps). At the spatial level, the soil property map derived using the soil similarity vectors shows more gradation across space. This spatial gradation of soil property values provides a more realistic representation of soil information than the step-function type of variation, which is often adopted in conventional soil maps. At the attribute level, for the steep mountainous region, the derived soil property values are more accurate than the soil property values derived from conventional soil maps but not as advantageous for more gentle landscapes.

Although the fuzzy representation of soil information does have advantages, the accuracy of the final soil property values depends on the accuracy of the soil similarity vectors, which in turn depends on the process generating these vectors. There are two key components during the soil similarity vector generation process: the precision of knowledge on soil–environment relationships and the degree to which the environmental data employed can be used to characterize the soil formative environment. We showed that SoLIM worked well in an area where there is a strong environmental gradient and a rich understanding of soil–environment relationships, and where the environmental data used can effectively characterize the soil formative environment. However, SoLIM did not perform as successfully in an area where the environmental gradient is gentle and the relationships between the gentle environmental gradients and the soils were not well established.

Table 5. Statistics on the performance of SoLIM vs. the soil map in Redhill.

Quantitative measures	Observed mean	Predicted mean	MAE	RMSE	AC
	$cm^2\ d^{-1}$				
SoLIM	1406	1383	838	1173	0.30
Soil map	1406	1622	902	1112	0.66

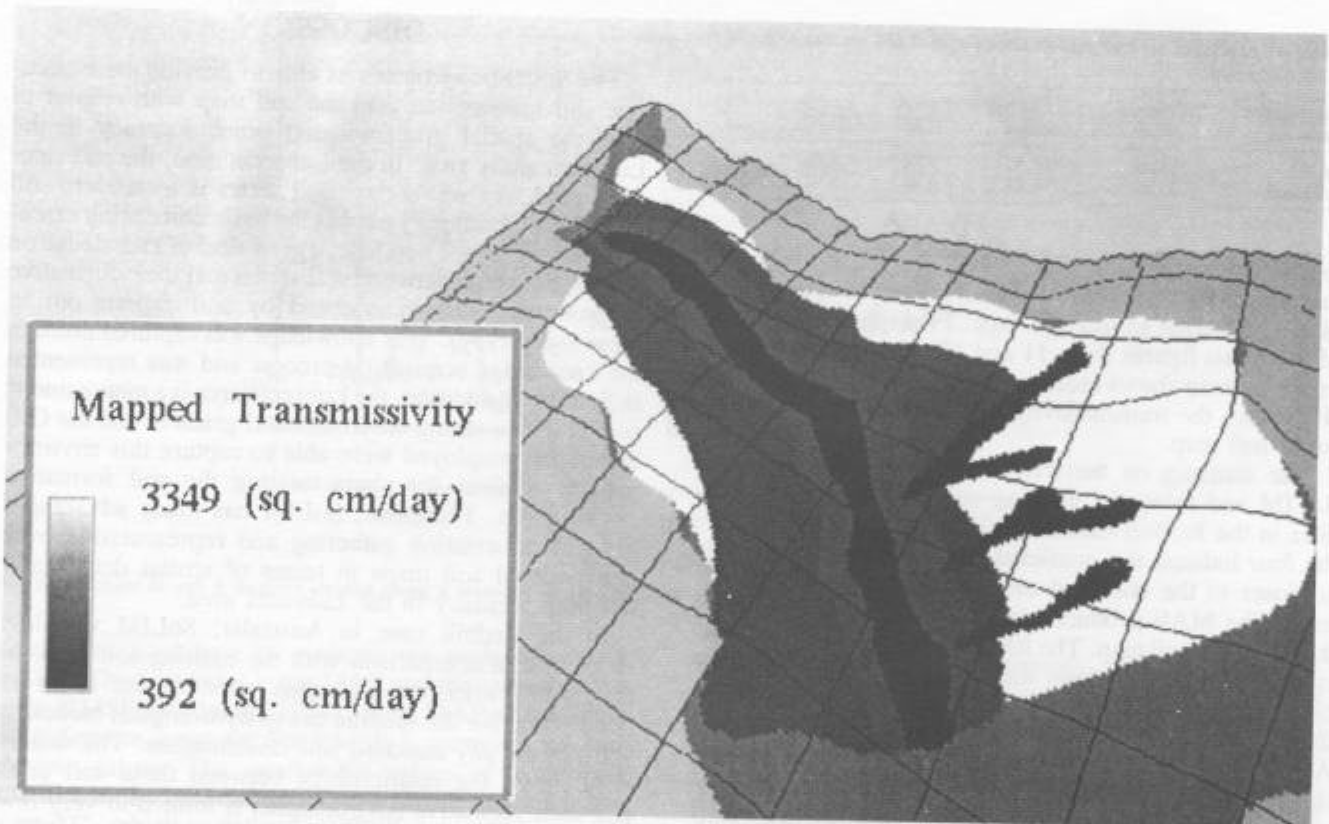


Fig. 12. Mapped transmissivity of the Redhill study area.

edge on the soil–environmental relationships was not utilized by SoLIM. In addition, the Redhill catchment is a small basin, about 173 ha, and has a very gentle environmental gradient. The GIS techniques were not able to provide enough details on the soil formative environment for SoLIM to utilize the knowledge extracted for soil inference. The soil property used in this exercise is the soil water transmissivity, which is highly variable across a very short distance and is difficult to map accurately. Under these circumstances, the SoLIM approach experienced difficulty in deriving an accurate continuous transmissivity map for the Redhill catchment. The result from SoLIM was only comparable to that from the soil map produced through many soil observations across a small area.

SUMMARY AND CONCLUSIONS

We derived soil property values across a region from the soil similarity vectors derived using SoLIM and demonstrated the usefulness of this fuzzy representation of soil information. From the above illustrations, this fuzzy representation of soil information does have some advantages over conventional representation of soil information under crisp logic (such as conventional soil survey maps). At the spatial level, the soil property map derived using the soil similarity vectors shows more gradation across space. This spatial gradation of soil property values provides a more realistic representation of soil information than the step-function type of variation, which is often adopted in conventional soil maps. At the attribute level, for the steep mountainous region, the derived soil property values are more accurate than the soil property values derived from conventional soil maps but not as advantageous for more gentle landscapes.

Although the fuzzy representation of soil information does have advantages, the accuracy of the final soil property values depends on the accuracy of the soil similarity vectors, which in turn depends on the process generating these vectors. There are two key components during the soil similarity vector generation process: the precision of knowledge on soil–environment relationships and the degree to which the environmental data employed can be used to characterize the soil formative environment. We showed that SoLIM worked well in an area where there is a strong environmental gradient and a rich understanding of soil–environment relationships, and where the environmental data used can effectively characterize the soil formative environment. However, SoLIM did not perform as successfully in an area where the environmental gradient is gentle and the relationships between the gentle environmental gradients and the soils were not well established.

Table 5. Statistics on the performance of SoLIM vs. the soil map in Redhill.

Quantitative measures	Observed mean	Predicted mean	MAE	RMSE	AC
	$cm^2\ d^{-1}$				
SoLIM	1406	1383	838	1173	0.30
Soil map	1406	1622	902	1112	0.66

Future development of SoLIM needs to be focused on two major areas: the representation of soil-forming factors, particularly geomorphic objects in the GIS database, and extraction of knowledge from multiple sources (such as human experts, existing soil maps, and research reports). At this stage, SoLIM only employs primitive topographic indices (such as slope gradient, slope aspect, curvature, and elevation). These indices can be used to predict the general distribution of soils in an area where environmental variation is very strong but would not be enough to capture the detailed functional effects of landform objects (such as mid-slope or floodplain) on soil formation processes. It is necessary to incorporate landform objects in SoLIM in order to provide more accurate soil information. SoLIM was only able to extract the knowledge on soil-environment relationships from a single expert. SoLIM would be easily applicable for other areas and the extracted knowledge would be more reliable if the knowledge were extracted from different human experts and/or from different sources (such as paper maps and survey reports). However, the knowledge from different sources needs to be properly integrated. The integration of knowledge from different sources would be a challenging research topic since all sources have inherent errors. Methodology needs to be developed so that these inherent errors would be removed during the knowledge integration (knowledge fusion) stage.

ACKNOWLEDGMENTS

Support from the University of Toronto, NSERC, NASA, CRC for Catchment Hydrology, and CSIRO is gratefully acknowledged. The GIS data on the Lubrecht study areas were mostly provided by the GIS Laboratory, School of Forestry, University of Montana. The start-up fund provided to the senior author by the Graduate School at the University of Wisconsin-Madison made it possible to complete this report.

REFERENCES

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

Band, L.E., D.L. Peterson, S.W. Running, J.C. Coughlan, R.B. Lammers, J. Dungan, and R. Nemani. 1991. Ecosystem processes at the watershed scale: Basis for distributed simulation. Ecol. Modell. 56:151–176.

Beven, K. 1986. Runoff production and flood frequency in catchments of order *n*: An alternative approach. p. 107–131. *In* V.K. Gupta et al. (ed.) Scale problems in hydrology. D. Reidel Publ. Co., Dordrecht, the Netherlands.

Brenner, R.L. 1968. The geology of Lubrecht Experimental Forest. Lubrecht Ser. 1. Montana For. Conserv. Exp. Stn., School of Forestry, Univ. of Montana, Missoula.

CSIRO Division of Water Resources. 1992. TOPOG user guide. Version 4.0. CSIRO, Canberra, ACT, Australia.

Dunne, T., and L.B. Leopold. 1978. Water in environmental planning. W.H. Freeman and Co., New York.

Freeman, T.G. 1991. Calculating catchment area with divergent flow based on a regular grid. Comput. Geosci. 17:413–422.

Griffith, D.A., and C.G. Amrhein. 1991. Statistical analysis for geographers. Prentice-Hall, Englewood Cliffs, NJ.

Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. McGraw-Hill, New York.

Jenny, H. 1980. The soil resource: Origin and behavior. Springer-Verlag, New York.

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

Mayer, D.G., and D.G. Butler. 1993. Statistical validation. Ecol. Modell. 68:21–32.

Missoula County Soil Survey. 1983. Soil series description. Missoula County Soil Survey, Missoula, MT.

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

Nemani, R., L. Pierce, S. Running, and L. Band. 1993. Forest ecosystem processes at the watershed scale: Sensitivity to remotely sensed leaf area index estimates. Int. J. Remote Sens. 14:2519–2534.

Nimlos, J.T. 1986. Soils of Lubrecht Experimental Forest. Misc. Publ. 44. Montana For. Conserv. Exp. Stn., Missoula.

Petersen, C. 1991. Precision GPS navigation for improving agricultural productivity. GPS World, 2(1):38–44.

Power, M. 1993. The predictive validation of ecological and environmental models. Ecol. Modell. 68:33–50.

Quinn, P., K. Beven, P. Chevallier, and O. Planchon. 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital elevation models. Hydrol. Process. 5:59–79.

Talsma, T., and P.M. Hallam. 1980. Hydraulic conductivity measurement of forest catchments. Aust. J. Soil Res. 18:139–148.

U.S. Geological Survey. 1990. Digital elevation models: Data users guide. USGS, Reston, VA.

Vertessy, R., 1990. Soils of the Redhill and Kylies Run catchments. Internal report. Cooperative Australian Center for Catchment Hydrology, CSIRO Division of Water Resources, Canberra, ACT, Australia.

Willmott, C.J. 1984. On the evaluation of model performance in physical geography. p. 443–460. *In* G.L. Gaile and C.J. Willmott (ed.) Spatial statistics and models. D. Reidel Publ. Co., Dordrecht, the Netherlands.

Willmott, C.J., S.G. Ackleson, R.E. Davis, J.J. Feddema, K.M. Klink, D.R. Legates, J. O'Donnell, and C.M. Rowe. 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. 90:8995–9005.

Zevenbergen, L.W., and C.R. Thorne. 1987. Quantitative analysis of land surface topology. Earth Surf. Processes Landforms 12:47–56.

Zhu, A.-X. 1994. Soil pattern inference using GIS under fuzzy logic. Ph.D. diss. Dep. of Geography, Univ. of Toronto, Toronto, ON, Canada.

Zhu, A.-X. 1996. A similarity model for representing soil spatial information. Geoderma (in press).

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

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