ELSEVIER

Geoderma 103 (2001) 181–201

GEODERMA

www.elsevier.com/locate/geoderma

Soil layer models created with profile cone penetrometer data

S. Grunwald *, K. McSweeney, D.J. Rooney, B. Lowery

Department of Soil Science, University of Wisconsin-Madison, 1525 Observatory Drive, Madison, WI 53706-1299, USA

Received 2 February 2000; received in revised form 30 January 2001; accepted 9 February 2001

Abstract

In creating soil layer models for our study site, we were challenged (i) to express vagueness of our soil data, while at the same time maintaining adherence to systematic classification principles, and (ii) to describe continuously the spatial distribution of soil materials and layers in three dimensions. We developed a method to create 3-dimensional (3-D) continuous soil layer models describing the distribution of soil materials, reworked loess vs. glacial till. Soil attribute data such as texture, bulk density and water content, in combination with penetration resistance obtained with a profile cone penetrometer on a 10-m grid, were used to describe soil materials and layers. We compared crisp hierarchical clustering with fuzzy k-mean classification in creating soil layer models for a 2.73-ha site in southern Wisconsin. The continuous 3-D soil layer models were developed using horizontal ordinary kriging and vertical linear interpolation. Validation proved that the crisp 3-D soil layer model predicted soil layers more accurately than the fuzzy 3-D soil layer model. We conclude that at the working scale, the crisp classification is superior to the fuzzy classification. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Clustering; Crisp classification; Fuzzy k-mean; Horizon; Landscapes; Penetrations; Three-dimensional models

1. Introduction

1.1. Crisp soil data

Conventionally, soil information systems contain polygons representing soil mapping units and associated representative soil attributes. Horizons of these

* Corresponding author. Current address: Soil and Water Science Department, University of Florida, 2171 McCarty Hall, P.O. Box 110290, Gainesville, FL 32611, USA. Tel.: +1-419-448-2089; fax: +1-419-448-2345.

E-mail address: sabgru@pocketmail.com (S. Grunwald).

PII: S0016-7061(01)00076-3

soil map units supposedly differ from adjacent and genetically related layers in physical, chemical and biological properties, e.g. texture, structure, color, soil organic matter and degree of acidity. As such, soil horizons and profiles of these properties correspond to discrete, sharply delineated (crisp) units, which are internally uniform. For example, Soil Taxonomy (Soil Survey Staff, 1998) defines an argillic horizon, as characterised by illuvial accumulation of phyllosilicate clays that has a certain minimum thickness depending on the thickness of the solum, a minimum quantity of clay in comparison with an overlying eluvial horizon depending on the clay content of the eluvial horizon, and usually has coatings of oriented clay on the surface of pores or peds or bridging sand grains. This definition is vague and ambiguous, and cannot be adequately modelled by crisp methods. Classification of soils into crisp classes has been reported by Oeborn (1989) and Simonson (1994).

The crisp or traditional method ignores important aspects of reality indicated by the gradual and continuous spatial changes of soil properties and terrain characteristics across the landscape. Considerable loss of information can occur when data that have been classified by this method are retrieved or combined using methods of simple Boolean algebra. Crisps allow only binary membership functions (i.e. true or false); an individual is a member or it is not a member of any given set as defined by exact limits. The advantage of a crisp method is that it is exploratory and may lead to testable hypotheses about the nature of soil and landscape (Burrough et al., 1992). However, vagueness in defining soil and terrain characteristics cannot be fully expressed. Crisp sets do not allow ambiguities and they are too inflexible to take account of genuine uncertainty (McBratney and de Grujiter, 1992).

Classification is an essential part of the data reduction process, whereby complex sets of observations are made understandable and transparent. Although classification involves a loss of information, it provides a convenient mean of information transfer. Optimal classification aims to reduce information, while identifying natural groups of individuals that have common properties. There are two distinct approaches to classify data. The first seeks a ‘natural partition’ of the observations in multivariate space into stable, mutually exclusive classes. Once identified, these crisp classes can be used as carriers of important information about other properties of the individuals in the group. For example, this approach is used in cluster and discriminant analysis (Webster and Oliver, 1990; Sinowski and Auerswald, 1999).

The second approach is to specify class limits on the basis of experience or imposed definitions, and then determine how many multivariate-defined individuals match the requirements (Burrough et al., 1992). For example, fuzzy methods allow the matching of individuals to be determined on a continuous scale instead on a Boolean binary or an integer scale. In contrast to crisp sets, a fuzzy set is a class that admits the possibility of partial membership. Fuzzy sets are generalizations of crisp sets. They are suitable for situations where the class membership or class boundaries are not, or cannot, be sharply defined. In short, fuzzy set theory provides an alternative approach, expressing the vagueness of soil properties over a landscape. Fuzzy c-mean clustering has been used to investigate soil–landscape relationships by De Bruin and Stein (1998), Odeh et al. (1992) and Irwin et al. (1997). A survey of fuzzy clustering is given by Yang (1993) and the general applications of fuzzy sets in soil science are given by McBratney and Odeh (1997). Continuous soil classification with fuzzy c-mean algorithm has been reported by McBratney and de Grujiter (1992) and Mazaheri et al. (1995). Burrough et al. (1992) and Lagacherie et al. (1997) compared crisp and fuzzy clustering on sets of soil survey data.

As both topography and soil are continuous phenomena, geostatistical methods have been used for the description of continuous variation of soil attributes in geographical space since the early 1980s. Applied at first to single soil attributes, the geostatistical approach has been extended to deal with multivariate indices of soil variation (Webster and Oliver, 1990), to qualitative variables (Bierkens and Burrough, 1993), and to model the multiple-nested scales of spatial variation (Goovaerts and Webster, 1994). Goovaerts (1997) presents the common kriging techniques, such as simple kriging, ordinary kriging, kriging with a trend model, block and factorial kriging. Cokriging (Isaaks and Srivastava, 1989; Goovaerts, 1997) and other kriging methods (Burrough et al., 1997; McBratney et al., 1991) that account for secondary information are useful when the primary data are sparse or poorly correlated in space.

Many of the phenomena, such as soil, investigated in earth sciences cannot be adequately described using two-dimensional (2-D) models. Examples include geological structures, mineral deposits, soil patterns, and hydrological processes. These phenomena are three-dimensional (3-D), and if changes over time are considered, abstractions have to include the fourth dimension (time). Recent developments in spatial modelling have focused on 3-D representations of data (Goderya et al., 1996; Garcia and Froidevaux, 1997). The relatively few 3-D representations of the soil currently available are striking (Vitek et al., 1996). At microscale level, Heijs et al. (1995) used computed tomography to create a 3-D model of water content and macropore structure. More recently Pereira and FitzPatrick (1998) presented a 3-D model of tubular horizons in sandy soils and Barak and Nater (1999) developed a gallery of 3-D soil minerals and molecules. At landscape scale Grunwald et al. (2000) presented 3-D models showing the spatial distribution of cone indices and soil layers. At the Cooperative Research Centre in Australia, research on mapping of regolith (and associated attributes) is in transition from 2-D to 3-D (Cooperative Research Centre for Landscape Evolution and Mineral Exploration, 1999). In order to account for the continuous nature of soil variation, Ameskamp (1997) developed a 3-D rule-based approach, which uses fuzzy rules defined by soil experts to represent relationships between the soil and the landscape. Ameskamp visualised soil horizons, as independent 3-D volumes.

Scientific visualization (SciVis) is one of the most powerful communicators of spatial information (Barraclough and Guymer, 1998) in 3-D. Research by Stibbard (1997) indicates that information is absorbed best when using more than one human sense, i.e. 10% is taken by reading, 30% by reading and visual, 50% reading, visuals sound, and 80% by reading, visuals, sound and interaction. Barraclough and Guymer (1998) indicate that advanced visualization techniques served to better communicate spatial information between people of different backgrounds, ranging from scientists, administrators, educators and the public. Just as maps can visually enhance the spatial and temporal understanding of phenomena, 3-D representations can enhance our understanding of soil patterns.

In creating soil layer models for our study site, we were challenged (i) to express vagueness of our soil data, while at the same time maintaining adherence to systematic classification principles and (ii) to describe continuously the spatial distribution of soil materials and layers in three dimensions. We compared crisp and fuzzy classification methods to create soil layer models for a 2.73-ha site in southern Wisconsin. Soil attribute data, such as texture, bulk density and water content in combination with penetration resistance, were used to describe soil materials and layers. Continuous extent of soil layers in three-dimensions was modelled using the spatial interpolation technique of ordinary kriging and 3-D surface rendering.

2. Material and methods

2.1. Soil layers and materials in southern Wisconsin

In relatively young glaciated landscapes of the southern Wisconsin (Wisconsin-age, $<15{,}000$ years BP), characterized by restricted and immature drainage networks, soil landscape variability imparts a major control on productivity and environmental quality associated with production agriculture. Apart from the influence of topography, climate and biological activity, soil material is a dominant factor influencing land use and management and, subsequently, their effects on soil and water quality. Two common soil materials found in glaciated landscapes are glacial till and reworked loess. Glacial till is unsorted and unstratified material, deposited by glacial ice, which consists of a mixture of clay, silt, sand, gravel and stones in any proportion. In contrast, loess—transported and deposited by wind—consists of predominantly silt-sized particles. Over the ages, the loess deposits have been subjected to a variety of processes such as erosion, deposition and decalcification. Thus, it is denoted as ‘reworked loess’. Glacial till and reworked loess differ in terms of drainage behaviour, water-holding capacity, compaction, adsorptivity and leaching of nutrients and agrichemicals. Currently, our knowledge about the three-dimensional spatial distribution of these soil materials at fine and coarse scale is limited.

2.2. The profile cone penetrometer system

We used a constant rate PCP with 60°-cone angle and 2-cm diameter surface area to measure soil profile cone penetrations. The PCP components include the PCP probe, a hydraulic truck-mounted push system (Gidding #9HD; Fort Collins, CO[1]), a load cell (1360-kg capacity; Omegadyne LC 101; Sunbury, OH), a depth transducer (Unimeasure HX-EP; Corvallis, OR), and a data acquisition system (data logger, Campbell Scientific 21X; Logan, UT). Profile cone measurements were very fast, with an average of 1.30 min/profile when compared to soil core sampling with about 30 min/profile.

2.3. Data set used to create the soil layer models

The continuous cone indices (CI), measured with the PCP system, were obtained at 273 locations at depths down to 1.30 m on a 10×10-m grid. A total area of 2.73-ha was covered. A Trimble 4600 LS differential global positioning system (GPS), single frequency, dual port, with an internal 4600 LS antenna (Trimble, 1996; Sunnyvale, CA) was used to georeference sampling locations. The horizontal resolution error for each mapped coordinate was ± 4 cm and the vertical resolution error was ± 8 cm. The GPS unit was also used to conduct a kinematic survey to derive a DEM for the study area. A grid size of 1 m was chosen for interpolation of point elevations. We derived primary topographic attributes, slope and upslope drainage area, using the ArcView–Spatial Analyst Geographic Information System (ESRI™, Redlands, CA).

Soil cores were collected at 21 penetration locations, somehow simultaneously (1–2 h) with penetration measurements, and samples from the horizons were analyzed for texture, bulk density and water content. Soil texture was analyzed by the University of Wisconsin Soil and Plant Analyses Laboratory, based on the hydrometer method. Soil water content (θ) and bulk density (ρ_b) was derived by collecting known volumes of samples and oven-drying them to obtain volumetric values.

2.4. Crisp cluster analysis

Supposing a survey of n sites lists values of p soil variables. The list constitutes a data matrix, $\mathbf{X}$, of size $n \times p$ and the hard partitioning of the data into k classes would produce an $n \times k$ matrix, $\mathbf{M}$, of membership values. Given that $\mathbf{M} = (m_{ic})$, where $m_{ic} = 1$ if individual i belongs to class c and $m = 0$

[1] Mention of company or trade name does not constitute endorsement by the University of Wisconsin-Madison or the authors.

otherwise. To ensure that the classes are mutually exclusive, jointly exhaustive and non-empty, the following conditions on **M** apply:

$$\sum_{c=1}^{k} m_{ic} = 1, \quad i = 1,\dots,n \tag{1}$$

$$\sum_{i=1}^{n} m_{ic} > 0, \quad c = 1,\dots,k \tag{2}$$

$$m_{ic} \in \{0,1\}, \quad i = 1,\dots,n; c = 1,\dots,k. \tag{3}$$

Since Eq. (3) corresponds to the all-or-nothing status of the memberships, this method is termed as 'crisp' classification.

We used the input variables: (i) CI/depth, i.e. CI was averaged over 3-cm depth increments along the profile for each penetration, (ii) elevation, (iii) slope and (iv) upslope drainage area in the hierarchical cluster analysis. We assumed that these topographic attributes influence the distribution of soil patterns in southern Wisconsin based on studies by Irwin et al. (1997) and Slater (1994). We used a hierarchical cluster analysis to form groups of similar penetrations and topographic attributes. Agglomerative, hierarchical clustering starts with crisp, discontinuous classes, whereby each individual belongs exactly to only one class. Variables were standardized to weight each variable equally in the cluster analysis. Key measures in cluster analysis are 'distance' and 'similarity'. Distance is a measure of how far apart two individuals are, and similarity measures closeness. Distance and similarity measures are small for cases that are similar. In cluster analysis, these concepts are especially important, since individuals are grouped on the basis of their 'nearness' in multivariate space. We used 'average linkage within groups' to combine clusters (SPSS Professional Statistics, 1994). This method combines clusters so that the average distance between all cases in the resulting cluster is as small as possible. Thus, the distance between two clusters is taken to be the average of the distances between all possible pairs of cases in the resulting cluster. We used the Pearson Correlation Coefficient to test similarity between variables. In agglomerative, hierarchical clustering, clusters are formed by grouping cases into bigger and bigger clusters until all individuals are members of a single cluster. The agglomeration schedule and dendrogram can be used to follow up the grouping process of all individuals at different stages. However, there are no objective guidelines to determine the optimal number of classes. We used the following criteria to make a decision about the number of classes: (i) each class was supposed to have a minimum amount of individuals, (ii) small coefficients of similarity indicated that fairly homogeneous clusters are being merged, whereas large coefficients of similarity indicated that clusters containing quite dissimilar members are being combined and (iii) resulting clusters should represent different physical characteristics. We used a two-tailed *t*-test to test whether the means of variables CI, clay, silt and sand content, ρ_b and θ were significantly different at the 95% confidence level between layers within each cluster and between clusters. To distinguish layers and clusters, at least one variable had to be significantly different.

2.5. Fuzzy clustering

We used the fuzzy c-mean clustering method (Bezdek and Pal, 1992), also known as fuzzy k-mean (McBratney and de Grujiter, 1992), for classifying our penetration and topographic data into fuzzy groups. Our analysis was conducted with the FuzMe software package (Minasny and McBratney, 1999), based on the algorithms of McBratney and de Grujiter (1992). Fuzzy k-mean clustering is a centroidal grouping method to create continuous classes and it accounts for vagueness of the data. We used the same input variables as in the case of crisp classification. Before running the analysis, we standardized the variables to account for equal weighting of the variables for fuzzy clustering.

Continuous or fuzzy clustering is simply a generalization of discontinuous classes, where the indicator function of conventional set theory, with value 0 or 1, is replaced by the membership function of fuzzy set theory (Zaheh, 1965). The theory relaxes Eq. (3) so that memberships are allowed to be partial, i.e. to take any value between and including 0 and 1. Thus, Eq. (3) is replaced by Eq. (4).

$$m_{ic} \in [0,1] \quad i = 1,\ldots,n; c = 1,\ldots,k. \tag{4}$$

Now any $n \times k$ matrix, $\mathbf{M}$, that satisfies the conditions as expressed in (Eqs. (1), (2) and (4), represents the so-called fuzzy partition of the n individuals into k fuzzy classes (Zaheh, 1965). As Eq. (3) partially agrees with Eq. (4), crisp partitions are a special case of fuzzy partitions. The key algorithm used in fuzzy k-mean clustering is termed the fuzzy k-mean objective function (FKM) defining $J(\mathbf{M},\mathbf{C})$ (Eq. 5). The $J(\mathbf{M},\mathbf{C})$ is the sum-of-square errors (expressed as distances) due to the representation of each individual by the center of its class. Fuzzy k-mean minimizes the within-class sum-of-square errors function $J(\mathbf{M},\mathbf{C})$ under condition (1), (2) and (3):

$$J(\mathbf{M},\mathbf{C}) = \sum_{i=1}^{n}\sum_{c=1}^{k} m_{ic}^{\varphi} d_{ic}^{2}(x_i,c_c), \tag{5}$$

where $\mathbf{C} = c_{cv}$ is a $k \times p$ matrix of class centres, c_{cv} denotes the value of the centre of class c for variable v, $x_i = (x_{i1},\ldots,x_{ip})^T$ is the vector representing individual i, $c_c = (c_{c1},\ldots,c_{cp})^T$ is the vector representing the center of class c, and $d_{ic}^2(x_i,c_c)$ is the square distance between x_i and c_c according to a chosen definition of distance, further denoted by d_{ic}^2 for simplicity, and φ is a coefficient which describes the degree of fuzziness of the solution. With the smallest meaningful value $\varphi = 1$, the solution of Eq. (5) is a crisp partition, i.e. the result is not fuzzy at all. As φ approaches infinity, the solution approaches its maximum degree of fuzziness, with $m_{ic} = 1/k$ for every pair of i and c (McBratney and de Grujiter, 1992). Since our variables are highly correlated with each other (e.g. CI/depth–elevation), we used Mahalanobis' metric to optimize the performance of the objective functions $J(\mathbf{M},\mathbf{C})$ since the metric allows for differences in variance and correlations among variables (Odeh et al., 1992).

Generally, fuzzy k-mean clustering produces many realizations while grouping each individual into classes and calculating membership values. Fuzzy realizations are influenced by (i) the value of the fuzzy exponent φ and (ii) the number of groups chosen. The degree of fuzziness is somewhat related to substructures in the **X** matrix, and hence the optimal number of classes. Ward et al. (1992) suggest that when using the Mahalanobis' metric, a fuzzy exponent φ close to 1 should be used. Odeh et al. (1990) suggest for soil data to set φ between 1.12 and 1.50, and McBratney and de Grujiter, (1992) suggest $\varphi = 2$. However, no objective method for determining the optimal degree of fuzziness exists. We used a sensitivity analysis varying φ between 1.0 and 1.9 and varying the number of classes between 2 and 8. We evaluated the fuzzy clustering using the fuzziness performance index (F') and the partition entropy (H') suggested by Roubens (1982), McBratney and Moore (1985) and Odeh et al. (1992):

$$F' = \frac{1 - (cF - 1)}{F - 1}, \tag{6}$$

where $F = F(\mathbf{M},\mathbf{C}) = \sum_{i=1}^{n}\sum_{c=1}^{k} m_{ic}\varphi^2/n$.

$$H' = \frac{H}{\log_\beta c} \tag{7}$$

where $H = -\sum_{i=1}^{n}\sum_{c=1}^{k} m_{ic}\log_\beta(m_{ic})/n \; 0 < \beta < \infty$.

F' describes the membership sharing between any pair of fuzzy classes. $F' = 1$ corresponds to maximum fuzziness and $F' = 0$ means non-fuzziness. H' describes the certainty (or uncertainty) of fuzzy k-partition of events of sample space **X**. Minimization of H' is consistent with maximizing the amount of information about the substructure in **X** that is generated by the FKM algorithm. The least fuzzy and least disorganized number of classes is considered optimal. Indices, F' and H', have the tendency of monotonicity and, therefore, lack appropriate thresholds against which their values can be judged as 'good'. Thus, a subjective decision is necessarily guided not only by the performance of the indices, but also by examination of membership values, where classification should be not 'too fuzzy' and not 'too crisp'.

2.6. Continuous soil layer models

2.6.1. Clusters identified by the hierarchical cluster analysis

The thickness of layers was determined by vertical point inflection method, which is described in detail by Grunwald et al. (2001). We calculated mean and standard deviation (SD) for soil properties, sand, silt and clay content, bulk density and water content for each layer of each cluster. We used these soil properties to describe soil material of each layer. On the basis of expert knowledge, we defined materials as 'reworked loess' if (i) sand content $< 53\%$, silt content $\geq 40\%$, (ii) bulk density ≥ 1.30 and < 1.6 Mg m^{-3} and (iii) water content ≥ 0.17 m^3 m^{-3}. We classified a material as 'glacial till' if (i) sand content $\geq 53\%$, silt content $\leq 50\%$, (ii) bulk density ≥ 1.60 Mg m^{-3} and (iii) water content < 0.17 m^3 m^{-3}. Typically, the top layer of soils is highly influenced by tillage operations resulting in changes in bulk density, water content, and structure.

We used soil layer and material information along with elevation data and 273 CI profiles to create a 3-D continuous soil layer model for the study area. We then used Environmental Visualization Software, EVS (EVS-NT Standard Version CTech, Huntington Beach, CA) for spatial modelling. Surfaces of each layer were created using 2-D ordinary kriging, and volumes (layers) were created with linear interpolation in the vertical direction between these surfaces. A vector geographic data model was used to describe the layer boundaries.

2.6.2. Classes identified by the fuzzy k-mean cluster analysis

We used the same procedure as described in Section 2.6.1, i.e. for each class, we defined layers and materials based on our soil core analysis. Centroids for topographic attributes and cone indices were output by the FuzMe program. Our goal was to create a soil layer model; thus, we had to defuzzify our data. We used a defuzzification method, suggested by Bandemer and Gottwald (1995), using weighted average to calculate layer depth at each penetration location (wlay_n), according to the following algorithm:

$$\mathrm{wlay}_n = \sum_{c=1}^{k} m_{nc}\, p_{nc}, \tag{9}$$

where m_{nc} are membership values for each site n, and class c and p denotes the layer thickness, for example, depth of reworked loess. At each of the 273 penetration locations, weighted layer depths for each layer were calculated and along with material information, we created a soil layer model for the study site.

2.7. Validation procedure

We validated the crisp and the fuzzy methods in predicting the soil layer depths using a different set of penetration measurements, at 77 locations along transects at varying spacing between them. The measurements were done on six transects running from the upper slopes to the lower slopes in the study area. We evaluated the accuracy of the model predictions using the Root Mean Square Error (RMSE). The RMSE is a measure of accuracy and precision, and is defined as the mean square root of the sum of squared prediction error:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \hat{y}_i)^2}{n}}, \tag{10}$$

where x_i are measured values and $\hat{y}_i$ are predicted values. The RMSE emphasizes the deviation between measured and predicted values.

We also used the Coefficient of Efficiency, E (Nash and Sutcliffe, 1970), to evaluate the fit of measured vs. predicted data. E is expressed as:

$$E = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 - \sum_{i=1}^{n}(\hat{y}_i - x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \, i = 1,2,\ldots,n, \tag{11}$$

where $\bar{x}$ is the arithmetic mean of x_i for all events $i = 1$–n.

The estimated E essentially is the sum of the deviations of the observations from a linear regression line with slope of 1. If the measured variable is predicted with high precision for all observations, then E approaches 1.

3. Results

3.1. Results of crisp hierarchical cluster analysis

We classified the penetration and topographic data using the hierarchical cluster analysis. We determined the optimal number of classes using the agglomeration schedule and dendrogram analysis. The results are illustrated in Fig. 1. The figure shows the 273 data records, each record comprising of penetration and topographic data, the Pearson Correlation Coefficients (r) and number of classes at different stages of the grouping process. While grouping the data records, r decreased exponentially. Large r indicated grouping of relatively similar records in terms of CI/depth, elevation, slope and upslope drainage area. We used the following criteria to make a decision about the optimal number of classes (i) $r > 0.8$ and (ii) $r > 14$ data records in each cluster, which equals $> 5\%$ individuals. Based on these criteria, we identified four different clusters (denoted by A, B, C and D), where clusters showed small variability within clusters and large variability between clusters in terms of

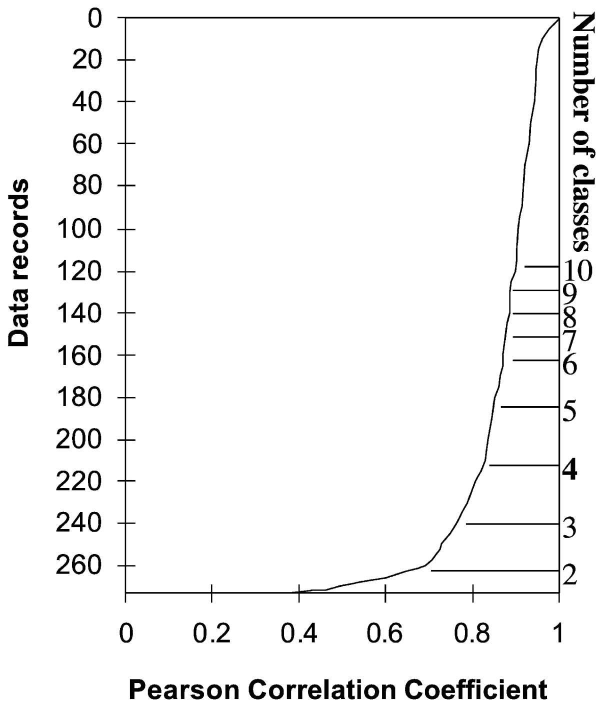

Fig. 1. Similarity between data records (penetration and topographic data) evaluated with the Pearson Correlation Coefficient and number of classes distinguished by the hierarchical cluster analysis.

CI/depth, elevation, slope and upslope drainage area. The mean and SD for the topographic attributes are presented in Table 1, and the mean and SD for the soil attributes and CI are listed in Table 2. SD is indicative of the variability of a variable. Clusters A and B were found on higher elevated areas and on the

Table 1
Mean and standard deviation of topographic attributes for each cluster identified by the hierarchical cluster analysis

Topographic attributes	Cluster A ($n=34$)	Cluster B ($n=100$)	Cluster C ($n=102$)	Cluster D ($n=37$)
Elevation (m)	**327.0**[a]; 1.1[b]; a–c[c]	**326.9**; 0.9;	**323.1**; 0.8; c–d	**325.9**; 0.6;
Slope (°)	**3.3**; 1.5; a–c	**3.0**; 1.2;	**1.1**; 0.5;	**2.4**; 0.8;
Upslope drainage area (m^2)	**1565**; 234; a–b, a–c	**6500**; 133; b–c	**15,450**;199; c–d	**4100**; 82;

[a]**Mean**.
[b]SD.
[c]Letters indicate significant differences between clusters of the same variable.

Table 2
Mean and standard deviation for cone index, clay, silt and sand content, bulk density and water content for layers identified by the hierarchical cluster analysis

Layer	Attribute	Cluster A[e]	Cluster B	Cluster C	Cluster D
1	CI (kPa)	**1,335.3**[a]; 482.8[b]; 1-2[c], 1-3	**1,207.4**; 440.4; 1-2, 1-3	**1,219.1**; 450.7; 1-2, 1-3	**1,441.5**; 491.1; 1-2
	Clay (%)	**13.8**; 1.2; 1-3, a-b[d], a-d	**21.2**; 1.2; 1-3, b-c	**13.8**; 1.4; 1-2, 1-3, c-d	**21.5**; 1.9;
	Silt (%)	**63.0**; 2.0; 1-3	**56.1**; 2.1;	**60.8**; 3.2; 1-3	**60.2**; 1.5; 1-2
	Sand (%)	**23.1**; 0.9; 1-2, 1-3	**21.1**; 1.5; 1-3	**25.4**; 1.9; 1-2	**21.1**; 1.4; 1-2
	ρb ($Mg\ m^{-3}$)	**1.30**; 0.2; 1-2; 1-3, a-b, a-d	**1.45**; 0.2; 1-3, b-c	**1.31**; 0.25; 1-3, c-d	**1.49**; 0.25; 1-2
	θ ($m^3\ m^{-3}$)	**0.191**; 0.015; 1-3	**0.186**; 0.012; 1-3	**0.192**; 0.02;	**0.186**; 0.022;
2	CI (kPa)	**1,825.6**; 502.1; 2-3, a-b, a-c	**1,514.5**; 478.8; 2-3, b-d,	**1,615.2**; 498.2; 2-3, c-d	**1,913.2**; 410.2;
	Clay %	**13.5**; 1.5; a-b, a-c, a-d	**21.5**; 1.8; 2-3	**24.6**; 1.9; 2-3	**26.0**; 1.1;
	Silt %	**63.2**; 2.1; 2-3	**63.0**; 2.1; 2-3	**63.5**; 3.7; 2-3	**67.2**; 2.2;
	Sand %	**36.4**; 4.5; 2-3, a-b, a-c, a-d	**21.2**; 1.1; 2-3, b-c, b-d	**12.3**; 2.1; 2-3	**12.1**; 2.1;
	ρb ($Mg\ m^{-3}$)	**1.38**; 0.25; 2-3; a-b	**1.51**; 0.2; 2-3, b-c, b-d	**1.38**; 0.2; 2-3	**1.31**; 0.23;
	θ ($m^3\ m^{-3}$)	**0.198**; 0.014; 2-3	**0.200**; 0.011; 2-3	**0.191**; 0.017;	**0.197**; 0.023;
3	CI (kPa)	**4,020.0**; 2,927.0; a-b; a-c	**2,759.8**; 1,225.1;	**2,866.0**; 1,390.2;	-
	Clay %	**9.0**; 1.6; a-c	**10.4**; 1.5; b-c	**34.1**; 2.1;	
	Silt %	**9.1**; 1.8; a-c	**11.0**; 1.9; b-c	**50.2**; 2.2;	
	Sand %	**80.3**; 4.6; a-c	**78.5**; 3.6; b-c	**25.4**; 27;	
	ρb ($Mg\ m^{-3}$)	**1.71**; 0.1; a-c	**1.65**; 0.2; b-c	**1.45**; 0.2;	
	θ ($m^3\ m^{-3}$)	**0.125**; 0.015; a-c	**0.123**; 0.009; b-c	**0.212**; 0.011;	

[a] Mean.
[b] SD.
[c] Figures indicate significant differences between layers of the same variable.
[d] Letters indicate significant differences between clusters of the same variable.

steeper slopes with mean elevation of 327.0 and 326.9 m and mean slope of 3.3° and 3.0°. In contrast, cluster C were found on the low elevated areas (mean elevation = 323.1 m) and low slope (mean: 1.1°). Cluster D occurred on the medium elevated areas (mean elevation = 325.9 m) and medium slope (mean: 2.4°). While cluster C is predominantly characterised by large upslope drainage area (15,450 m^2), cluster A showed small values for upslope drainage area (1565 m^2). Clusters B and D occurred on medium-sized upslope areas, with a mean of 6500 and 4100 m^2, respectively.

Table 3
Thickness of layers in centimeters for clusters identified by the hierarchical cluster analysis

Layer	Cluster A ($n=34$)	Cluster B ($n=100$)	Cluster C ($n=102$)	Cluster D ($n=37$)
1	0–25	0–25	0–25	0–40
2	25–85	25–80	25–80	40–130
3	85–130	80–130	80–130	–

We identified three different layers, with layer 1 corresponding to the topsoil layer typically formed in reworked loess material, but largely influenced by tillage operations. Layer 2 was formed in reworked loess and layer 3 in glacial till material. The different soil materials are shown in Table 2, where the table columns with light gray background indicate reworked loess material, and the columns with dark gray background indicate glacial till. We found glacial till layers in clusters A and B, whereas all other layers were formed in reworked loess material. Layers 1 and 2 were distinguished based on soil texture (clusters A, C and D) and bulk density (cluster B). Soil water content was similar in layers 1 and 2 for all clusters. However, layer 3 showed low water content, 0.125 m^3 m^{-3} for cluster A and 0.123 m^3 m^{-3} for B. We distinguished only two layers in cluster D. The thickness of layers is shown in Table 3. The crisp soil layer model is visualized in Fig. 4. The model resulting from the crisp analysis showed a shallow layer of reworked loess on backslope positions, and an overthickened layer of reworked loess on lower elevated areas (Fig. 4). We defined the latter areas as colluvium. Glacial till occurred close to the soil surface on higher elevated areas, and no glacial till was found on lower elevated areas at a depth of 1.30 m.

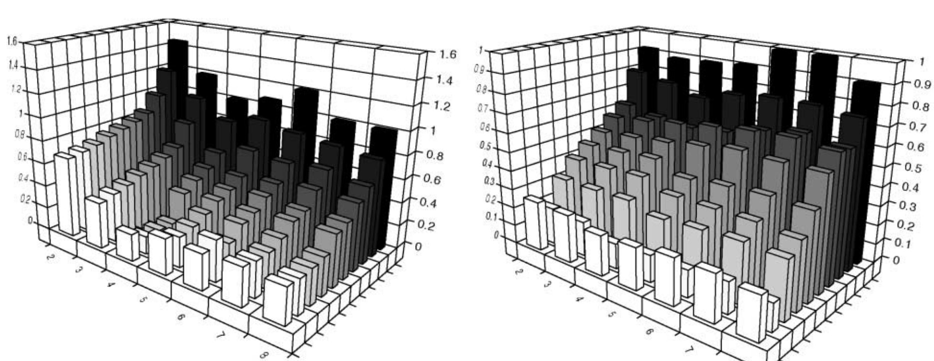

Fig. 2. Fuzziness performance index (F') and partition entropy (H') for different classifications, and fuzzy exponents (φ) using fuzzy k-mean classification. Left, x-axis: number of classes; y-axis: fuzzy exponent; z-axis: F' (fuzziness performance index). Right, x-axis: number of classes; y-axis: fuzzy exponent; z-axis: H' (partition entropy).

3.2. Results of fuzzy k-mean classification

We applied the fuzzy k-mean algorithm as expressed in Eq. (5) to our data set with 273 records of CI/depth, elevation, slope and upslope drainage area using fuzzy exponents in the range 1.0–1.9 and number of classes between 2 and 8. F' and H' for each of the different factor combinations are shown in Fig. 2. The optimal value of F' and H' for a given number of classes and φ is at the minimum of the blocks in the 3-D bar diagram (Fig. 2). We calculated the minimum of F' for 4 classes or clusters using φ of 1.3 with $F' = 0.150$. The combination of 4 classes and $\varphi = 1.2$ yielded $F' = 0.151$. The minimum of H' for 4 classes and $\varphi = 1.2$ was $H' = 0.212$. The minima for F' and H' coincide when the number of classes is 4 and φ is 1.2. We therefore used these specifications for further analysis. Class centroids of CI/depth for fuzzy k-mean classification are shown in Fig. 3. Mean and SD for topographic attributes of the 4 classes, denoted by classes A, B, C and D, are shown in Table 4. Class A occurred on high-elevated areas with a mean elevation of 328.2 m, class C on

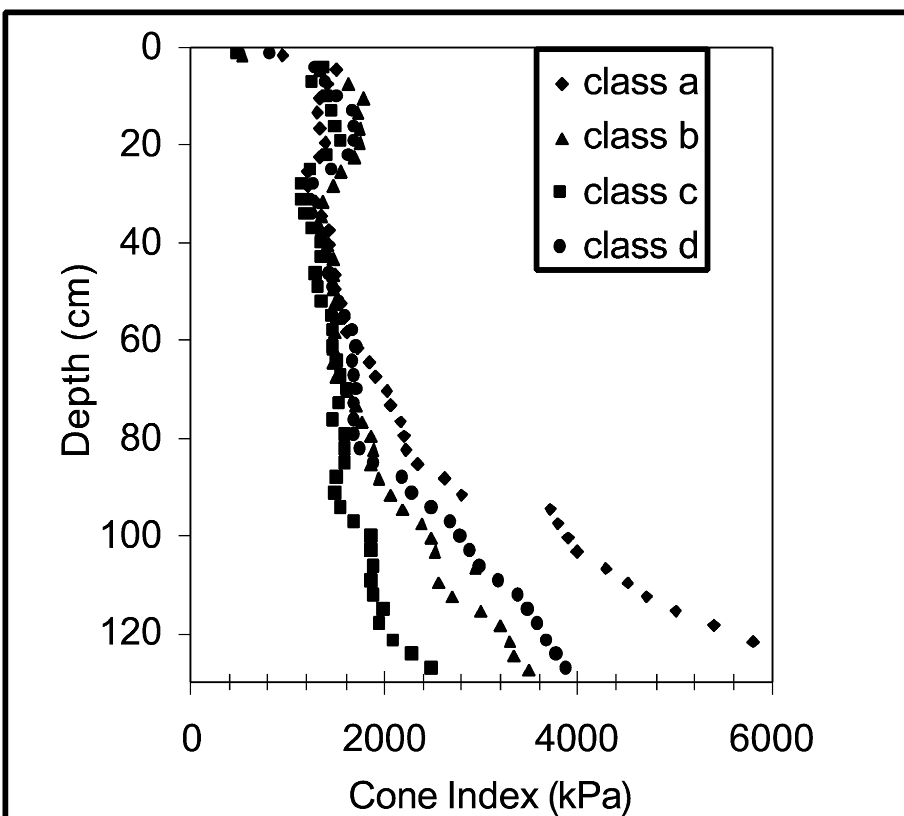

Fig. 3. Class centroids of CI/depth for fuzzy k-mean classification.

Table 4
Class centroids for each class identified by the fuzzy c-mean analysis

Topographic attributes	Class A ($n=51$)	Class B ($n=99$)	Class C ($n=77$)	Class D ($n=46$)
Elevation (m)	328.2; a–c	325.3; b–c	321.1; c–d	326.8
Slope (°)	4.3; a–c	2.6; b–d	0.9; c–d	4.8
Upslope drainage area (m^2)	1300; a–b, a–c	8568; b–c	20,511; c–d	3544

low-elevated areas with mean elevation of 321.1 m, and classes B and D on medium-elevated areas with means 325.3 and 326.8 m, respectively. Mean slope

Table 5
Mean and standard deviation for cone index, clay, silt and sand content, bulk density (ρ_b) and water content (θ) for layers identified by the fuzzy k-mean analysis

Layer	Attribute	Class A[e]	Class B	Class C	Class D
1	CI (kPa)	**1,221.3**[a]; 450.8[b]; 1-2[c], 1-3	**1,336.5**; 478.3; 1-2, 1-3	**1,119.1**; 411.7; 1-2, 1-3	**1,291.5**; 501.0; 1-2
	Clay %	**15.1**; 1.5; a-b[d]	**23.5**; 2.0; 1-3, b-c	**12.4**; 2.4; 1-2, 2-3	**15.5**; 2.1; 1-2
	Silt %	**65.1**; 2.2; 1-3	**59.1**; 1.9; 1-3	**62.5**; 2.2;	**63.2**; 1.7;
	Sand %	**18.2**; 1.3; 1-2	**20.3**; 1.9; 1-3	**22.4**; 2.1; 1-2	**19.1**; 1.9; 1-2
	ρb (Mg m^{-3})	**1.30**; 0.2; 1-2, 1-3, a-b, a-d	**1.52**; 0.2; b-d	**1.30**; 0.21; 1-3, c-d	**1.42**; 0.2; 1-2
	θ (m^3 m^{-3})	**0.201**; 0.02; 1-3	**0.181**; 0.019; 1-3	**0.191**; 0.021;	**0.201**; 0.019;
2	CI (kPa)	**1,899.3**; 499.8; 2-3, a-c	**1,714.4**; 505.8; 2-3	**1,499.1**; 500.2; 2-3, c-d	**1,988.2**; 422.2;
	Clay %	**14.2**; 2.2; a-c, a-d	**21.4**; 2.1; 2-3, b-d	**22.5**; 2.0; 2-3	**30.0**; 2.1;
	Silt %	**61.8**; 2.4; 2-3	**63.2**; 2.1; 1-3	**64.5**; 2.4; 2-3	**67.2**; 1.9;
	Sand %	**33.2**; 3.3; 2-3, a-b, a-c, a-d	**18.4**; 1.9; 2-3, b-d	**13.1**; 2.1;	**9.4**; 1.9;
	ρb (Mg m^{-3})	**1.41**; 0.2; 2-3	**1.45**; 0.2; 2-3, b-d	**1.38**; 0.15;	**1.32**; 0.2;
	θ (m^3 m^{-3})	**0.178**; 0.021; a-c	**0.193**; 0.017; 2-3	**0.201**; 0.021;	**0.200**; 0.020;
3	CI (kPa)	**3890.2**; 2,209.2; a-c	**3,108.1**; 1,424.2; b-c	**2,500.3**; 1,795.1;	-
	Clay %	**11.4**; 2.0; a-c	**14.4**; 2.0;	**31.2**; 2.3;	
	Silt %	**10.1**; 1.8; a-c	**15.8**; 2.0; b-c	**55.1**; 2.0;	
	Sand %	**75.2**; 3.5; a-c	**68.2**; 3.9; b-c	**21.1**; 2.7;	
	ρb (Mg m^{-3})	**1.69**; 0.15; a-c	**1.61**; 0.12; b-c	**1.41**; 0.2;	
	θ (m^3 m^{-3})	**0.122**; 0.018; a-c	**0.153**; 0.017; b-c	**0.213**; 0.021;	

Table 6
Thickness of layers in centimeters for classes identified by the fuzzy c-mean analysis

Layer	Class A ($n=51$)	Class B ($n=99$)	Class C ($n=77$)	Class D ($n=46$)
1	0–25	0–25	0–30	0–25
2	25–85	25–95	30–130	25–90
3	85–130	80–130		90–130

was quite different for classes, ranging from 4.3° for class A, 2.6° for class B, 0.9° for class C and 4.8° for class D. Upslope drainage area ranged from 1300 m^2 for class A, 8568 m^2 for class B, 20,511 m^2 for class C and 3544 m^2 for class D.

Classes A and B showed glacial till material in layer 3, where large mean sand contents were found with 75.2% and 68.2%, respectively (Table 5). Bulk density in layer 3 of these classes was large, with 1.69 and 1.61 Mg m^{-3}. Water content was very small in layer 3 of classes A and B, with 0.122 and 0.153 m^3 m^{-3}. Layers 1 and 2 of classes A and B were distinguished by CI and bulk density values. Class C showed three different layers, which differed in mean clay and sand content and CI. For example, CI increased from 1119.1 kPa in layer 1 to 1499.1 kPa in layer 2, and to 2500.3 kPa in layer 3 of class C. Class D showed only two distinct layers, which differed only in CI, mean clay content and bulk density. The thickness of layers used to create the 3-D continuous

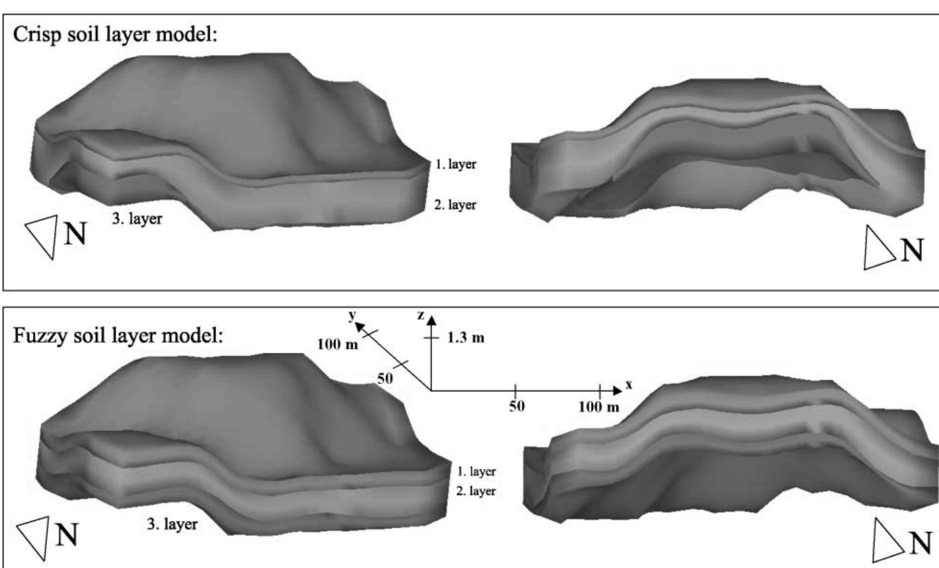

Fig. 4. Continuous 3-D soil layer models using crisp hierarchical cluster analysis and fuzzy k-mean classification.

Table 7
Root Mean Square Errors (RMSE) and Coefficient of Efficiency (E) used to evaluate the accuracy of soil layers identified by the hierarchical cluster and the fuzzy k-mean analysis

Layer	Crisp classification		Fuzzy classification	
	RMSE	E	RMSE	E
1	3.81	0.96	3.67	0.97
2	4.18	0.90	15.69	0.51
3	4.98	0.86	17.33	0.40

fuzzy soil layer model is listed in Table 6. The fuzzy soil layer model is shown in Fig. 4. When compared to the crisp soil layer model, the fuzzy model showed a thin layer of glacial till on lower landscape positions, and a thicker layer of reworked loess on the shoulder position. The spatial distribution of layers 1 and 2, representing reworked loess and glacial till was very different in the crisp and the fuzzy soil layer model.

3.3. Validation results

The results of validation of the models using RMSE are shown in Table 7. Predictions of the depth of layer 1 showed small RMSE for the hierarchical cluster and the fuzzy k-mean analysis, which means that predictions of layer 1 were very good. Similar small values for RMSE were obtained for layers 2 and 3 of the crisp model. Predictions were poor for layers 2 and 3 using the fuzzy k-mean method. We used E to evaluate the fit of actual vs. predicted data. The Coefficient of Efficiency was larger for layer 1 using crisp and fuzzy layer predictions and layers 2 and 3 using the crisp analysis, when compared to an ideal E of 1. Smaller Es were obtained for layers 2 and 3 using the fuzzy classification for layer prediction.

4. Discussion

Results indicate that the crisp method overall produced better predictions than the fuzzy method. Four classes, as identified with simple threshold criteria for similarity and sample size, were sufficient to describe clusters which differed in their soil physical (penetration resistance, texture, bulk density and water content) and topographic (elevation, slope, upslope drainage area) characteristics.

There are several reasons for the poor performance of the fuzzy k-mean classification in developing an accurate 3-D soil layer model, as exemplified here. First, as the membership values were transformed into layer depth information by defuzzification, it contradicts with the concept of vagueness of fuzzy set theory because it transforms fuzzy output back into crisp output. Second, the use of indices such as F' and H' may not have produced the optimal combination of number of classes and fuzzy exponent φ. Several studies indicated successful uses of these indices (Bezdek, 1981; Odeh et al., 1992; Lagacherie et al., 1997). However, in some other studies, these indices did not yield satisfactory results, because they do not give clear optimal results (Vriend et al., 1988; Irwin et al., 1997; Triantafilis and McBratney, 1993). Furthermore, the least fuzzy or least disorganized classification is considered to be the optimal solution. Odeh et al. (1990) suggest that this is rather paradox since choosing a fuzzy approach means that the idea of a crisp solution is rejected a priori. Some studies used trial and error methods (Mazaheri et al., 1995; Hendricks Franssen et al., 1997) to identify the "best and most reasonable" balance between continuous soil variation and soil classes. De Bruin and Stein (1998) used an external criterion analysis, where variables not used in the classification guide the evaluation of validity of the fuzzy c-partition. Molenaar (1993) and Milligan (1996) stress that a partition is useful only if the results can be understood within the context of the research question at issue. In spite of the poor performance by the fuzzy model, it was able to explain the four different classes in relation to the landscape and the class centroids of different topographic attributes, CI/depth values and soil attributes look plausible.

5. Summary and conclusions

In this study, we compared the performance of crisp horizontal hierarchical cluster analysis and a fuzzy k-mean classification for creating 3-D continuous soil layer models. The classification methods used the cone penetrometer measurements and topographic attributes to identify soil layers. Validation results showed higher accuracy for the predicted crisp soil layer model in comparison with the fuzzy soil layer model. A reason for the excellent performance of the crisp classification might be the dense penetrometer sampling with a spacing of 10 m between sampling points to create the models. Sparser sampling on the same study site might have produced poorer results, where the description of vagueness of data would have become an important factor. Scale factor might be important while making a decision about whether to use crisp or fuzzy classification to create soil layer models.

An advantage of using PCP is rapid data collection. These data, combined with limited soil coring can be used to create 3-D soil layer models. We hope to apply our method to other sites with different soil and landscape characteristics. Additionally, more testing at different scales is necessary to validate our approach.

Generally, soil layer models describe the spatial variability of changes in layer depth and/or soil material across a landscape. Because optimal site-specific management and crop growth is highly dependent on soil characteristics such as texture, bulk density and penetration resistance, such models are extremely useful. Continuous 3-D soil-layer models will benefit farmers to manage their land and soil, and scientists and researchers to realistically describe the distribution of soil patterns. Furthermore, soil layer and material information is used as input for environmental assessment and water quality simulation models. These models require crisp soil data input and currently, they are not able to take advantage of membership functions or fuzzy information.

References

Ameskamp, M., 1997. Three-dimensional rule-based continuous soil modeling. PhD Thesis, Christian-Albrechts-University Kiel, Germany.

Bandemer, H., Gottwald, S., 1995. Fuzzy Sets, Fuzzy Logic and Fuzzy Methods. Wiley, Chichester.

Barak, P., Nater, E., 1999. The virtual museum of minerals and molecules. Available on-line at http://www.soils.wisc.edu/virtual_museum/index.html (posted 1998; verified 25 Oct. 1999).

Barraclough, A., Guymer, I., 1998. Virtual reality—a role in environmental engineering education? Water Sci. Technol. 38 (11), 303–310.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York, NY.

Bezdek, J.C., Pal, S.K., 1992. Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data. Institute of Electrical and Electronics Engineers, New York.

Bierkens, M.F.P., Burrough, P.A., 1993. The indicator approach to categorical soil data: I. Theory, II. Application to mapping and land use suitability analysis. J. Soil Sci. 44, 361–381.

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

Burrough, P.A., van Gaans, P.F.M., Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77, 115–135.

Cooperative Research Centre for Landscape Evolution and Mineral Exploration, 1999. Annual Report 1998/1999. Available at: http://leme.anu.edu.au/annual/ar99/mapping99.html (verified June 20, 2000).

De Bruin, S., Stein, A., 1998. Soil-landscape modelling using fuzzy c-means clustering of attribute data derived from a digital elevation model (DEM). Geoderma 83, 17–33.

Garcia, M., Froidevaux, R., 1997. Application of geostatistics to 3D modelling of contaminated sites: a case study. In: Soares, A. (Ed.), Geostatistics for Environmental Applications. Kluwer Academic Publishing, Netherlands, pp. 309–325.

Goderya, F.S., Dahab, M.F., Woldt, W.E., Bogardi, I. et al., 1996. Spatial patterns analysis of field measured soil nitrate. In: Srivastava, S. (Ed.), Geostatistics for Environmental and Geotechnical Applications. ASTM STP 1283, Am. Soc. Testing and Materials, pp. 248–261.

Goovaerts, P. et al., 1997. Geostatistics for natural resources evaluation. Applied Geostatistics Series. Oxford Univ. Press, New York.

Goovaerts, P., Webster, R., 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur. J. Soil Sci. 45, 79–95.

Grunwald, S., Barak, P., McSweeney, K., Lowery, B., 2000. Soil landscape models at different scales portrayed in Virtual Reality Modeling Language (VRML). Soil Sci. 65, 598–614.

Grunwald, S., Rooney, D.J., McSweeney, K., Lowery, B., 2001. Development of pedotransfer functions for a profile cone penetrometer. Geoderma 100, 25–47.

Heijs, A.W.J., de Lange, J., Schoute, J.F.Th., Bouma, J., 1995. Computed tomography as a tool for non-destructive analysis of flow patterns in macroporous clay soils. Geoderma 64, 183–196.

Hendricks Franssen, J.J.W.M., van Eijnsbergen, A.C., Stein, A., 1997. Use of spatial prediction techniques and fuzzy classification for mapping soil pollutants. Geoderma 77, 243–262.

Irwin, B.J., Ventura, S.J., Slater, B.K., 1997. Fuzzy and isodata classification of landform elements form digital terrain data in pleasant Valley, Wisconsin. Geoderma 77, 137–154.

Isaaks, E.H., Srivastava, R.M., 1989. Applied Geostatistics. Oxford Univ. Press, New York.

Lagacherie, P., Cazemier, D.R., van Gaans, P.F.M., Burrough, P.A., 1997. Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area. Geoderma 77, 197–216.

Mazaheri, S.A., Koppi, A.J., McBratney, A.B., 1995. A fuzzy allocation scheme for the Australian Great Groups Classification system. Eur. J. Soil Sci. 46, 601–612.

McBratney, A.B., de Grujiter, 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., Moore, A.W., 1985. Application of fuzzy sets to climatic classification. Agric. For. Meteorol. 35, 165–185.

McBratney, A.B., Odeh, I.O.A., 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77, 85–113.

McBratney, A.B., Hart, G.A., McGarry, D., 1991. The use of region partitioning to improve the representation of geostatistically mapped soil attributes. J. Soil Sci. 42, 513–532.

Milligan, G.W., 1996. Clustering validation: results and implications for applied analysis. In: Arabie, P., Hubert, L.J., De Soete, G. (Eds.), Clustering and Classification. World Scientific Publ., River Edge, NJ, pp. 341–375.

Minasny, B., McBratney, A.B., 1999. Fuzzy k-mean with extragrade program—FuzME version 1.0. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006. Available at http://www.usyd.edu.au/su/agric/acpa/fkme/FkME.html (posted March 1999; verified Nov. 1999).

Molenaar, M., 1993. Object hierarchies and uncertainty in GIS, or why is standardization so difficult? Geogr. Inf. Syst. 6 (4), 22–28.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models—Part I: A discussion of principles. J. Hydrol. 10, 282–290.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1990. Design of optimal sampling spacings for mapping soil using fuzzy-k-means and regionalized variable theory. Geoderma 47, 93–122.

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

Oeborn, I., 1989. Properties and classification of some acid sulfate soils in Sweden. Geoderma 45, 197–219.

Pereira, V., FitzPatrick, E.A., 1998. Three-dimensional representation of tubular horizons in sandy soils. Geoderma 81, 295–303.

Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. Eur. J. Oper. Res. 10, 294–301.

Simonson, R.W., 1994. Morphology and composition of specimen soils, Okinawa. Geoderma 63, 19–42.

Sinowski, W., Auerswald, K., 1999. Using relief parameters in a discriminant analysis to stratify geological areas with different spatial variability of soil properties. Geoderma 89, 113–128.

Slater, B.K., 1994. Continuous classification and visualization of soil layers: a soil-landscape model of Pleasant Valley Wisconsin. Doctoral dissertation, Department of Soil Science, University of Wisconsin-Madison.

Soil Survey Staff, 1998. Keys to Soil Taxonomy. Available at http://www.statlab.iastate.edu/soils/keytax/ (verified 28 Dec. 1999).

SPSS Professional Statistics, 1994. SPSS User's Manual Version 6.1. SPSS.

Stibbard, A., 1997. Warwick University Forum, No. 6.

Triantafilis, J., McBratney, A.B., 1993. Application of Continuous Methods of Soil Classification and Land Suitability Assessment in the Lower Namoi Valley. Dept. of Agricultural Chemistry and Soil Science, University of Sydney, Australia.

Vitek, J.D., Giardino, J., Fitzgerald, J.W., 1996. Mapping geomorphology: a journey from paper maps, through computer mapping to GIS and virtual reality. Geomorphology 16, 233–249.

Vriend, S.P., van Gaans, P.F.M., Middelburg, J., de Nijs, A., 1988. The application of fuzzy c-means cluster analysis and non-linear mapping to geochemical datasets: examples from Protugal. Appl. Geochem. 3, 213–224.

Ward, A.W., Ward, W.T., McBratney, A.B., de Gruijter, J.J., 1992. MacFUZZY a user friendly program for data analysis by generalized fuzzy k-means on the Macintosh. CSIRO Div. of Soils Div. Rep. No. 116, CSIRO, Australia.

Webster, R., Oliver, M.A., 1990. Statistical Methods in Soil and Land Resource Survey. Oxford Univ. Press, Oxford.

Yang, M.-S., 1993. Survey of fuzzy clustering. Math. Comput. Modell. 18 (11), 1–16.

Zaheh, L.A., 1965. Fuzzy sets. Inf. Control 8, 338–353.