Geoderma 155 (2010) 164–174

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Construction of membership functions for predictive soil mapping under fuzzy logic

A-Xing Zhu [a,b,c], Lin Yang [b,d,*], Baolin Li [b], Chengzhi Qin [b], Tao Pei [b], Baoyuan Liu [e]

[a] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
[b] State Key Laboratory of Environment and Resources Information System, Institute of Geographic Sciences and Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[c] Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA
[d] Graduate University of the Chinese Academy of Sciences, Beijing 100049, China
[e] College of Geography and Remote Sensing Sciences, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 31 May 2008
Received in revised form 20 April 2009
Accepted 27 May 2009
Available online 21 June 2009

Keywords:
Fuzzy membership function
Digital soil mapping
Purposive sampling
Knowledge on soil-environment relationships
SoLIM

ABSTRACT

Fuzzy membership function is an effective tool to represent relationship between soil and environment for predictive soil mapping. Usually construction of a fuzzy membership function requires knowledge on soil-landscape relationships obtained from local soil experts or from extensive field samples. For areas with no soil survey experts and no extensive soil field observations, a purposive sampling approach could provide the descriptive knowledge on the relationships. However, quantifying this descriptive knowledge in the form of fuzzy membership functions for predictive soil mapping is a challenge. This paper presents a method to construct fuzzy membership functions using descriptive knowledge. Construction of fuzzy membership functions is accomplished based on two types of knowledge: 1) knowledge on typical environmental conditions of each soil type and 2) knowledge on how each soil type corresponds to changes in environmental conditions. These two types of knowledge can be extracted from catenary sequences of soil types and the associated environment information collected at a few field samples through purposive sampling. The proposed method was tested in a watershed located in Heshan farm of Nenjiang County in Heilongjiang Province of China. A set of membership functions were constructed to represent the descriptive knowledge on soil-landscape relationships, which were derived from 22 field samples collected through a purposive sampling approach. A soil subgroup map and an A-horizon soil organic matter content map for the area were generated using these membership functions. Forty five field validation points were collected independently to evaluate the two soil maps. The soil subgroup map achieved 76% of accuracy. The A-horizon soil organic matter content map based on the derived fuzzy membership functions was compared with that derived from a multiple linear regression model. The comparison showed that the soil organic content map based on fuzzy membership functions performed better than the soil map based on the linear regression model. The proposed method could also be used to construction membership functions from descriptive knowledge obtained from other sources.

1. Introduction

The need for detail and continuous spatial soil information and the availability of spatial information processing techniques have promoted the development of digital soil mapping techniques (Peterson, 1991; Band and Moore, 1995; Zhu and Mackay, 2001). Fuzzy set theory has been widely used in soil science for soil classification and mapping, land evaluation, fuzzy soil geostatistics, soil quality indices (Chang and Burrough, 1987; Burrough, 1989; Zhu et al., 1996; McBratney and Odeh, 1997; McBratney et al., 2003; Zhang et al., 2004; Lagacherie, 2005). The development of fuzzy logic-based digital soil mapping techniques has attracted much attention in the digital soil mapping community due to its ability to capture and represent the continuous nature of soil spatial variation (Zhu and Band, 1994; Burrough, 1996; Dobermann and Oberthur, 1997; McBratney and Odeh, 1997; Zhu, 1997; Zhu et al., 2001; Yang et al., 2007). In fuzzy logic-based approaches, soil spatial variation is expressed as spatial variation of membership in soil classes (Zhu, 1997; McBratney et al., 2000; Qi et al., 2006), which is then used to produce conventional soil class maps and to predict spatial variation of specific soil properties (Zhu et al., 1996). Membership in soil classes is generally derived in two ways (McBratney et al., 2000): continuous classification using techniques such as fuzzy *c*-means (FCM) (De Gruijter and McBratney, 1988; McBratney et al., 1992; De Gruijter et al., 1997) and Semantic Import Model (SI) (Burrough et al., 1992). The former determines membership by partitioning observations into relatively stable natural groups based on multivariate attributes. In other words, it is a data-driven approach. The latter determines membership based on fuzzy membership functions derived from class limits which are based on

* Corresponding author. State Key Laboratory of Environment and Resources Information System, Institute of Geographic Sciences and Resources Research, Chinese Academy of Sciences, Beijing 100101, China. Tel.: +86 10 64889461.
E-mail address: Yanglin@lreis.ac.cn (L. Yang).

0016-7061/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2009.05.024

expert knowledge or conventionally imposed definitions. It is a knowledge-based approach.

The key issue in this knowledge-based approach to fuzzy membership function definition is the determination of class limits and membership gradation within these class limits. Based on the assumption that there is a relationship between soil and environment, Zhu (1999) developed a personal construct-based approach to extract the knowledge on soil-environment relationships from local soil experts and represent the knowledge as optimality curves (membership curves), which are then used to approximate the needed fuzzy membership functions for digital soil mapping under fuzzy logic. Lagacherie (2005) proposed a procedure based on possibility theory and fuzzy pattern matching to translate soil class description in soil database into a set of membership functions. Qi et al. (2006) developed a prototype-based fuzzy soil mapping approach to represent soil-environment knowledge as fuzzy membership functions, which were also constructed based on the knowledge obtained from soil experts. Qi et al. (2008) developed a data mining method using the Expectation Maximization (EM) algorithm to define membership functions based on the information extracted from conventional soil class maps. Liu and Zhu (2009) developed a mapping with words approach based on computational theory of perceptions to define membership functions.

These techniques either require local soil experts or large amount of field soil samples or the existence of conventional soil maps. For areas with no soil survey experts, no extensive soil field observations and no existing soil maps, Zhu et al. (2008) and Yang et al. (2007) developed a purposive sampling approach based on a fuzzy c-means classification method to determine the few typical locations for field investigation from which a descriptive knowledge on relationship between soil and environment can be obtained. The approach has been successful in mapping spatial variation of discrete soil classes. However, to be able to map spatial continuity of soils using the so-acquired descriptive knowledge under fuzzy logic, methods for constructing fuzzy membership functions quantifying the descriptive knowledge on soil-landscape relationships are needed. This paper presents a method to construct fuzzy membership function from descriptive knowledge which can be obtained through purposive sampling or other knowledge elicitation methods (Bui et al., 1999; Qi and Zhu, 2003) for predictive soil mapping.

2. Materials and methods

2.1. Study area and environmental data

The study area is located in Heshan farm of Nenjiang County in Heilongjiang province of China (Fig. 1). Its area is 60 km^2 with elevation ranging from 276 m to 363 m and slope gradient mostly under 4°, which is indicative of a generally gentle environmental gradient. The original vegetation is meadow, but it has been cultivated as cropland over the past 40 years. Crop in the watershed at present is generally limited to soya bean and wheat. The soils in the area are formed on deposits of silt loam loess and have a thick *A* horizon with high organic matter content. Due to this thick and dark *A* horizon people (both soil scientists in China and local people in the area) call them "black soils". The parent materials for the area are the same over the whole area except in the valley bottom which is mainly occupied by fluvial deposits. The land use and soil management have been pretty uniform and no organic fertilizer has been applied to the area because of the naturally high contents of organic matter in these soils.

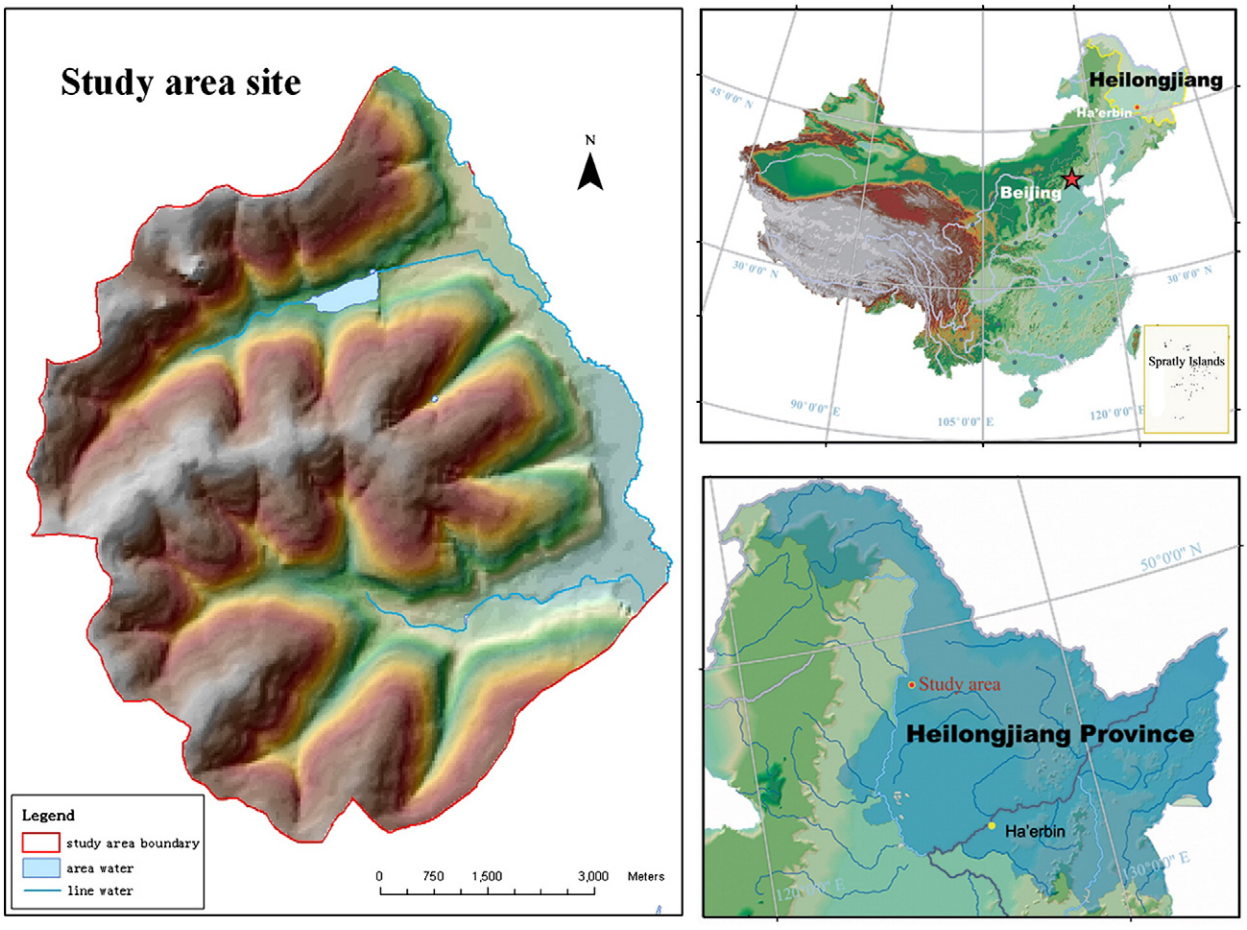

Fig. 1. Location and DEM of the study area.

The following four topographic variables (slope gradient, contour curvature, profile curvature and topographic wetness index) were used in this study to characterize the environment conditions. The selection of four topographic variables is based on the fact that the area is small (60 km^2) and the macro-climate is pretty similar over the area and micro-climate conditions can very much be captured using the topographic conditions captured by the topographic variables used. The vegetation and parent materials are similar across the study area. As result only the above four topographic variables are needed to separate the soils in the area (Yang et al., 2007). We want to point out that we are not arguing nor saying or suggesting that four topographic variables are only what we need for predictive soil mapping for any area. In fact, what variables are needed depends on the nature of pedogenesis in the specific area.

Information of slope, planform curvature, and profile curvature were derived from a 10 m resolution DEM (Fig. 1) which was created from the 1:10,000 topographic map of the area using a terrain analysis software 3DMapper (www.terrainanalytics.com). Topographic wetness index was calculated according to the following equation (Beven and Kirkby, 1979): $w = \ln(a/\tan\beta)$, where a is the cumulative upslope area draining through a point (per unit contour length), β is the slope gradient at the point. Because the relief of our study area is gentle and the floodplain is wide, multiple-flow strategy MFD-fg was used to calculate the upslope drainage area (a) (Qin et al., 2006). Max down slope was used as (β) because it is considered better to express the effect of the relief on surface water distribution than the average slope (Hjerdt et al., 2004). A post process procedure was applied to overcome the problem of high wetness values concentrated along the flow line in the wide floodplain and allow wetness index value change gradually from the flow line to the edge of the floodplain (Qin et al., 2006).

Chinese soil taxonomy is chosen as the soil taxonomy system (Chinese Soil Taxonomy Research Group, 2001). The Chinese soil taxonomy is a system based on diagnostic horizons and diagnostic characteristics. It introduces some diagnostic horizons concepts from US Soil Taxonomy and World Reference Base for Soil Resources and also defined new diagnostic horizons and diagnostic characteristics specific to soils in China. It has six categories: Order, Suborder, Group, Subgroup, Family and Series. The first four are the high categories and the last two are the basic (low) levels. The Chinese soil taxonomy system is rather new and levels of soil units lower than subgroup have not been defined. Subgroup is currently used as the basic soil unit for soil mapping in this study. The subgroup is the auxiliary unit of soil Group and is defined according to whether the soils deviate from the central concept of a Group, or if they have some characteristics resulting from additional processes, or have remnant features inherited from the parent materials. The subgroup that corresponds to the central concept of a Group is defined as Typic.

2.2. Descriptive knowledge obtained from purposive sampling

We use the purposive sampling method proposed by Yang et al. (2007) and Zhu et al. (2008) as an illustration of how to obtain descriptive knowledge for areas with no existing soil information (such as our study area). Extensive discussion on purposive sampling is not necessary here because it is not the topic of this paper. Interested readers are referred to the above references for details. A brief overview of it and how it is used in this paper are given as follow. The basic idea of purposive sampling is to sample the locations where the soils are typical of the soil categories (Yang et al., 2007; Zhu et al., 2008). Based on the soil-landscape model theory (Hudson, 1992), the assumption is that typical instances of soil classes correspond to unique configurations (combinations) of environment conditions. A fuzzy classification technique (fuzzy c-means classifier, FCM) was used to identify the unique combinations (or environment classes) that existed in the environmental data set. 13 environment classes were identified to be the optimal number of classes in the study area based on the improvements in partition coefficient and entropy of classification (see Yang et al., 2007 for detail). Membership maps of the derived 13 environment classes were generated and locations with high membership values in these environment classes were considered as locations of typical soil instances. For each environmental class, two or three points were selected. Field investigation was then made at these points to identify the soil type (soil subgroups in this case). Soils type was identified at each site by a soil classification expert. If the soil types of first two points were the same, then this environmental class was indicated to be associated with this soil type. If not, the third point was collected to indicate the soil type of environmental class. This collection of field points is called as the "explanation set" because they used to associate the environmental combinations (classes) to soil types. Twenty two points were finally selected (Fig. 2) and 6 soil subgroups were identified.

Once association between soil types and environment combinations were established through explanation points, a catenary sequence of soil types relating to environment classes was built using the spatial adjacency between environment classes.

Environment classes those correspond to the same soil type and are spatially adjacent can be considered as one instance of this soil type. For example, Classes 1 and 3 located at ridge tops and shoulders. Due to their spatial adjacency, these two classes were considered to be one instance of Mollic Bori–Udic Cambosols. Environment classes those correspond to the same soil type but are not spatially adjacent are considered as separate instances of this soil type, as if they were different soil types. Environment Classes 9, 4, 7, and 11 corresponded to Typic Hapli–Udic Isohumosols. All located at backslopes, these classes were marked as 'Typic Hapli–Udic Isohumosols-1' as one instance. Environment Class 10 also corresponded to Typic Hapli–Udic Isohumosols but was located at footslopes and separated spatially from "Typic Hapli–Udic Isohumosols-1", thus it was treated as a different instance, marked as 'Typic Hapli–Udic Isohumosols-2'.

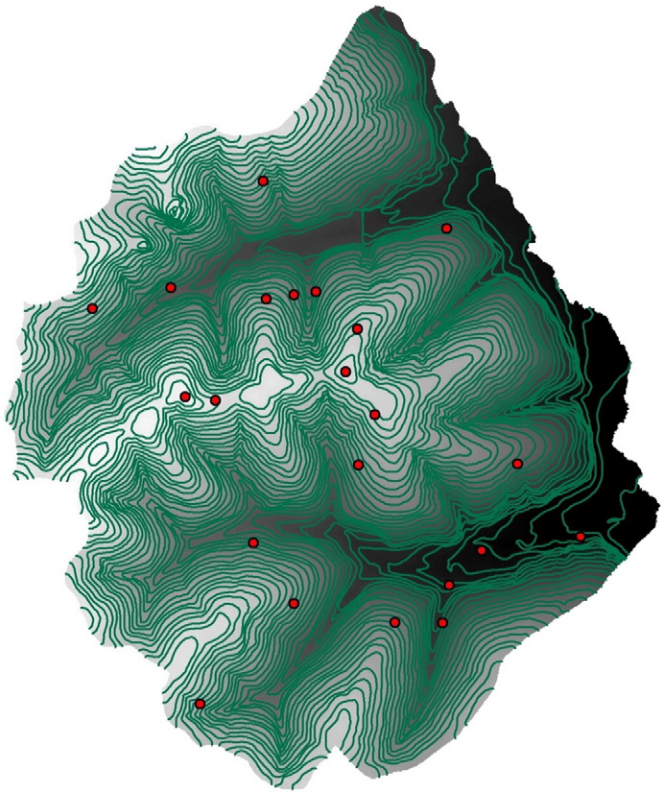

Fig. 2. Location map of the 22 explanation points.

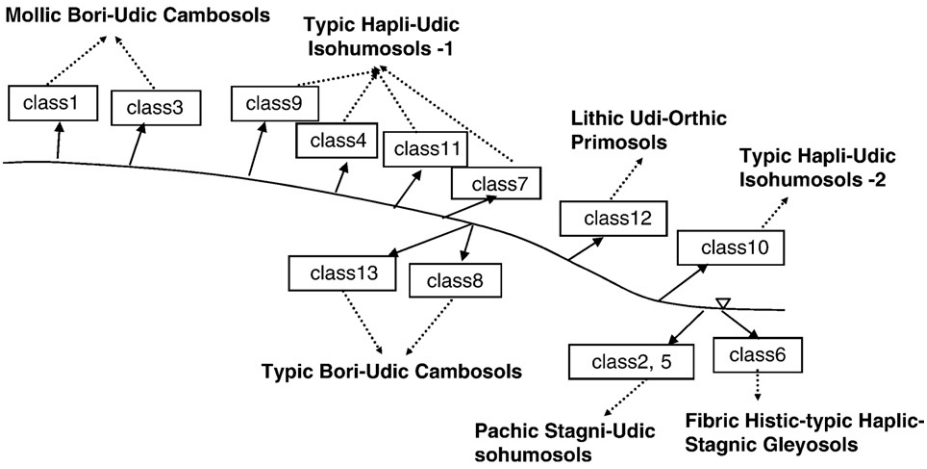

Fig. 3. General catenary sequences of environment classes and soil subgroups.

A catenary sequence of environment classes together with the 6 soil subgroups was developed (Fig. 3). This catenary sequence together with the environment conditions of the explanation points for each environment class (Table 1) constituted the descriptive knowledge about the soil-landscape relationships over the area and formed the basis for constructing membership functions.

2.3. Construction of fuzzy membership functions

Fuzzy membership function describes how similarity between a local soil and the typical case of the given soil type will change as environmental conditions change. The similarity value varies from 0 (which means that local soil is very different from the given soil type) to 1 (which means that local soil is exactly the same with the given soil type). Relationships between soil and its environment can be captured using some combination of three basic forms: bell-shaped, s-shaped and z-shaped (Zhu, 1999). In this paper we used a Gaussian-like function as the basic form of fuzzy membership functions (Eq. (1)) (Zhu, 1999; Shi et al., 2004; Qi et al., 2006):

$$S_{ij,v}^{k} = e^{-\left(\frac{|z_{ij,v} - z_{0,v}^{k}| \times 0.8326}{D_{v}^{k}}\right)^{2}} \quad (1)$$

where $S_{ij,v}^{k}$ is the similarity of the local soil at point (i,j) to soil type k based on environmental variable (factor) v; $z_{ij,v}$ is the value of environmental variable v at the point; $z_{0,v}^{k}$ is the typical value of environmental variable v when the similarity of local soil to soil type k is 1.0; and D_{v}^{k} is the difference between $z_{0,v}^{k}$ and the value of environmental variable value at which the similarity value is 0.5.

Two types of knowledge were needed to construct a membership function (or a membership curve) (Zhu, 1999). The first type, referred to as Type I Knowledge (the $z_{0,v}^{k}$), defines the typical environmental conditions under which a particular soil type would occur. This means that local soils at the locations with these conditions will have maximum similarity (1.0) to the given soil type. The second type, referred to as Type II Knowledge, defines how similarity will change as environmental conditions deviate from the typical conditions. To obtain this type of knowledge, the curve type and the difference between the value of the environmental condition when the similarity value is 1 and 0.5, which is called 'width' (the D_{v}^{k}), need be determined.

Type I knowledge is extracted from the information observed at the explanation points where the local soils are supposed to be typical instances of the soil types. One soil type may be associated with one or more environment classes. If one soil type is associated with only one environment class, the environmental values at the explanation points of the environment class can be considered as the typical environmental conditions of this soil type. If there are more than one explanation point involved, either the average value of environmental condition at all the explanation points or the value of environmental condition of the explanation point with the highest fuzzy membership in that environment class is taken as the typical environmental condition. Typically, we suggest the following guideline when choosing which of the above two strategies to take when determining Type I Knowledge: if the explanation points have similar membership values in that environment class, the average strategy is taken. If one of the explanation points has much higher membership value than the other explanation points for this environmental class, the maximum membership strategy is adopted.

If the soil type is associated with two or more environment classes and these classes are spatially adjacent to each other, that is, they are considered as one instance of the soil type, then the maximum and

Table 1
Environment classes and the environmental conditions of the respective explanation points.

Environment class	Explanation point ID	Slope gradient	Planform curvature (10^{-3}) (1/m)	Profile curvature (10^{-3}) (1/m)	Wetness index	Class membership
Class 1	16	0.00	0.00	0.00	9.17	0.91
	17	0.01	−2.87	−0.08	8.35	0.91
Class 3	6	0.87	−67.48	−1.01	8.49	0.92
	10	0.86	−82.84	−1.25	8.72	0.87
Class 4	22	1.66	−2.53	0.45	9.19	0.61
Class 7	9	2.23	−2.59	0.32	9.23	0.77
	21	2.18	−3.76	−0.04	9.35	0.71
Class 9	8	1.44	−6.01	−1.01	8.62	0.57
	14	1.46	−42.57	0.28	8.92	0.50
Class 11	20	2.06	−17.37	−0.75	9.02	0.74
	3	2.13	−15.72	−1.10	8.89	0.59
Class 8	2	2.44	−3.13	0.30	8.78	0.76
	19	2.60	21.71	0.02	8.81	0.56
Class 13	13	2.36	−11.69	−0.76	8.77	0.85
	15	2.40	3.70	−0.67	8.66	0.70
Class 12	1	3.38	−6.78	0.14	9.00	0.55
	18	3.13	−19.90	0.49	8.50	0.77
Class 10	7	2.35	−0.56	3.15	9.26	0.54
	11	2.06	10.37	3.90	9.76	0.91
Class 2	12	1.21	34.78	1.66	15.71	0.40
Class 5	5	1.19	3.40	1.57	15.97	0.49
Class 6	4	0.09	−20.97	−0.04	18.63	0.86

Table 2
Type I Knowledge with respect to slope gradient for all soil subgroups.

Soil subgroup	Type I Knowledge (typical values)
Mollic Bori–Udic Cambosols	0.005–0.87
Typic Hapli–Udic Isohumosols-1	1.45–2.21
Typic Bori–Udic Cambosols	2.38–2.52
Lithic Udi–Orthic Primosols	3.26
Typic Hapli–Udic Isohumosols-2	2.06
Pachic Stagni–Udic Sohumosols	1.19–1.21
Fibric Histic-typic Haplic–Stagnic Gleyosols	0.09

minimum values of the typical environmental condition values are used to specify the range of the typical environmental condition for the given soil type. For example, Mollic Bori–Udic Cambosols was related to two environment classes (Classes 1 and 3), but those two were treated as one instance because they were spatially adjacent (as shown in Fig. 3). Environment Class 1 had two explanation points (as shown in Table 1). The average strategy was taken in this case to determine the typical slope gradient value for this class because the membership values in Class 1 for the two explanation points were similar (Table 1). Thus the typical gradient value for Class 1 was 0.005. Similarly, the typical gradient value for Class 3 was determined to be 0.865. The optimal slope gradient range (Type I Knowledge) for Mollic Bori–Udic Cambosols was then 0.005 through 0.865.

If the soil type is associated with two or more environment classes but these classes are not spatially adjacent to each other, these classes are treated as separate instances of this soil types. The process used to determine the typical values for soil type associated with one environment class as described above is adopted to deal with each instance. Table 2 contains all of the typical values or ranges for all the subgroups.

To determine the width, we need to know the environment condition where membership value is 0.5. To accomplish this task, for each environment variable we first organize the typical values of that environment variable for all soil types in such a way that each group (sequence) would contain the largest number of the typical values possible and these typical values can be arranged in an ascending order that does not interrupt the order in which they appear in the catenary sequence. For example, the six soil subgroups in our study area were divided into two separate sequences: the first contains Mollic Bori–Udic Cambosols, Typic Hapli–Udic Isohumosols-1, Typic Bori–Udic Cambosols, and Lithic Udi–Orthic Primosols with the typical slope gradient values in an ascending order from 0.005–0.865, through 1.45–2.205, through 2.38–2.52, to 3.255; the second contains Fibric Histic-typic Haplic–Stagnic Gleyosols, Pachic Stagni–Udic Sohumosols, Typic Hapli–Udic Isohumosols-2, and Lithic Udi–Orthic Primosols with the typical slope gradient values in an ascending from 0.09, through 1.19–1.21, through 2.06, to 3.255. To determine the widths for each membership function, it is reasonable to assume that for two soil types which are adjacent along this environment variable their similarity curves with respect to that environment variable overlap and that the crossover point between the two similarity curves is at the middle of their adjacent typical values. It means that the middle value of the two adjacent typical values can be used as the environmental condition where similarity to both soil types is 0.5. For example, the crossover point in the overlap region of the membership functions for Soil Types Mollic Bori–Udic Cambosols and Typic Hapli–Udic Isohumosols-1 along the slope gradient variable is $1/2(0.865+1.45)$ (equals to 1.1575). So the width for the right half of the membership function for Mollic Bori–Udic Cambosols is (1.1575–0.865) (equals to 0.2925) and the width for the left half of the membership function for Typic Hapli–Udic Isohumosols-1 is (1.45–1.1575) (equals to 0.2925).

The curve type of a membership function for each soil type can be determined by their respective order in sequence. For example, typical slope gradient value for Mollic Bori–Udic Cambosols is located at the leftmost end of the slope gradient axis, its fuzzy membership function with respect to this environmental variable (slope gradient here) is considered to be Z-shaped. The membership function for Typic Hapli–Udic Isohumosols-1, Typic Bori–Udic Cambosols, should be the form of asymmetric bell-shaped and the membership function for Lithic Udi–Orthic Primosols should be that of S-shaped curve.

For each catenary sequences, the above procedure is employed. Finally, one can get all the fuzzy membership functions for all the soils. Figs. 4 and 5 illustrate the membership functions for the first sequence and the second sequence, respectively. The soil type (Lithic Udi–Orthic

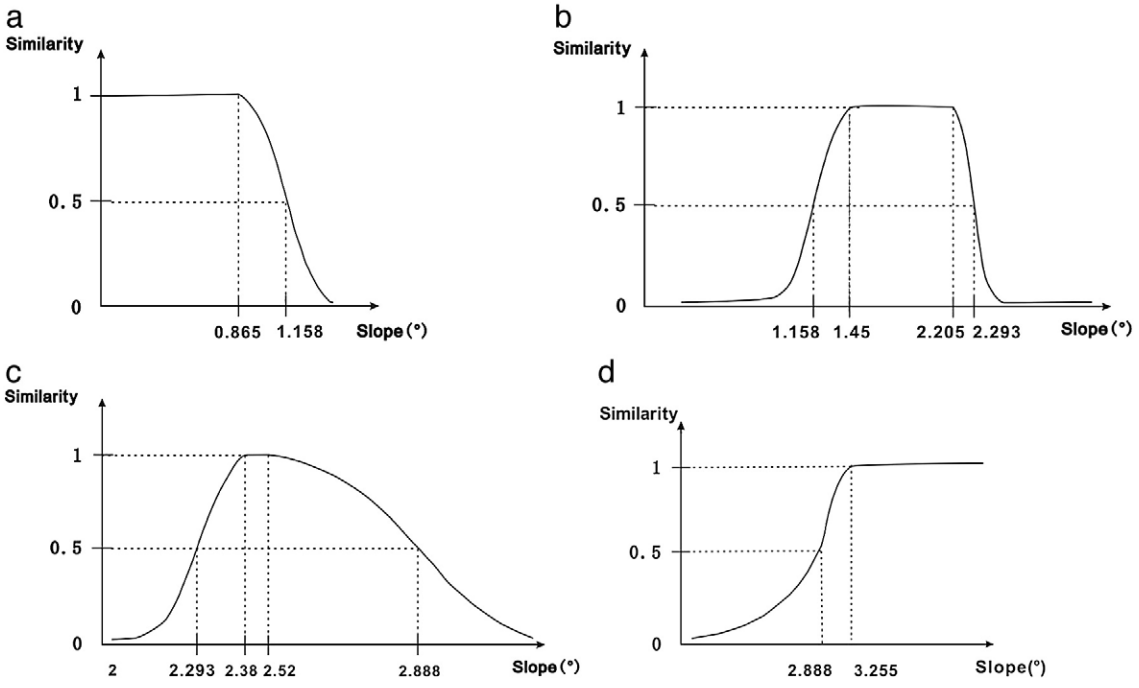

Fig. 4. Fuzzy membership curves with respect to slope for the first sequence: a. Mollic Bori–Udic Cambosols, b. Typic Hapli–Udic Isohumosols-1, c. Typic Bori–Udic Cambosols, d. Lithic Udi–Orthic Primosols.

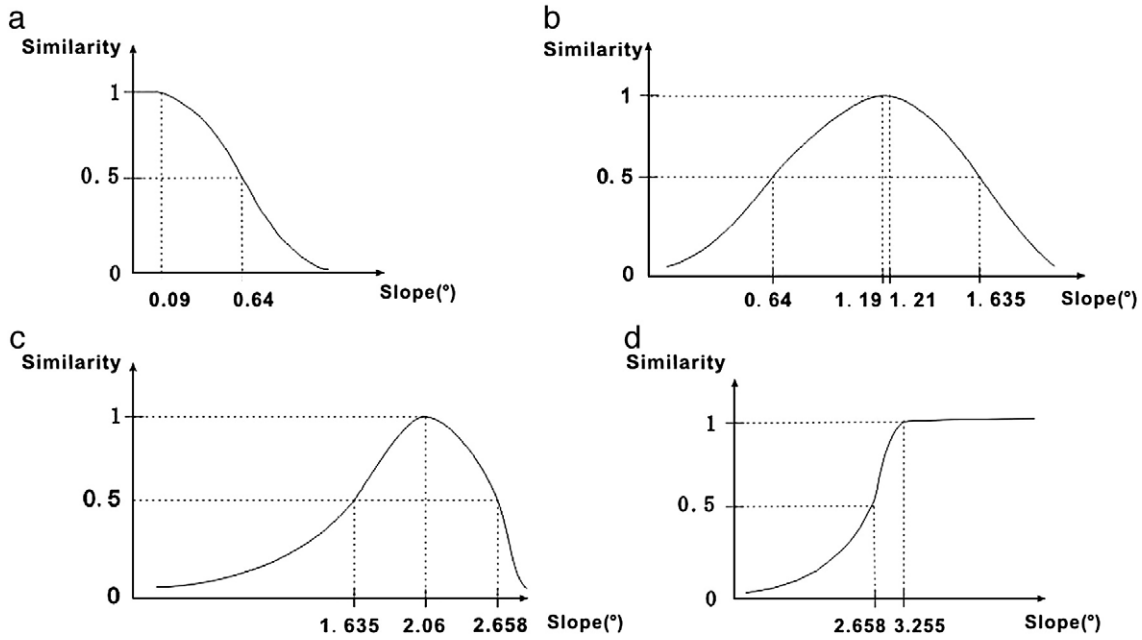

Fig. 5. Fuzzy membership curves with respect to slope for the second sequence: a. Fibric Histic-typic Haplic–Stagnic Gleyosols, b. Pachic Stagni–Udic Sohumosols, c. Typic Hapli–Udic Isohumosols-2, d. Lithic Udi–Orthic Primosols.

Primosols in our study area) which occurs in multiple sequences would have multiple membership functions, one from each sequence, and the final membership function with respect to the given environment variable for this soil type is the membership function which has the narrowest width. This decision for selecting membership function for this soil type is to increase the separability of the final membership functions.

3. Predictive soil mapping using the constructed fuzzy membership functions

Fuzzy membership functions were used under the SoLIM framework to generate fuzzy soil maps for the area. The SoLIM framework is an automated soil inference system that combines the knowledge on soil-environment relationship, expressed as fuzzy membership functions, with environmental conditions characterized using GIS/remote sensing techniques to infer the spatial distribution of soils (Zhu, 1997, 1999; Zhu et al., 2001).

Six fuzzy membership maps, one for each soil type, were generated for the study area. Fig. 6 shows three of them as example. These fuzzy membership maps need to be examined and evaluated to assess the success of the presented approach for constructing membership functions. Direct evaluation of the membership maps is still a challenging research issue. Evaluation of its products might provide good sights to the validity of these membership maps and in turn provide indirect assessment of the presented approach for predictive

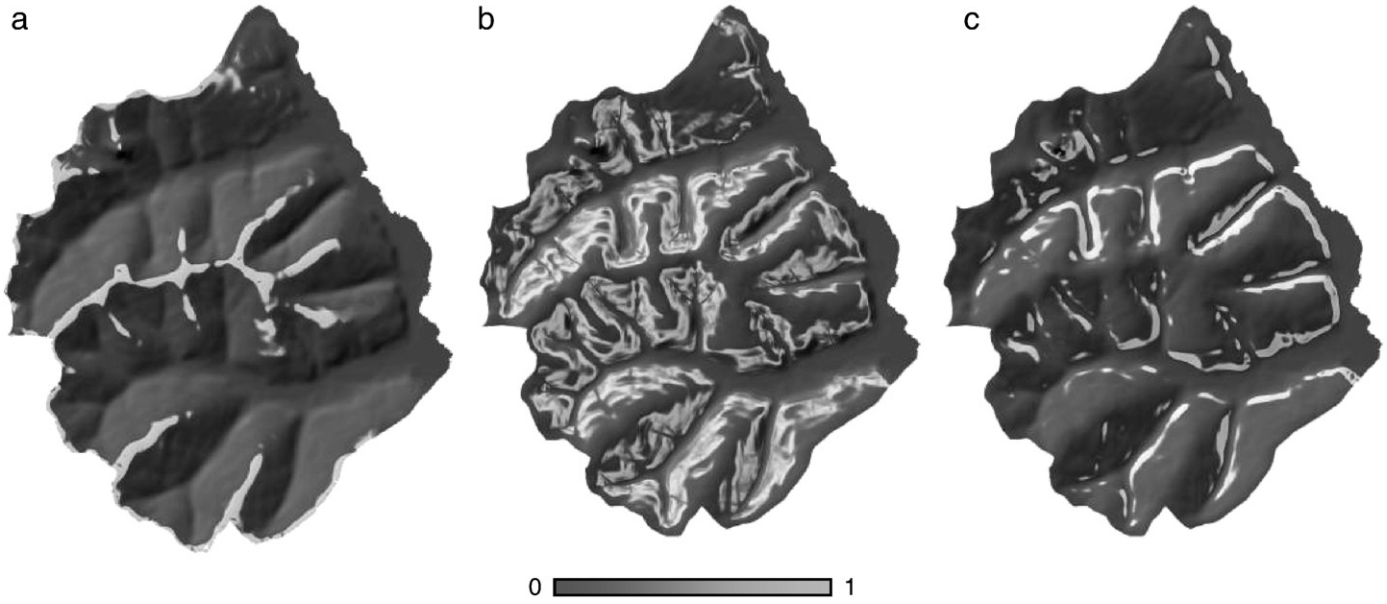

Fig. 6. Membership maps of soil types: a. Mollic Bori–Udic Cambosols, b. Typic Bori–Udic Cambosols, c. Lithic Udi–Orthic Primosols.

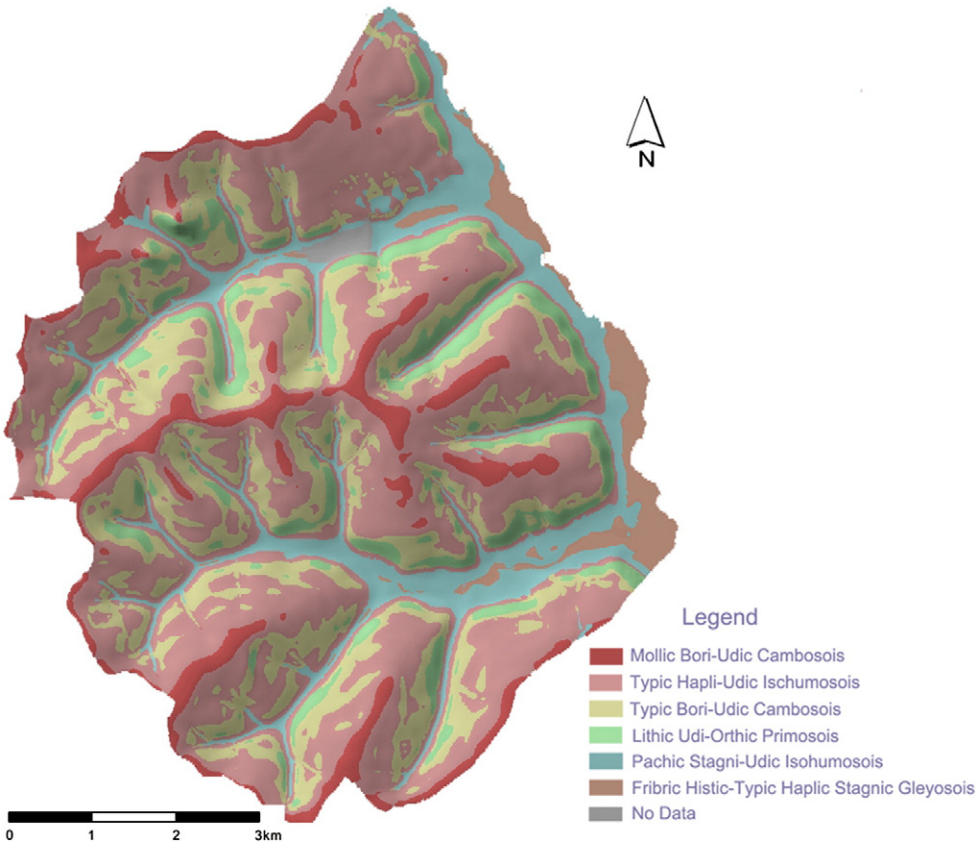

Fig. 7. Soil class map using the membership functions under the SoLIM framework.

soil mapping. In this regard, we derived a soil subgroup map and a soil A-horizon organic content map from the fuzzy membership maps for evaluation.

3.1. Soil subgroup map

A hardened soil subgroup map for the area was created by hardening the fuzzy membership maps. The hardening is done by assigning each location the label of the soil subgroup having the highest membership value for that point (Zhu, 1997). The hardened soil subgroup map derived from SoLIM is shown in Fig. 7.

3.2. A-horizon soil organic matter content map

A weighted average model (Zhu et al., 1997) was used to derive an A-horizon soil organic matter content map over our study area. In this model, it was assumed that the more the local soil environment resembles the environment of a given soil category, the closer the property values of the local soil to the typical property of that soil category. An *A*-horizon soil organic matter content at location (*i*,*j*) can be computed using the following equation:

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} . V^{K}}{\sum_{k=1}^{n} S_{ij}^{k}} \quad (2)$$

where V_{ij} is the organic matter content at site (i, j); V^k is the typical value of the organic matter content of soil subgroup k; S_{ij}^k is the fuzzy membership value of soil subgroup k at (i, j); n is the total number of soil subgroup, which is 6 in our study area.

The typical value of the organic matter content of soil class (V^k) was approximated by the measured organic matter content values of typical points (explanation points) of the soil subgroup. For each soil type, the organic matter content value at the point with the maximum membership value was considered to be the typical organic matter content value of that soil type. There were six soil types in the study area, and soil type Typic Hapli–Udic Isohumosols has two instances which are assumed to have different *A*-horizon organic matter contents. Therefore, we used 7 explanation points, one for each instance of six soil subgroups, to determine the typical organic matter contents (Table 3).

4. Validation and evaluation of the soil maps

4.1. Validation of soil subgroup map

Field validation of the inferred soil subgroup map was conducted in order to evaluate the presented approach for fuzzy membership function construction. We collected 45 validation points in the study area, and soils from all the points were classified and assigned to soil subgroup types by a soil classification expert. Field sites were selected through three sampling strategies: regular sampling, subjective sampling, and transect sampling. A regular sampling with 1100 m

Table 3
Typical soil organic matter contents for all instances of the six soil subgroups.

Soil instances	Typical soil organic matter content (%)
Mollic Bori–Udic Cambosols	3.86
Typic Hapli–Udic Isohumosols-1	3.88
Typic Bori–Udic Cambosols	2.85
Lithic Udi–Orthic Primosols	4.82
Typic Hapli–Udic Isohumosols-2	8.35
Pachic Stagni–Udic Sohumosols	6.20
Fibric Histic-typic Haplic–Stagnic Gleyosols	4.98

by 740 m grid arrangement was used for collecting validation points which were aimed to validate the overall performance of the soil maps (Fig. 8a). The regular sampling strategy produces a total of 30 validation points. Subjective sampling was conducted to investigate areas with unique characteristics where soils were not covered by regular sampling. These unique areas are mostly locations in footslope or floodplain positions. Only 5 subjective points were collected (Fig. 8b). A transect sampling was conducted in such a way that it covered major environmental variations within the shortest distance from ridge top to valley bottom (Fig. 8c). It was used to evaluate how well soil maps capture spatial variation of soil information. The transect contains 10 validation points.

The field observed soil subgroups at these validation sites were compared with the soil subgroups obtained from the inferred soil map at these locations. Soil subgroups from the inferred soil map match field observed soil subgroups at 34 of all the 45 sites, which accounts for about 76% of accuracy. The accuracies of the three sampling strategies were listed in Table 4. The results indicated that the hardened soil map can capture the local variation of soil and the overall spatial distribution of soil well. Unfortunately, there is no large scale soil map which can be used to compare with the inferred soil map in our study area. Given that the accuracy of most 1:24,000 scale soil maps produced in U.S. is about 60% (Zhu et al., 2001), 76% accuracy is acceptable for an initial soil mapping. This suggests that

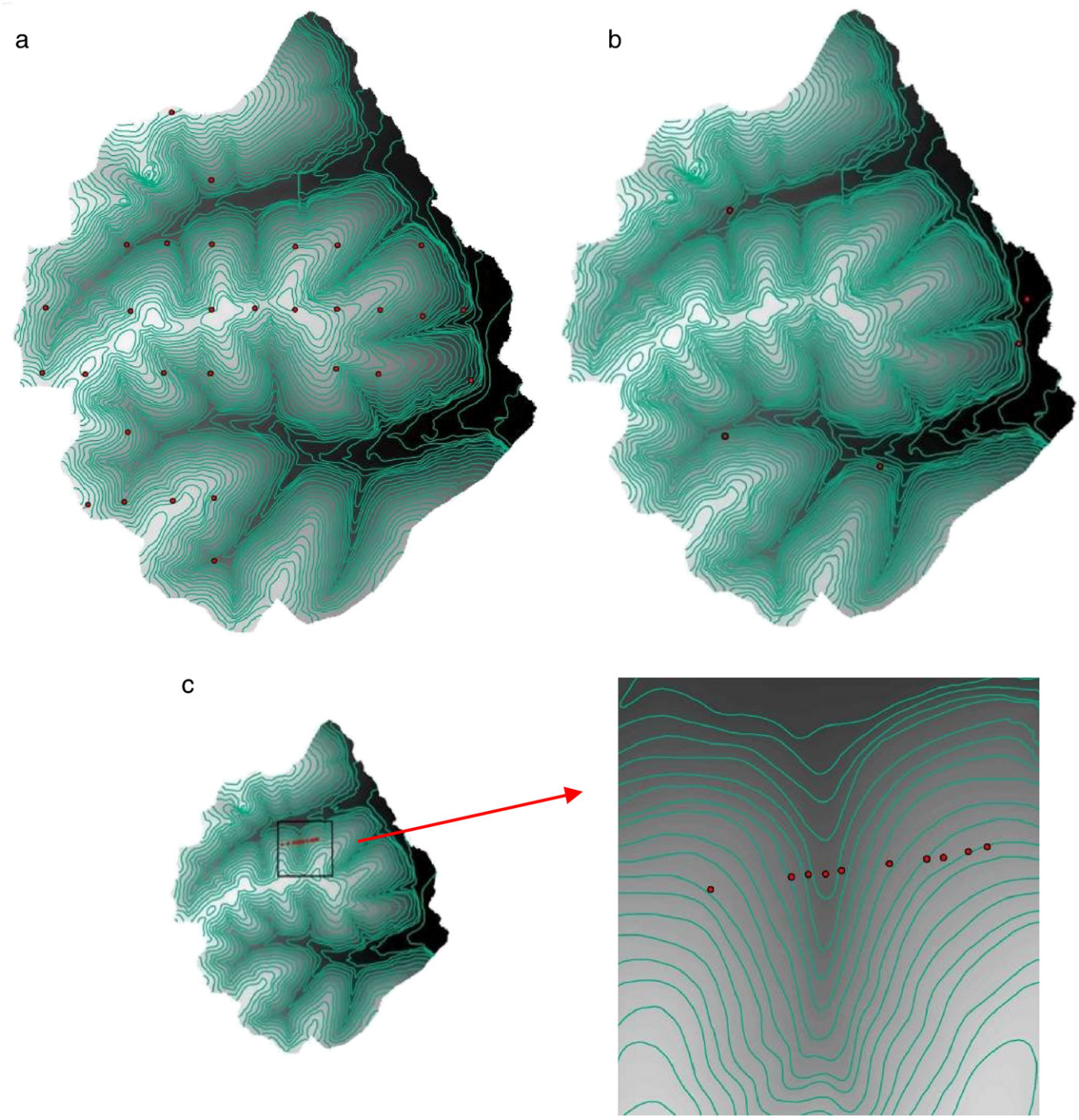

Fig. 8. Location maps of validation points (green lines are the contour lines): a. Regular sampling, b. Subjective sampling, c. Transecting sampling.

Table 4
The accuracy of three sampling strategies.

	Number	Accuracy (%)
Regular sampling	30	68.9
Subjective sampling	5	100
Transect sampling	10	80
Total	45	76

the membership functions constructed in this study do capture the major pattern of soil-landscape relationships over the area.

4.2. Evaluation of soil organic matter content map

The evaluation of the soil organic matter content map based on fuzzy membership method is conducted in two steps. The first step consists of two parts: the evaluation of spatial patterns of the predicted organic matter content distribution and the validation of the predicted map against field observed values at the validation points. The second step is to compare the predicted map based on the fuzzy approach and that derived from a statistical approach.

Fig. 9 shows the top soil organic matter content variation over the study area which was derived based on the fuzzy membership maps. It shows that areas with steep slope and areas with divergent slope tend to have low organic matter content, while areas with convergent slope tend to have high A-horizon soil organic matter content. This spatial variation pattern of soil A-horizon organic matter content matches the expected spatial variation of soil organic matter content well. Areas with steep slope gradient and areas with divergent slopes tend to be dominated by erosive process which reduces the organic matter content in the soil while areas with convergent slope tend to be associated with depositional processes which typically lead to higher organic matter content.

To validate the soil organic matter content map soil organic matter contents were measured at 43 validation points. Two indices were used for this validation: root mean square (RMSE) and agreement coefficient (AC). AC is defined as (Willmott, 1984):

$$\mathrm{AC} = 1 - \frac{n^{*}\mathrm{RMSE}^2}{\mathrm{PE}} \tag{3}$$

$$\mathrm{PE} = \sum_{j=1}^{n} \left(|P_i - \overline{O}| + |O_i - \overline{O}| \right)^2 \tag{4}$$

where n is the number of observations, PE is the potential error variance. O is the observed mean, P_i and O_i are the estimated and observed value, respectively. AC value varies between 0 and 1, with 1 indicating perfect agreement between the predicted and observed values and 0 meaning complete disagreement between the two. This measure is considered to be a more objective measure than the coefficient of determination of a linear regression because the latter does not distinguish a perfect match between two sets of values and a perfect correlation between the two sets of values (Willmott, 1984).

The RMSE for 43 points from the predicted map is 1.04% while the standard deviation of the validation points is 1.17%. This means that our prediction of soil organic matter values at these points is better than predicting these values using the average of the sample points. This is further supported by the computed AC which is 0.71 which indicates a good agreement (or a good match) between the predicted values and the observed values at these points.

In order to compare the soil organic matter content map based on fuzzy membership method with that produced with a statistical

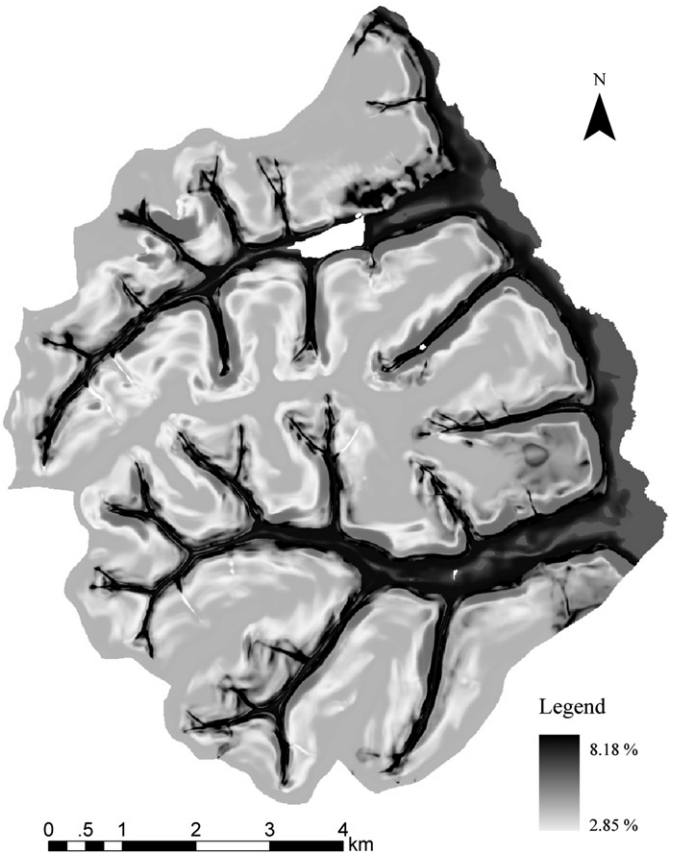

Fig. 9. The A-horizon soil organic matter map based on the derived fuzzy membership functions.

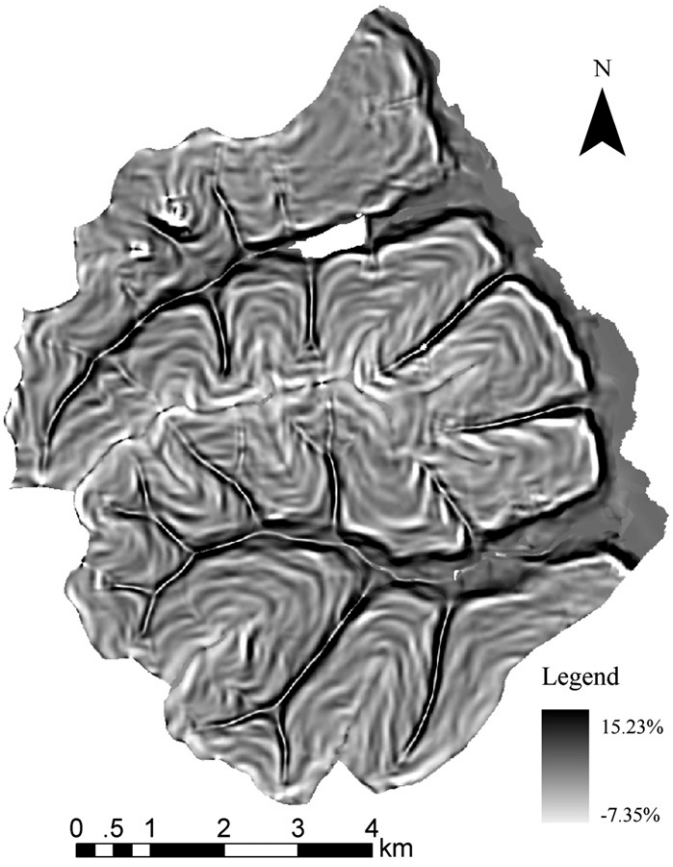

Fig. 10. The A-horizon soil organic matter content map based on the regression model.

Table 5
Comparison of the fuzzy membership function approach with the linear regression model.

	RMSE	AC
Fuzzy membership approach	1.04	0.71
Linear regression model	1.47	0.49

model we developed a multiple linear regression model for predicting A-horizon soil organic matter content using four terrain attributes (slope gradient, contour curvature, profile curvature and topographic wetness index) after Moore et al. (1993) and Gessler et al. (1995). 41 modeling points independent of the 43 validation points, were used to construct the statistical model. Below is the regression model:

$$y_{\text{A-organic}} = 3.509 + 0.1{*}x_{\text{slope}} - 43.325{*}x_{\text{planform}} + 6867.359{*}x_{\text{profile}} + 0.072{*}x_{\text{wetness}} \left(R^2 = 0.512\right) \quad (5)$$

where $y_{\text{A-organic}}$, x_{slope}, x_{planform}, x_{profile}, x_{wetness} is the A-horizon soil organic matter, slope gradient, planform curvature, profile curvature, and wetness index, respectively. Then the A-horizon soil organic matter content over the study area was calculated using the above function (Fig. 10).

The map from the regression model shows different distribution of the soil organic matter from the map based on fuzzy membership functions. The map from the regression model has an overly strong imprint of the profile curvature. It might be true that profile curvature has impact on A-horizon soil organic matter content, but such strong influence of profile curvature over this area cannot be realistic. In addition, the regression-based map shows some negative organic matter content values which are artifacts introduced by the regression model.

The values of MAE (mean absolute error), RMSE, and AC for the statistical model and those for the fuzzy model are listed in Table 5 for comparison, The MAE and RMSE statistics for the fuzzy membership inferred map are consistently lower than those for the soil map derived from the regression model, and the AC value for fuzzy membership method is significantly higher. These observations support the conclusion that is the fuzzy membership approach is an effective way for constructing fuzzy membership functions for predictive soil mapping. In turn it again supports the claim the membership function construction method is effective.

5. Discussion

The validity of the membership functions produced using the new method depends on two major factors: the method itself and the quality of the descriptive knowledge used. Clearly, the purposive sampling approach asserts its impact on the validity of the derived membership functions through the quality of descriptive knowledge derived from the purposive sampling approach. The quality of the descriptive knowledge is dependent on the implementation of the purposive sampling approach such as the selection of the environmental variables, number of environmental clusters, the selection of classification algorithm and parameters which are discussed in another paper (Zhu et al., 2008).

In this paper the new method has been presented under the context of purposive sampling. We must point out that purposive sampling is one method for deriving the descriptive knowledge, mostly for areas with limited existing soil information. Descriptive knowledge on soil-environmental relationships can also be obtained from soil profile descriptions or soil type descriptions (such as soil series descriptions) obtained from conventional soil survey. Field soil survey experts can be another major source for such descriptive knowledge.

Our study area is located in the black soil region in Northeast China. The typical black soil is called Typic Hapli–Udic Isohumosols in the Chinese soil taxonomy system. Black soil is famous for its high fertility of soil, which is especially important for agricultural development. The total area of the Northeast black soil is about 1.02 million km^2 according to the national agriculture soil survey of China (Fan et al., 2004). The typical terrain of the black soil region is undulating mounds and the parent materials are mainly loess. From the soil forming environment point of view, our study area is very representative of the Northeast black soil region. From soil mapping perspective the proposed approach with purposive sampling strategy is potentially an effective way for soil survey in the region because there are no local soil experts nor detailed soil maps in the black soil region and obtaining membership functions from these sources is not an option. Coupling with purposive sampling which only requires few soil field points to derived the descriptive knowledge, the proposed approach not only improve the accuracy of predictive mapping but also save time and resources needed for extensive field work.

6. Conclusions

This paper presents an approach to the construction of fuzzy membership functions from descriptive knowledge generated by a purposive sampling approach which is suited for areas with no soil survey experts, no sample data, and no existing soil maps. From the results of a case study in Northeast China, we concluded that:

(1) The constructed fuzzy membership functions were able to produce good quality soil spatial information (soil type map and soil property map). Accuracy of the soil class map generated using fuzzy membership functions was at about 76%. The A-horizon soil organic matter content map produced from the fuzzy membership functions is at a better quality than that produced from a linear regression model which even uses more modeling points.
(2) Together with the purposive sampling technique, the proposed method provides an effective way to quantify knowledge on soil-environment relationships for predictive soil mapping, especially for those areas with limited data.

Acknowledgements

This study is supported by National Basic Research Program of China (No. 2007CB407207); Chinese Academy of Sciences International Partnership Project "Human Activities and Ecosystem Changes" (No. CXTD-Z2005-1); State Key Laboratory of Soil and Sustainable Agriculture (No. 0551000015); National Natural Science Foundation of China (No. 40501056); National Natural Science Foundation of China (No. 40601078); 'Hundred Talents' Program of Chinese Academy of Sciences; and field forefront program of LREIS. The authors wish to thank Prof. Du Guohua for his cooperation in this research as a soil classification expert. The authors would like to thank Dr. Budiman Minasny for his generous help with the revisions of this paper.

References

Band, L.E., Moore, I.D., 1995. Scale: landscape attributes and geographical information systems. Hydrological Processes 9, 401–422.
Beven, K.J., Kirkby, N.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 43–69.
Bui, E.N., Loughhead, A., Corner, R., 1999. Extracting soil–landscape rules from previous soil surveys. Australian Journal of Soil Research 37, 495–508.
Burrough, P.A., 1989. Fuzzy mathematics methods for soil survey and land evaluation. Journal of Soil Science 40, 447–492.
Burrough, P.A., 1996. Natural objects with indeterminate boundaries. In: Burrough, P.A., Frank, A.U. (Eds.), Geographic Objects with Indeterminate Boundaries. Francis and Taylor, London.
Burrough, P.A., MacMillian, R.A., van Deusen, W., 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. Journal of Soil Science 43, 193–210.

Chang, L., Burrough, P.A., 1987. Fuzzy reasoning: a new quantitative aid for land evaluation. Soil Survey and Land Evaluation 7, 69–80.
Chinese Soil Taxonomy Research Group, 2001. Keys to Chinese Soil Taxonomy, 3rd edition. University of Science and Technology of China Press, Hefei.
De Gruijter, J.J., McBratney, A.B., 1988. A modified fuzzy *k*-means method for predictive classification. In: Bock, H.H. (Ed.), Classification and Related Methods of Data Analysis. Elsvier, Amsterdam, pp. 97–104.
De Gruijter, J.J., Walvoort, D.J.J., Van Gaans, P.F.M., 1997. Continuous soil maps: a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77, 169–195.
Dobermann, A., Oberthur, T., 1997. Fuzzy mapping of soil fertility — a case study on irrigated riceland in the Philippines. Geoderma 77, 317–339.
Fan, H.M., Cai, Q.G., Wang, H.S., 2004. Condition of soil erosion in phaeozem region of Northeast China (In Chinese). Journal of Soil and Water Conservation 18, 66–70.
Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modeling and spatial prediction of soil attributes. International Journal of Geographical Information Systems 9, 421–432.
Hjerdt, K., McDonnell, J., Seibert, J., Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. Water Resources Research 40, W05602. doi:10.1029/2004WR003130.
Hudson, B.D., 1992. The soil survey as a paradigm-based science. Soil Science Society of America Journal 56, 836–841.
Lagacherie, P., 2005. An algorithm for fuzzy pattern matching to allocate soil individuals to pre-existing soil classes. Geoderma 128, 274–288.
Liu, J., Zhu, A.X., 2009. Mapping with words: a new approach to automated digital soil survey. International Journal of Intelligent Systems 24, 293–311.
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., de Gruijter, J.J., Brus, D.J., 1992. Spatial prediction and mapping of continuous soil classes. Geoderma 54, 39–64.
McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97, 293–327.
McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57, 443-452.
Peterson, C., 1991. Precision GPS navigation for improving agricultural productivity. GPS World 2, 38–44.
Qi, F., Zhu, A.X., 2003. Knowledge discovery from soil maps using inductive learning. International Journal of Geographical Information Science 17, 771–795.
Qi, F., Zhu, A.X., Harrower, M., Burt, J.E., 2006. Fuzzy soil mapping based on prototype category theory. Geoderma 136, 774–787.
Qi, F., Zhu, A.-X., Pei, T., Qin, C., Burt, J.E., 2008. Knowledge discovery from are-class resource maps: capturing prototype effects. Cartography and Geographic Information Science 35, 223–237.
Qin, C.Z., Li, B.L., Zhu, A.X., Yang, L., Pei, T., Zhou, C.H., 2006. Multiple flow direction algorithm with flow partition scheme based on downslope gradient (in Chinese). Advances in Water Science 17, 450–456.
Shi, X., Zhu, A.X., Burt, J.E., Qi, F., Simonson, D., 2004. A case-based reasoning approach to fuzzy soil mapping. Soil Science Society of America Journal 68, 885–894.
Willmott, C.J., 1984. On the evaluation of model performances in physical geography. In: Gaile, G.L., Willmott, C.J. (Eds.), Spatial Statistics and Models. D. Reidel Publi., Dordrecht, the Netherlands, pp. 43–460.
Yang, L., Zhu, A.X., Li, B.L., Qin, C.Z., Pei, T., Liu, B.Y., Li, R.K., Cai, Q.G., 2007. Extraction of knowledge about soil-environment relationship for soil mapping using fuzzy *c*-means (FCM) clustering. Acta Pedologica Sinica 44, 16–23.
Zhang, B., Zhang, Y., Chen, D., White, R.E., Li, Y., 2004. A quantitative evaluation system of soil productivity for intensive agriculture in China. Geoderma 123, 319–331.
Zhu, A.X., 1997. A similarity model for representing soil spatial information. Geoderma 77, 217–242.
Zhu, A.X., 1999. A personal construct-based knowledge acquisition process for natural resource mapping. International Journal of Geographical Information Science 13, 119–141.
Zhu, A.X., Band, L.E., 1994. A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing 20, 408–418.
Zhu, A.X., Band, L.E., Vertessy, R., Dutton, B., 1997. Derivation of soil properties using a soil land inference model (SoLIM). Soil Science Society of America Journal 61, 523–533.
Zhu, A.X., Mackay, D.S., 2001. Effects of spatial detail of soil information on watershed modeling. Journal of Hydrology 248, 54–77.
Zhu, A.X., Band, L.E., Dutton, B., Nimlos, T., 1996. Automated soil inference under fuzzy logic. Ecological Modelling 90, 123–145.
Zhu, A.X., Hudson, B., Burt, J.E., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal 65, 1463–1472.
Zhu, A.X., Yang, L., Li, B.L., Qin, C.Z., English, E., Burt, J.E., Zhou, C.H., 2008. Purposive sampling for digital soil mapping for areas with limited data. In: Hartemink, A.E., McBratney, A.B., Mendonca Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. InSpringer-Verlag, New York, pp. 233–245.