Computers & Geosciences 37 (2011) 1425–1436

Contents lists available at ScienceDirect

Computers & Geosciences

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

Comparing three methods for modeling the uncertainty in knowledge discovery from area-class soil maps

Feng Qi [a,*], A-Xing Zhu [b,c]

[a] Department of Geology and Meteorology, Kean University, 1000 Morris Ave. Union, NJ 07083, USA
[b] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences Building 917, Datun Road, An Wai, Beijing 100101, China
[c] Department of Geography, University of Wisconsin-Madison, 550 North Park Street, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 2 July 2010
Received in revised form
18 October 2010
Accepted 20 October 2010
Available online 10 November 2010

Keywords:
Knowledge discovery
Uncertainty
Fuzzy
Prototype theory
Soil classification

ABSTRACT

Knowledge discovery has been demonstrated as an effective approach to extracting knowledge from existing data sources for soil classification and mapping. Soils are spatial entities with fuzzy boundaries. Our study focuses on the uncertainty associated with class assignments when classifying such entities. We first present a framework of knowledge representation for categorizing spatial entities with fuzzy boundaries. Three knowledge discovery methods are discussed next for extracting knowledge from data sources. The methods were designed to maintain information for modeling the uncertainties associated with class assignments when using the extracted knowledge for classification. In a case study of knowledge discovery from an area-class soil map, all three methods were able to extract knowledge embedded in the map to classify soils at accuracies comparable to that of the original map. The methods were also able to capture membership gradations and helped to identify transitional zones and areas of potential problems on the source map when measures of uncertainties were mapped. Among the three methods compared, a fuzzy decision tree approach demonstrated the best performance in modeling the transitions between soil prototypes.

1. Introduction

Soil is a fundamental natural resource. In the United States and many other countries of the world, the spatial distribution of soils is routinely collected, modeled, and archived in inventories during soil surveys. Previous research has indicated that valuable knowledge was embedded in the archived soil maps and such knowledge could be revealed through knowledge discovery (Moran and Bui, 2002; Qi and Zhu, 2003). The knowledge is about the relationships between soil classes and the underlying environmental conditions. Such knowledge can be used for soil classification and mapping during soil survey updates, when information on the environmental conditions of an area is available at greater detail or higher accuracies.

Largely influenced by nineteenth century biological taxonomy and geological survey (Scull et al., 2003), the current soil survey system defines discrete soil classes under crisp logic, which represent only a limited number of modal soil profiles without being able to capture the full amount of soil variability (Campbell and Edmonds, 1984). The resulting product is usually an area-class map, on which locations within each area share a membership of unity to a soil category. However, soils, like many other geospatial entities, have fuzzy boundaries in both attribute space and geographic space (Burrough, 1996; Greve and Greve, 2004). Fitting such entities into discrete categories with crisp boundaries induces uncertainties in the class assignments (Burrough, 1996).

Previous studies have developed models and visualization methods to conceptualize, measure, represent, and present the uncertainties associated with the classification of soils and other fuzzy geographic entities (Fisher, 1994; Goodchild et al., 1994; Ehlschlaeger et al., 1997; Davis and Keller, 1997; Van Der Wel et al., 1998; Fisher et al., 2005). As knowledge discovery and data mining has been recognized as an effective approach to extract and apply knowledge for clustering or classifying geospatial entities in recent years (Canty, 2009; Zhang et al., 2005; and see Miller and Han, 2001 for reviews of earlier studies), the issue of uncertainty in this context is worth exploring. This paper first reviews the conceptualization of errors and uncertainties in geographic classification and examines the existence and nature of uncertainties in soil classification, based on which we introduce a knowledge representation for classifying spatial categories in a way that explicitly reflects the fuzzy class boundaries and membership gradations within classes. We then discuss three data mining methods that capture the uncertainties associated with class assignments during

* Corresponding author. Tel.: +1 908 737 3668; fax: +1 908 737 3699.
E-mail addresses: fqi@kean.edu (F. Qi), azhu@wisc.edu (AX. Zhu).

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2010.10.016

the knowledge discovery process. The methods are finally illustrated and compared through a case study on updating the soil classification in a watershed in Wisconsin, USA.

2. Uncertainties in soil classification

Uncertainty is common in the representation and processing of geographic information. Uncertainty exists because we either have imperfect understanding of the geographic phenomena we study or we have imperfect data to study it (Harrower, 2003). Specific errors and uncertainties in categorical maps have been well studied by GIScientists. Goodchild et al. (1992) identified two major types of errors that cause uncertainties in categorical maps: inclusions and generalizations of transition zones. Ehlschlaeger and Goodchild (1994) added two additional forms of mapping errors: incorrect labels and misplacement of class boundaries (assuming that an objective boundary exists). Fisher (1994, 2010) and colleagues (Fisher et al., 2005) established a formal conceptualization of the types of uncertainty in geographic information based on three causes: (1) generalization (or conceptualization) of the spatial phenomena leads to "ambiguity" (often due to the discord of different classification schemes used); (2) poorly defined classes not being able to capture the natural gradations of geographic properties leads to inherent "vagueness" of the classes; and (3) "errors" in measurement or any steps in the classification and mapping process leads to inevitable uncertainties in the mapping product.

Soil maps, the product of soil classification, are one kind of such categorical maps that are prone to uncertainties. Soil maps are created by soil experts to capture the spatial variation of soils using a limited number of soil classes. The delineation of soil polygons are based on expert knowledge of the local soil-environment relationship, a process that is known as predictive soil mapping (Scull et al., 2003). Qi (2004) studied errors present on soil maps resulting from the predictive soil mapping process and differentiated two kinds of errors. The first kind comes from the human classification model used to classify local soils and is referred to as the modeling error. Modeling errors can include: (1) generalization of the continuous soil variation as discrete categories, thus leading to an oversimplification of transitional zones as lines; and (2) inclusions (or mixing of soil classes) that cannot be eliminated by increasing map scale, that is, the model cannot discriminate between two soil classes using the current classification model or available environmental factors. The second kind of errors is introduced in the mapping process, and is referred to as the mapping error. Major mapping errors include: (1) misplacement of class boundaries; (2) mislabeling of polygons; and (3) inclusions that can be avoided by increasing map scale, that is, the missing of small patches of classes due to the limitation of map scale. These mapping errors lead to the uncertainty caused by "error" as identified by Fisher et al. (2005). The modeling errors, on the other hand, should correspond mostly to the "vagueness" uncertainty but also explains the "ambiguity" uncertainty in that conceptualization.

Our discussion here focuses on the modeling error and examines the nature of such error on the basis of cognitive psychology, one of whose central concerns have been the process of categorization (Tversky and Hemenway, 1984). The classical view of categories supported Boolean categorization, that is, an instance either is in or out of a category completely. As long-recognized that such classical logic is invalid in dealing with many natural categories in reality, new category theories were introduced in the 1970s (Smith and Medin, 1981). Emerged from Wittgenstein's earlier ideas on family resemblance, centrality and membership gradience (Wittgenstein, 1953) and Zadeh's fuzzy set theory (Zadeh 1965), prototype theory (Rosch, 1973, 1978) stresses the fact that category membership is not always homogenous and that some members are better representatives of a category than others, which is noted as the "prototype effects" (Lakoff 1987).

Prototype effects are present in many geographic categories that are generalized from the natural environment. For example, in the case of soil classification, the continuous soil body is categorized into soil classes. Prototype effects are shown when soil scientists think a certain pedon is more representative of a soil class than another one, although both are classified as belonging to the same class. And in the case of landform classification, the natural landscape is commonly seen as hills and valleys or summits, shoulders, back slopes, foot slopes, and drainage ways on a larger scale. Prototype effects are reflected such that a particular foot slope location may be perceived to be a better example of a foot slope than others. Plewe (2002) named these geographic categories "motivated entities", which are conceptual phenomena created from more complex phenomena by processes of simplification. Classification models used in these processes create simplifications of the often infinitely complex reality, and at the same time impose more order on real phenomena than is probably inherently there. This process results in discrepancies that lead to inherent degree of category memberships, which is one of the three major causes of prototype effects generalized by Lakoff (1987). And this inherent uncertainty is sometimes termed indeterminacy (Burrough, 1996) that corresponds to the vagueness uncertainty in Fisher's conceptualization (Fisher et al., 2005).

On the other hand, as highly simplified models of the rich detail in the complex reality, the definitions of these "motivated entities" cannot be perfectly reconciled, even conceptually (Plewe, 2002). The extent of a motivated entity can thus be ambiguous (when more than one realization is present and valid). As generalized by Lakoff (1987), the second major cause of prototype effect is the existence of cluster models of a category's cognitive structure that evolved and developed in different communities. In geographic classification, the objects resulting from classifying continuous geographic features often have no objective appearance, shape, and boundaries; rather, their perception depends upon the cognitive categories the culture of a community imposes on its members (Ferrari, 1996). Soil surveyors, for example, are such a community. After decades of experiments, discussions and research, people in this community reached a resolution on how to categorize soil bodies. Soil taxonomy evolved along with the process and was eventually used as the basis of soil classification and soil resource inventory. As have been noticed by previous studies classification systems established in such a way could be highly inconsistent across countries (Fisher et al., 2005) or even within a single country. Different soil experts with different regional experiences might well develop two different sets of soil maps given the same amount of information and field work in the area (Hodza, 2010). This discord of the perceptual structures that exist implicitly in soil experts' mind as their tacit knowledge (Shi et al., 2004) is one other cause of the prototype effects and explains the "ambiguity" uncertainty (Fisher et al., 2005).

As considerable research has been conducted on the conceptualization of uncertainties in geographic classification, the various conceptualization models developed do not exclude each other but emphasize different stages of the process for conceptualizing and classifying geographic entities. Although one might wish for a single framework for the representation and modeling of uncertainty in applications, it has been the case that uncertainty modeling is highly application-oriented (Zimmermann, 2000) and very much specific in different domains. In the case of modeling uncertainty associated with categorical maps such as soil maps, the two types of uncertainties caused by modeling errors and mapping errors (Qi and Zhu, 2003) have been studied using

different approaches. Mapping errors and the associated uncertainties are commonly modeled with stochastic methods. Fischer (1995) used stochastic methods to simulate the inclusion error and presented it using animation. Bivariate statistics were used to model the accuracy of planar spatial objects (Shi, 1998; Leugn and Yan, 1998). Qi and Zhu (2003) used probabilistic distribution of soil pixels to direct sampling inside soil polygons to reduce mapping errors in a study on knowledge discovery from soil maps. The uncertainty associated with modeling errors, on the other hand, has been commonly studied using fuzzy logic (Davidson et al., 1994; Burrough, 1996; Davis and Keller, 1997). Some studies have also investigated multi-valued logic (such as rough set theory as by Ola et al., 2000; Worboys, 2001) and supervaluation semantics (Bennett, 2001).

In this study we focus on the uncertainty resulting from assigning an instance to its class prototype in an area-class soil map. Our hope is to provide practical implications for guided use of the traditional products soil classification by modeling and presenting such uncertainties in a quantitative manner. Such uncertainty mostly corresponds to the "vagueness" uncertainty in previous conceptualizations (Fisher et al., 2005) but also is intrinsically related to the "ambiguity" uncertainty that is another cause of the prototype effects, as discussed above.

The generalization of the continuous soil body to discrete polygons overlooks the prototype effects of the classes. Fig. 1 illustrates the positions of two instances I_1 and I_2 both classified as class B. In the classification space that's defined by two environmental features X and Y to separate the classes, neither I_1 nor I_2 is the actual prototype of class B. First of all, by assigning these instances to class B and having them bear the properties of the prototype, we exaggerate the similarities between these instances and those of the class prototype and thus introduce an exaggeration uncertainty. On the other hand, during class assignment, we also ignore the fact that I_1 may bear some similarity to the prototypes of classes A and C, as does I_2 to some degree. This leads to an ignorance uncertainty (Zhu, 1997a). Such uncertainties could be approximated if the instances memberships to the different classes are estimated during soil classification. We illustrate here with a case study on knowledge discovery from soil maps the modeling of such uncertainties through explicit quantification of the membership gradations within the mapped classes and positioning an instance in relevance to its class prototype.

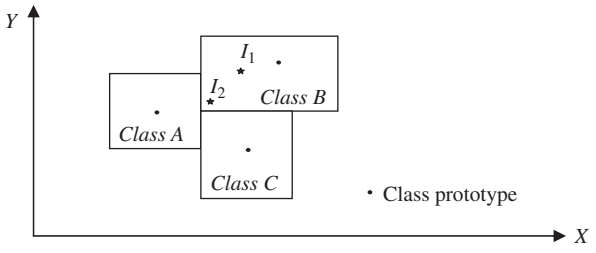

Fig. 1. Class assignments for instances I_1 and I_2.

3. Knowledge discovery for classifying entities with fuzzy boundaries

3.1. Knowledge representation

As different representation models are believed to be suitable for different cognitive tasks (Markman, 1999), featural models are often used for categorization. With a featural model, a category is represented by a composite set of properties (features). Based on prototype theory, such features should summarize the real instances of the category which serve as the cognitive reference points for inference (Minda and Smith, 2001).

Our knowledge representation scheme is based on the widely adopted knowledge representation of categories using the featural model in the form of 'frames' (Fillmore, 1985). In order for the knowledge to incorporate the prototypical properties of classes, we explicitly model the prototypes and membership gradations in the knowledge representation. The prototypes of classes are stored using a common frame structure while the membership gradations are represented with optimality functions (Zhu, 1999). Furthermore, spatial relationships can be represented with inter-frame links. Fig. 2 shows an example of such a representation for soil classes. The features that define the prototypes of a soil class are listed in the frames, and each feature also points to an optimality function that describes how membership responds when the value of the feature changes. Specifically, if the value of a feature corresponds to an optimality value of 1, the possession of such a feature will most probably lead to full membership in the class. On

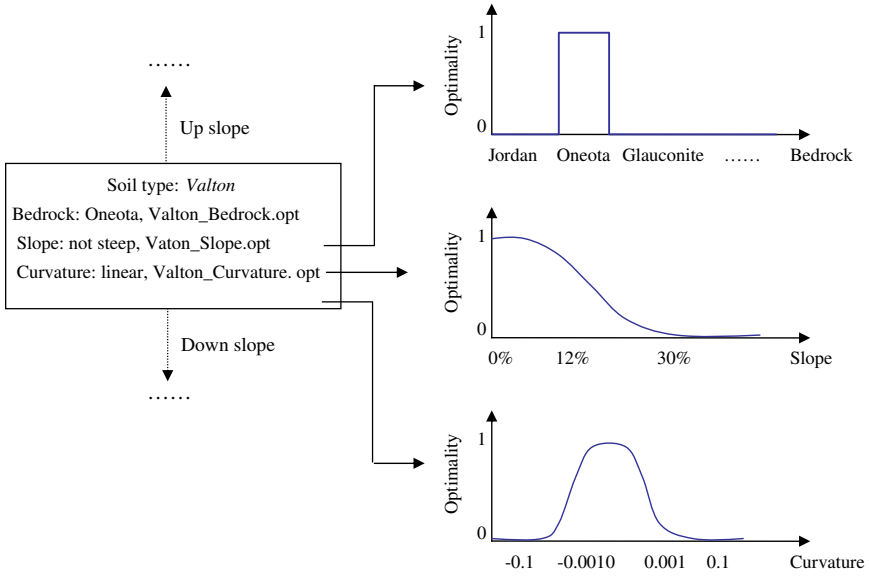

Fig. 2. Frame representation of knowledge for classifying four soil classes.

the other hand, an optimality of 0 means that the corresponding feature value does not favor the class at all. For soil class *Valton* which only occurs on a particular bedrock *Oneota*, (see Fig. 2), its membership drops immediately from 1 to 0 when the bedrock changes from *Oneota* to any other type. Fig. 2 also shows that steeper slopes limit the development of *Valton* and linear curvatures provide an optimistic condition for the soil.

3.2. Knowledge discovery

Knowledge represented in such a way can be obtained through knowledge engineering by interviewing domain experts as detailed in our earlier work (Qi et al., 2006). The current study focuses on an alternative approach to obtaining the knowledge embedded in existing data sources. Previous studies (Canty, 2009; Zhang et al., 2005; and see Miller and Han, 2001 for reviews of earlier studies) have demonstrated the robustness of knowledge discovery methods in clustering or classifying geospatial entities. The scenario is that when a large amount of data is available on classified examples and their related features, the knowledge of prototypes and membership gradations can be obtained empirically through knowledge discovery.

Our previous work has proposed two methods for knowledge discovery from soil maps. This study compares the efficiencies of these methods with a third one in the context of modeling uncertainties during soil classification from providing useful information on the use of the classification products. The three methods to be compared include one that is based on an empirical measure of the typicality of category members that were first introduced by psychologists Rosch and Mervis (1975) as the family resemblance measure (referred to as the *family resemblance approach*). The second builds on our previous results of decision tree induction (Qi and Zhu, 2003) and uses a Semantic Import Model (SI) (Burrough, 1989) to post-fuzzify the decision tree (referred to as the *fuzzy decision tree approach*). The third approach was inspired by a cognitive model that accounts for prototype effects and uses recent techniques in ensemble training in machine learning (referred to as the *ensemble approach*):

1) The family resemblance approach.
   The family resemblance measure developed by Rosch and Mervis (1975) quantifies an instance's typicality based on a sum of features that an instance possesses, weighted by how many other category members also possess them. An item is a typical member of a concept if it contains features shared by many other members of the same concept. In Rosch and Mervis' experiments, the features are binary features (absence or presence of an attribute) only. Qi et al. (2008) extended the concept to multi-valued and continuously valued features. Specifically, a set of modal values of the defining features define the class prototype. Frequency distribution curves of these features then model the membership gradations (optimality functions in the knowledge representation illustrated in Fig. 2). To derive optimality functions, histograms of all features are constructed from training samples of each individual class, and a smooth curve is then fitted to the histograms. We used the approach developed by Qi et al. (2008) to obtaining a holistic yet realistic representation of the often skewed optimality curve based on the Expectation-Maximization (EM) method (Dempster et al., 1977).
2) The fuzzy decision tree approach.
   The second method builds upon the results from a previous study on extracting knowledge from soil maps in the form of decision trees (Qi and Zhu, 2003). This study demonstrated that a traditional decision tree can be trained using preprocessed data to approximate the central concepts (prototypes) of the mapped soil classes with considerable accuracy. In order to account for the membership gradations in terms of optimality functions, we employed a post-fuzzification method (Chiang and Hsu, 2002) that uses SI-based fuzzy membership functions (FMFs) (Burrough, 1989) in the current study. Specifically, a decision tree is first trained using preprocessed data to capture the class prototypes. Optimality curves are then modeled as FMF curves during post-fuzzification. The shape of an FMF curve is determined by the number of bounds set to a variable in the decision tree, and the cross-over points (Burrough, 1989) of the FMF curve are determined by the break values in the decision tree. For example, if the values of a defining variable have two bounds (e.g. elevation $\in$[800, 1200]), the FMF curve is bell-shaped with the two cross-over points at 800 and 1200; if they are only bounded in one direction (e.g. elevation $<$600 ft, or slope $>$12%), the curve could be S-shaped or reverse S-shaped. Correspondingly, the single cross-over point needed is set to be at the break value (e.g. 600 ft, or 12%).
3) The ensemble approach.
   The third method is based on the machine learning strategy known as ensemble learning (Dietterich, 1997). In ensemble learning, many classifiers (e.g. decision trees) instead of one are trained and inference is achieved by letting the entire set of classifiers vote. Originally designed to increase classification accuracies (Canty, 2009), this strategy of learning and inference actually corresponds to one of the cognitive models that causes prototype effects in the resulting classes: the cluster model as defined by Lakoff (1987). With such a model, the categorization of an instance is based on the composite of categorization outputs from a set of individual cognitive structures.

Various approaches to construct ensembles have been investigated by researchers in the machine learning community. Our previous studies (Qi and Zhu, 2006) indicated that *AdaBoosting* (Freund and Schapire, 1996) is an effective method for modeling the fuzzy boundaries of soil classes when used to train multiple decision trees from soil samples and this method was thus used in our current study to compare with the two other methods mentioned above.

3.3. Inference and uncertainty modeling

Qi et al., 2006 used prototype-based inference to infer the spatial distributions of soils and their properties with knowledge provided by soil experts but represented in the same scheme as discussed in this paper. With prototype-based inference, the features of an instance to be classified are compared to the class prototypes, and the similarities based on all defining features are then combined using a fuzzy AND operator (Zadeh, 1965) to generate the overall similarity (membership) of the instance to the class prototype. Every instance is then associated with a set of membership values to all prescribed classes. We used the similarity model (Zhu, 1997b) to represent the spatial distribution of all class memberships. An instance at pixel location (i, j) is represented as an n-dimensional similarity vector, $S_{ij}=(S_{ij}^1, S_{ij}^2, \ldots, S_{ij}^k, \ldots, S_{ij}^n)$, where S_{ij}^k represents the similarity value or fuzzy membership of the instance to category k, and n is the total number of the prescribed categories.

As discussed in Section 2, two aspects of uncertainty are associated with the oversimplification of soil categories to their prototypes: the ignorance of individual instances' similarity to prototypes of other soil classes and the exaggeration of members' similarity to their own class prototypes. The ignorance uncertainty is clearly related to membership diffusion in the similarity vector in

our inference result. The more concentrated the membership in a particular class, the smaller the uncertainty. If a location has a high similarity to one single soil class but very low similarities to others, the classification of it to the dominant soil class will lead to a low ignorance uncertainty. The ignorance uncertainty can thus be approximated using an entropy measure (Goodchild et al., 1994; Zhu, 1997a)

$$U_{ij} = \frac{1}{\ln N} \sum_{k=1}^{N} (S_{ij}^k \ln S_{ij}^k) \quad (1)$$

where U_{ij} is the estimated ignorance uncertainty, S_{ij}^k is the similarity value of the instance at pixel (i, j) to category k, and N is the number of categories that the instance has similarity to. When soil at (i, j) has full membership to only one category, U_{ij} will obtain the value 0 meaning no ignorance uncertainty is in question. The highest U_{ij} at the value of 1, on the other hand, indicates that the soil is evenly similar to all categories, and that assigning the instance to any one of the categories would involve the greatest degree of ignorance uncertainty.

The uncertainty associated with the exaggeration of members' similarity to their own class prototypes is inversely related to the saturation of its membership to the assigned category. If an instance has full membership to a class or complete similarity to the class prototype, there should be no exaggeration for the instance to be categorized to that class. And the lower the membership of the instance to the assigned class (to which the similarity is already the highest among all soil classes), the greater is the exaggeration. The exaggeration uncertainty can thus be approximated with (Zhu, 1997a)

$$E_{ij} = 1 - S_{ij}^a \quad (2)$$

where E_{ij} is the estimated exaggeration uncertainty and S_i^a is the similarity value of the instance at pixel (i,j) to its assigned category (a).

4. Case study: knowledge discovery for soil classification

Digital soil mapping methods have been developed to map the gradations of soil memberships and the related classification uncertainties through knowledge-based approaches (Zhu, 1999; Shi et al., 2004; Qi et al., 2006). The knowledge used in soil inference in these previous studies was obtained directly from soil experts through either knowledge engineering (Zhu, 1999; Qi et al., 2006) or case-based reasoning (Shi et al., 2004). When experienced soil expert is not available an alternative approach is to obtain similar knowledge through data mining from large amount of classified examples and their related features. It is often the case, however, in soil mapping and the mapping of many other natural resources that large amounts of field samples are not available as classified examples to be used for knowledge discovery. Existing inventory maps thus provide an alternative source for such classified examples; each location enclosed within a polygon is an example of the class indicated by the polygon label. Using these classified examples, previous studies have extracted relationships between the mapped soil classes and their environmental conditions through data mining (Moran and Bui, 2002; Qi and Zhu, 2003). We used a similar soil map in the current study to compare our data mining methods for capturing membership gradations and modeling uncertainties.

4.1. Data and method

The soil map we used was created from a recent soil survey that mapped 16 soil series in the area (see Fig. 3). The soil-formative environmental conditions were captured with a GIS database as detailed in our previous study (Qi and Zhu 2003). Environmental data layers such as bedrock geology and topography as well as spatial data layers such as spatial neighbors are included in the database. The soil map was overlaid with the environmental data layers to create the set of classified examples with each pixel being labeled with the soil series name and associated with the values of all environmental variables.

With the family resemblance method, histograms were constructed for the pixels in a particular class based on every individual feature. The histogram can be either unimodal or bimodal. Data preprocessing was conducted to detect bimodal cases through human visualization. When a bimodal case was detected, the two modes were regarded as two prototypes, and the two prototypes were separated following Qi et al. (2008).

For the fuzzy decision tree approach, data preprocessing was conducted before the decision tree training through histogram sampling following Qi and Zhu (2003). Specifically, only the pixels that fall close to histogram modes of individual environmental features for each soil class get selected. This reduces the impact of possible errors on the original map and captures more accurately the prototype of each mapped soil class (Qi and Zhu, 2003). See5 program[1] was used to train a decision tree that represents the prototypes of the mapped soil classes. And optimality curves were modeled as Gaussian FMF curves on the basis of the decision tree output.

The same data preprocessing strategy employed for the fuzzy decision tree approach was also applied to the ensemble approach, since the ensemble training is basically iterations of decision tree training. With the ensemble approach, preprocessed training examples were fed to the *AdaBoosting* (Freund and Schapire, 1996) algorithm on the See5 platform to train ensembles of decision trees.

4.2. Inference results

Knowledge extracted with all three methods was used for soil inference for comparisons. A soil series map was eventually created with each method by assigning each pixel the soil series with the highest similarity score in the similarity vector. Fig. 4 shows the three defuzzified soil series maps. All three inferred maps exhibit spatial patterns of the 16 soil series that resemble that shown on the original map (Fig. 3) from which the knowledge was extracted. Soil series that occupy the most area appear on similar landscape positions to those on the original map. Minor differences exist where slightly different spatial extents (soil series *Orion*, for example) are observable for a few soil series. It indicates that the extracted knowledge with the data mining methods was able to capture the major characteristics of the soil series in the feature space, which is then reflected in the configurations in the physical space.

In order to evaluate and compare the inference results (and thus the extracted knowledge with the three methods), 99 field samples were collected from the watershed and classified by experienced soil scientists from the local soil survey agency. In terms of classification accuracy at the soil series level, the original map correctly classified soils at 83 out of the 99 sites, while the inferred soil series maps named 80, 77, and 83 sites correctly, respectively. The inferred maps misclassified zero (the ensemble approach) to six (the fuzzy decision tree approach) sites that were correctly mapped by the original map. This echoes the visual comparison of the inferred maps with the original map, indicating that the extracted knowledge with all three approaches is able to capture

[1] Data Mining Tools See5 and C5.0. http://www.rulequest.com/see5-info.html.

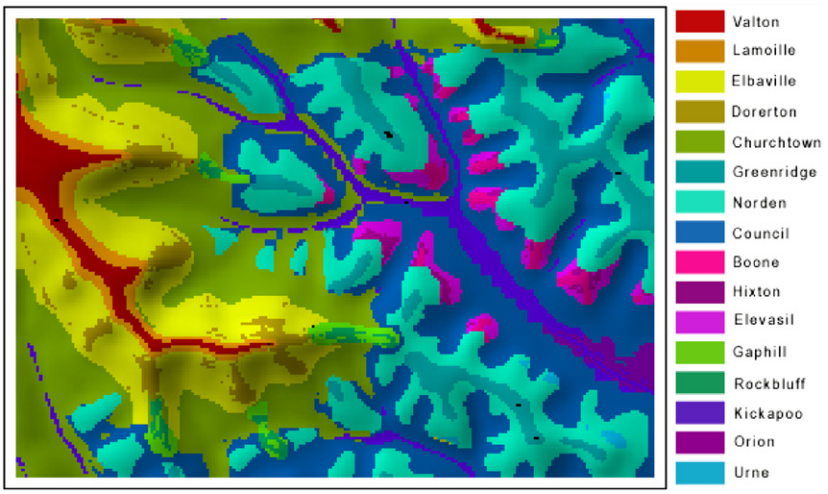

Fig. 3. Soil Series map of Raffelson watershed.

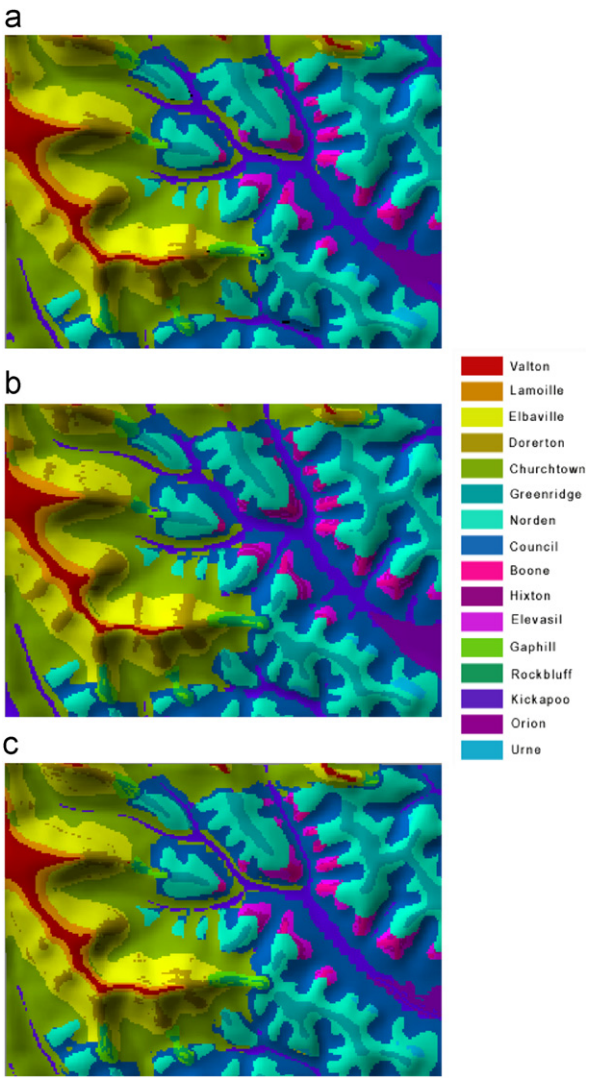

Fig. 4. Soil Series maps inferred with (a) family resemblance approach, (b) fuzzy decision tree approach, and (c) ensemble approach.

the major patterns of soil distribution at the soil series level over the mapped area to a considerable degree of accuracy, with the best being the ensemble approach.

4.3. Uncertainty analysis

Uncertainties associated with assigning specific soil series to a soil pixel were computed from the inferred similarity vectors following Eqs. (1) and (2). Fig. 5 shows the distribution of the ignorance uncertainty (left) and exaggeration uncertainty (right) with all three approaches. We see that the corresponding maps show different overall tones and varying patterns among the three approaches, indicating different uncertainty levels of the classifications. The maps, however, do exhibit some dominant spatial patterns that are shared by all three sets of maps: transitional zones between soil prototypes and the lower valleys are associated with the highest uncertainty while typical landscape positions of a soil series show relatively low uncertainty values.

(1) Ignorance uncertainty.
The areas marked as A in Fig. 5 are examples of high ignorance uncertainty on slope shoulder positions that are transitional zones between typical ridges and backslopes. The soils developed on such transitional zones bear similarities to both their upslope and downslope neighboring prototypes but are not fully qualified for either. Similarly, we see high uncertainty levels mapped in transitional zones between bedrock-controlled soils and colluvium-based soils in area B. Another observation from Fig. 5a is the exceptionally high uncertainty of small patches in the middle of the watershed (area C, particularly obvious for the family resemblance approach and ensemble approach). It turns out that these patches are on a unique bedrock that takes up very limited area where three different soil types were developed. The crowding of three soil types in space makes it difficult to separate the soil classes using available landscape characteristics. Mixture of similarities to all three types is thus inevitable.

(2) Exaggeration uncertainty.
Similar to what was seen on the ignorance uncertainty maps, high uncertainty is also dominant along boundaries of soil bodies on the exaggeration uncertainty maps. The reason is that soils in transitional zones bear similarities to multiple soil classes but similarity is not high in any one of the classes. We also notice the very high exaggeration uncertainty in the low valleys of the watershed (area D). This indicates that the distribution of soils in the wide flat low land here bear low similarities to even the soil class they are most similar to. This can happen in two circumstances. In the first, the soil prototypes and their distributions in this area are not well captured

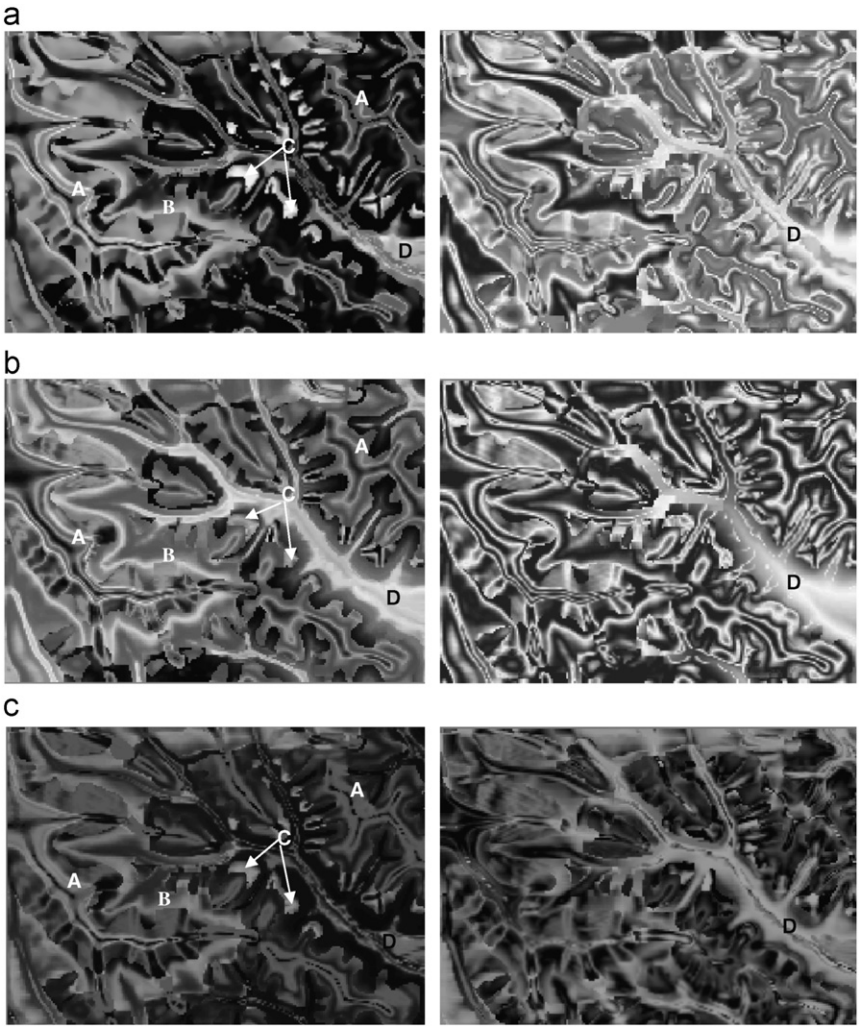

Fig. 5. Distribution of ignorance uncertainty (left) and exaggeration uncertainty (right) with the (a) family resemblance approach, (b) fuzzy decision tree approach, and (c) ensemble approach (light tones indicating high uncertainty values).

using the environmental variables included in our GIS database and thus the knowledge extracted using our data mining methods is not accurate enough. In the second, the soil classes mapped on the original map in this area are not adequate to capture the soil prototypes in the area, and the original map is not sufficient in portraying the real soil variations in the valley. In this study, the latter would be a more proper explanation, since an examination of the misclassified samples by the original soil map among the 99 field samples reveals that nearly half of the misclassifications occurred in this low elevation area. Thus the exaggeration uncertainty map communicates some information on the accuracy of the source map itself.

As shown above, the uncertainty values provide a measure of prototype effects of the classified soils and sometimes reveal information on the quality of the knowledge obtained through knowledge discovery. They help highlight areas where managerial decisions should be adjusted when using the extracted knowledge. For example, while the transitional areas in the watershed are all classified as certain soil series, they should not be treated the same as the prototypes of the soil series because of the high uncertainty associated with class assignments. The implication is that other managerial measures may have to be applied in using the soil resource in these areas. The example in the Raffelson watershed also shows that not only the knowledge extracted through knowledge discovery not always capture well the soil distribution but also that the original soil map may not be accurate in certain areas as indicated by the high exaggeration uncertainty in the low valleys. Such information could also offer a cautionary lesson in future use of the extracted knowledge as well as the original data source in these areas.

4.4. Comparison of the three approaches

A major difference we could observe from the comparison of the three sets of maps in Fig. 5 is that the family resemblance approach exhibits an overall lower level of ignorance uncertainty than the fuzzy decision tree approach but an overall higher level of exaggeration uncertainty. The ensemble approach shows the lowest among the three with both uncertainties. In addition to the visual differences of the uncertainty maps, we computed the uncertainty measures at the 99 locations where we collected field samples. Some of the sample points are on prototype locations with low uncertainties but some are clearly located in transitional zones where uncertainties are expected to be high. Table 1 lists the mean uncertainty measures computed for sample locations based on the three sets of uncertainty maps. It also indicates that the fuzzy decision tree approach has resulted in an overall higher uncertainty level while the ensemble approach is associated with the lowest. By

a quick examination of the optimality curves derived with the family resemblance and fuzzy decision tree methods, we found that the family resemblance approach tends to generate narrower curves than the fuzzy decision tree approach. This led to less extensive overlapping between adjacent soil series in the classified map and thus lower ignorance uncertainty. At the same time, similarity values are often computed to be higher with the wider optimality curves derived with the fuzzy decision tree approach, thus lower exaggeration uncertainties. With the ensemble approach, no optimality curves are involved. Fuzzy inference was achieved by letting the decision trees in the ensemble vote. In our current study, only 100 decision trees were trained in the ensemble (limited by software). We know that the less decision trees in the ensemble, the more the inference should be similar to a traditional classification based on crisp logic. The average uncertainty measures being fairly close to 0 with this approach (Table 1) shows that the one hundred iterations are not enough for capturing the fuzzy transitions in our study area and that the classification is only a slight deviation from the traditional crisp classification.

The uncertainty measures computed using the similarity values can be used to identify transitions between soil classes because the similarity vectors contain information on the full range of membership of the local soil to all potential soil classes. We thus used these similarity vectors to map soil properties in a fashion which captures the gradual change between class prototypes. In our case study, continuous soil property maps of percentage of sand and silt in the *A* horizon were generated for all three methods following our earlier work (Qi et al., 2006), in which the soil property at a particular location was calculated as the weighted average of the prototypical properties of the soil classes in the area, with the weights being the inferred similarity values to the soil classes. Property maps were also derived from the original map by assigning each pixel the typical property values recorded for the labeled soil series.

Forty nine field samples were sent to the National Soil Survey Center at Lincoln, NE for soil property analysis. The percentages of sand and silt in the *A* horizon of the samples were determined and then compared with those obtained from the inferred maps and evaluated using three indices: MAE, RMSE, and agreement coefficient (AC) (Willmott, 1984). The range of AC values is between 0 and 1, with 1 indicating perfect agreement and 0 meaning complete disagreement between the estimated and observed values (Willmott, 1984).

Table 2 lists the computed statistics from the original map and three inferred maps. , It shows that the error rates of the inference results are lower than those of the original soil map. Higher AC for the inference results also implies a better performance of the inferred property maps on estimating the selected continuous soil properties. This should be attributed to their ability to capture the transitions between soil prototypes, especially when the inferred maps' accuracy in predicting soil series names is even lower than that of the original map (except for the ensemble approach). Comparing the three methods, we see that the error rates of the texture maps derived with the fuzzy decision tree approach are notably lower than those of the family resemblance approach, despite the fact that the crisp classification performance at the soil series level with the fuzzy decision tree approach (77.8%) is actually worse than that with the family resemblance approach (83.8%). On the other hand, the RMSE rates of the ensemble approach also appear slightly higher than those of the fuzzy decision tree approach although the classification accuracy at the soil series level with the ensemble approach is much higher. Both pieces of evidence indicate that the soil texture maps inferred with the fuzzy decision tree approach might better capture the continuous soil properties even given a lower accuracy at the soil series level.

In order to further examine whether the better performance of the fuzzy decision tree approach over the family resemblance

Table 1
Average entropy and exaggeration uncertainty at the 99 sample locations based on the three data mining methods.

	Family resemblance	Fuzzy d tree	ensemble
Ignorance uncertainty	0.08	0.25	0.04
Exaggeration uncertainty	0.63	0.17	0.05

Table 3
Accuracies of the derived *A* horizon textures: the family resemblance approach *vs.* the fuzzy decision tree approach on transitional and non-transitional sets based on uncertainty measures calculated with the family resemblance approach.

	Percentage of sand			*Percentage of silt*		
	MAE	**RMSE**	**AC**	**MAE**	**RMSE**	**AC**
Family resemblance approach						
Transitional	9.55	13.80	0.80	7.95	11.09	0.80
Non-transitional	9.85	15.20	0.83	8.41	13.23	0.83
Fuzzy decision tree approach						
Transitional	9.47	11.47	0.84	7.92	9.27	0.85
Non-transitional	10.37	15.54	0.80	8.42	13.54	0.79

Table 4
Accuracies of the derived *A* horizon textures: the family resemblance approach *vs.* the fuzzy decision tree approach on transitional and non-transitional sets based on uncertainty measures calculated with the fuzzy decision tree approach.

	Percentage of sand			*Percentage of silt*		
	MAE	**RMSE**	**AC**	**MAE**	**RMSE**	**AC**
Family resemblance approach						
Transitional	10.40	15.65	0.81	8.93	13.25	0.81
Non-transitional	8.22	11.68	0.80	6.59	9.46	0.82
Fuzzy decision tree approach						
Transitional	9.35	11.79	0.84	7.60	11.58	0.86
Non-transitional	11.00	13.00	0.79	9.31	11.24	0.80

Table 2
Accuracies of the derived *A* horizon textures: the inference results *vs.* the original map.

	Accuracy of soil series prediction	***Percentage of sand***			***Percentage of silt***		
		MAE	**RMSE**	**AC**	**MAE**	**RMSE**	**AC**
Family resemblance approach	80.8% (80/99)	9.69	14.47	0.81	8.17	12.14	0.82
Fuzzy decision tree approach	77.8% (77/99)	9.38	12.60	0.82	7.99	11.10	0.82
Ensemble approach	83.8% (83/99)	8.47	13.46	0.82	7.44	11.93	0.82
Original soil map	83.8% (83/99)	10.66	16.63	0.67	9.51	14.31	0.67

approach (both involve optimality curves and thus the most comparable) is attributed to the ability to capture soil variations on transitional areas, we separated the 49 field samples to two testing sets: the transitional set and the non-transitional set. The transitional set contains samples for which either the ignorance or exaggeration uncertainty measure computed from the inference result is greater than the average uncertainty in the study area. The non-transitional set, on the other hand, contains those samples for which both uncertainties are lower than the mean. We obtained three pairs of such separated testing data: one based on the uncertainty values computed from the fuzzy decision tree approach and one from the family resemblance approach, with the third one considering both. That is, if a field point is deemed to be low in uncertainty by both methods, it is on a non-transitional position. Otherwise, it is transitional. The accuracies of this testing are listed in Tables 3–5. It shows that based on all three versions of the separated test sets, the fuzzy decision tree approach has similar or even lower performance (similar or higher RMSEs or similar or lower ACs) than the family resemblance approach. It is the apparent superiority on the transitional positions that attributes to the over-performance of the fuzzy d tree approach overall.

Table 5
Accuracies of the derived *A* horizon textures: the family resemblance approach *vs.* the fuzzy decision tree approach on transitional sets based on uncertainty measures calculated with both the family resemblance and the fuzzy decision tree approach.

	Percentage of sand			***Percentage of silt***		
	MAE	**RMSE**	**AC**	**MAE**	**RMSE**	**AC**
Family resemblance approach	13.41	16.67	0.81	10.62	13.16	0.83
Fuzzy decision tree approach	12.04	14.09	0.86	8.86	10.64	0.89

The *A* horizon soil texture maps created with the three approaches are juxtaposed in Figs. 6 and 7 show the soil texture maps based on the original soil map for comparison. We observe that the inferred texture maps tend to illustrate more continuous changes of the texture values than those based on the original map. With the inferred maps, abrupt changes of texture mostly occur only when parent material changes. Comparing the texture maps generated using the three different knowledge discovery approaches; it is notable that the property maps generated with the fuzzy decision tree approach (Fig. 6a) exhibit smoother transitions than those with the family resemblance approach (Fig. 6b) and ensemble approach (Fig. 6c). The property values in

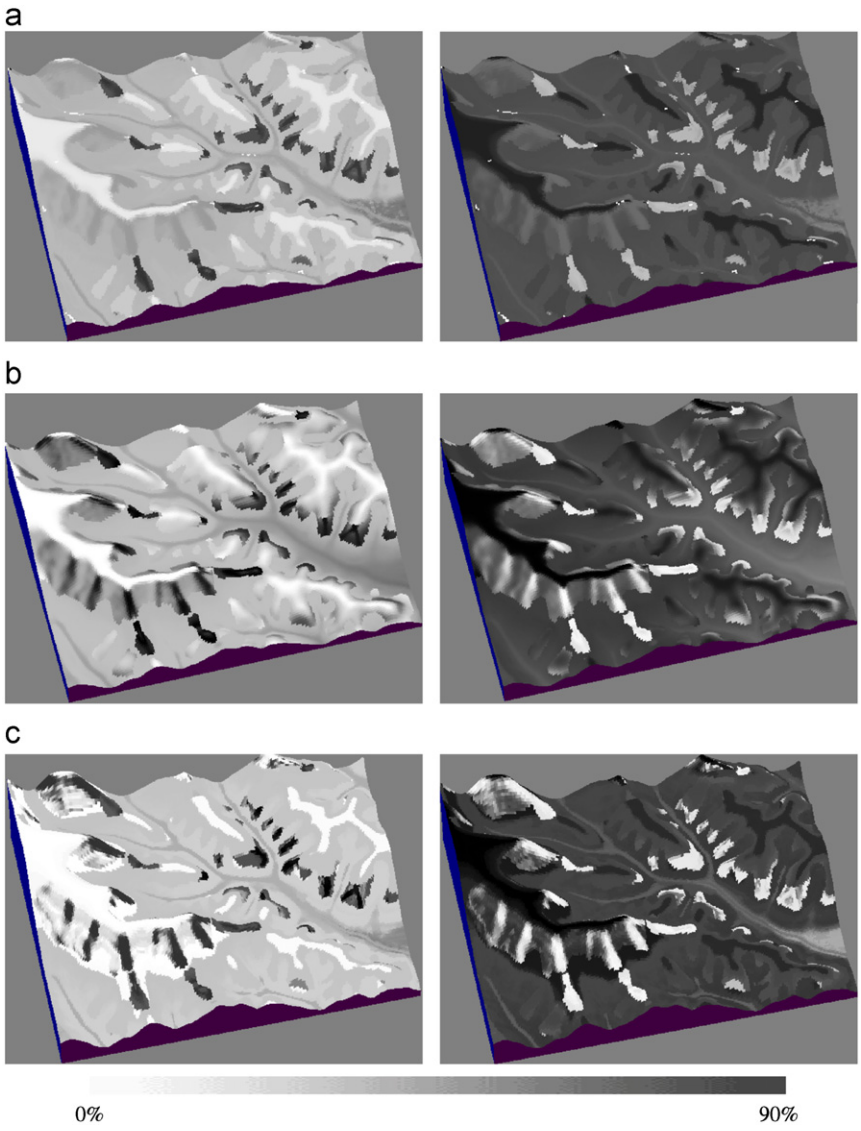

Fig. 6. *A* horizon sand (left) and silt (right) percentages derived from inference results with the (a) family resemblance approach, (b) fuzzy decision tree approach, and (c) ensemble approach.

Fig. 6a appear to change gradually and seamlessly, especially within the same regions of parent material. The maps based on the other two approaches (Fig. 6b and c) do show fuzzy transitional zones between soil types, but the extents of the transitional zones are rather limited, and most of the areas still appear to have uniform texture values. Although it is difficult to objectively determine how wide the transitional zones between soil classes should be, the accuracy measures computed for all three sets of maps indicate that the distributions of property values on the maps in Fig. 6a are closer to reality than the others. From a soil formation perspective, the transitions of soil classes should be in accordance with the transitions of the environmental conditions. Because the environmental variables (elevation, slope gradient, etc.) in this area gradually change across the mapped area, the soils formed in the area should exhibit similar gradual change in terms of their properties.

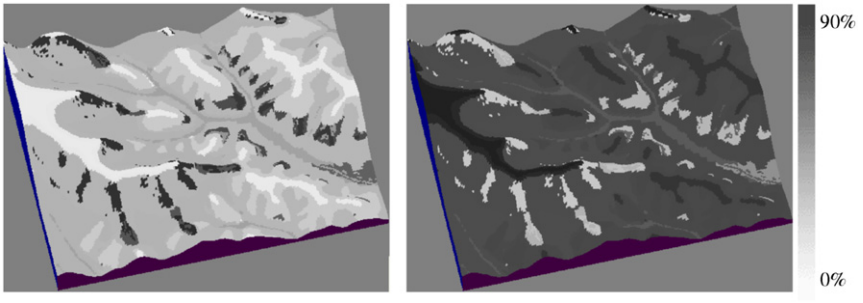

Fig. 7. *A* horizon sand (left) and silt (right) percentages based on the original soil map.

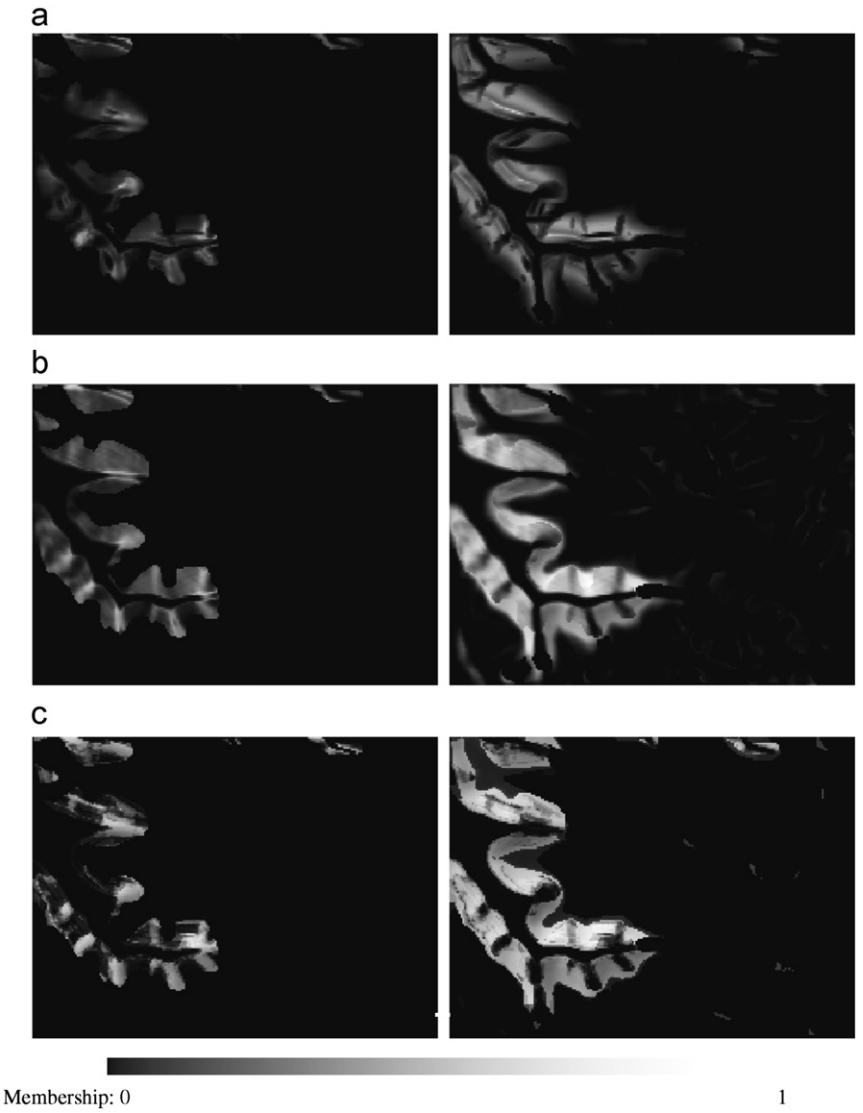

Fig. 8. Fuzzy membership maps of *Dorerton* (left) and *Elbaville* (right) inferred from knowledge extracted with the (a) family resemblance approach, (b) fuzzy decision tree approach, and (c) ensemble approach.

In order to examine the reason why the derived soil property maps appear different among the three approaches, the fuzzy membership values used to create the property maps were compared. Fig. 8 displays the fuzzy membership maps of soil series *Dorerton* and *Elbaville* inferred with the three approaches. On a fuzzy membership map, lighter pixels are those with higher membership values than darker ones. White zones are usually the typical positions at which to expect a particular soil series; and black zones are where memberships to the soil series are zero. Fig. 8 shows that the membership values tend to gradually fade away on the maps created with both the family resemblance (Fig. 8a) and fuzzy decision tree approach (Fig. 8b). The membership values inferred with the ensemble approach (Fig. 8), however, do not show as pronounced a fuzzy boundary as the other approaches. The reason could be that in this case study only 100 rounds of boosting were experimented with the ensemble approach. Unlike the other two approaches, which use continuous optimality curves to derive fuzzy memberships, the ensemble approach relies on the set of different decision trees to derive fuzzy memberships from count of votes. When the number of available trees is limited, therefore, the derived fuzzy membership values may not capture the full range of fuzzy gradations.

Another observation that can be made from the maps in Fig. 8 is that the membership values inferred from the family resemblance approach are consistently lower than those from the fuzzy decision tree approach, although both show apparent fuzziness. The difference in fuzzy membership values is actually a result of the difference of the optimality curves derived from the two approaches. Fig. 9 shows the optimality curves derived from both approaches for soil series *Dorerton* based on the feature slope gradient. The two curves are not only different in terms of their shapes, but also the locations of cross-over points. It is evident that the curve in Fig. 9a represents a more constrained fuzzy concept (in terms of limited gradient range) than that in Fig. 9b. The reason is that the curve in Fig. 9a was derived from a data histogram constructed using gradient values of *Dorerton* in the mapped area, which represents only the local distribution of the gradient for *Dorerton*. Although the global concept for *Dorerton* may be something that occurs on gradient greater than 25% (an S-shaped curve), the lack of pixels at certain slopes in the local area made the histogram cover only gradients from 35% to 50%. The curve derived with the fuzzy decision tree approach, however, is not limited by the lack of high gradient pixels in the local area since the curve was based on global partition of the feature space. Therefore, optimality curves based on post-fuzzified decision trees are broader than those based on local data histograms. As a result, inference using the narrower curves of the family resemblance approach will result in lower membership values of the soil class and possibly a more restrained spatial extent.

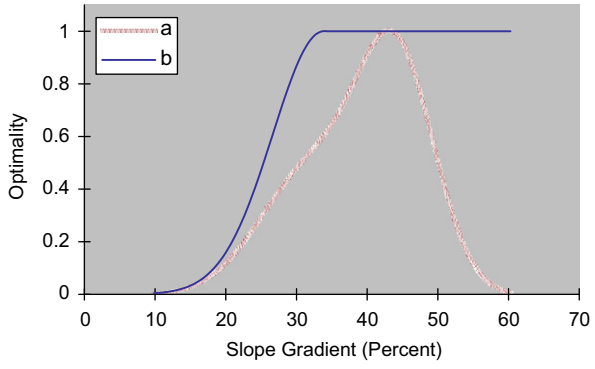

Fig. 9. Optimality curves of slope gradient for soil *Dorerton* derived for the (a) family resemblance approach, and (b) fuzzy decision tree approach.

5. Conclusion

All three methods were able to extract knowledge embedded in the soil map to classify soils based on their soil-formative environmental features. The knowledge, when used to classify soils, was able to predict soil series names of field samples at accuracies that are close to or the same as the accuracy of the original map. Moreover, the extracted knowledge also captures prototype effects, and thus can be used to infer an instance's memberships to different soil series. Such information can be used to derive measures of two types of uncertainty: the ignorance uncertainty and the exaggeration uncertainty. Uncertainty images helped to identify transitional areas and areas of potential problems on the inferred soil series map.

The three knowledge discovery methods employed in this study are quite different in terms of both the algorithms and theoretical basis. Field evaluation showed that the performances of the three approaches are slightly different. Inference results with the ensemble approach gave the most accurate prediction on soil series names (soil classification in the traditional manner). The fuzzy decision tree approach, however, gave the lowest error rates in terms of estimating continuous soil texture values. The reason why it is more accurate than the ensemble approach is that the inferred fuzzy membership values with this approach exhibit smoother and more continuous patterns than those with the ensemble approach (and the family resemblance approach) and capture better the transitions between soil prototypes. The ensemble approach we employed here used one hundred iterations. It may make a difference if more iterations are allowed. The reason why the property maps generated from the fuzzy decision tree approach are more accurate than those from the family resemblance approach lies in the fact that the optimality curves fitted from data histograms with the latter approach are affected by the local environment conditions and do not reflect the global characteristics of the soil classes.

The case study employs a simple weighted average method to estimate the soil property at a location, based on the typical properties of all soil classes it is similar to. With the weights being the local soil's similarity values to all prescribed soil classes, this method takes into consideration membership diffusion and thus reduces the omission error. It does not, however, deal with the exaggeration of membership. Future study may address this problem and investigate better ways of utilizing the similarity vectors.

References

Bennett, B., 2001. What is a forest? On the vagueness of certain geographic concepts. Topoi 20, 189–201.
Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40, 477–492.
Burrough, P.A., 1996. Natural objects with indeterminate boundaries. In: Burrough, P.A., Frank, A. (Eds.), Geographic Objects with Indeterminate Boundaries. Francis and Taylor, London, pp. 3–28.
Campbell, J.B., Edmonds, W.J., 1984. The missing geographic dimension to soil taxonomy. Annals of the Association of American Geographers 74, 83–97.
Canty, M.J., 2009. Boosting a fast neural network for supervised land cover classification. Computers and Geosciences 35, 1280–1295.
Chiang, I.J., Hsu, J.Y.J., 2002. Fuzzy classification trees for data analysis. Fuzzy Sets and Systems 130, 87–99.
Davis, T.J., Keller, C.P., 1997. Modelling and visualizing multiple spatial uncertainties. Computers & Geosciences 23, 397–408.
Davidson, D.A., Theocharopoulos, S.P., Bloksma, R.J., 1994. A land evaluation project in Greece using GIS, and based on Boolean, and fuzzy set methodologies. International Journal of Geographical Information Systems 8, 369–384.
Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B 39, 1–38.
Dietterich, T.G., 1997. Machine learning research: four current directions. AI Magazine 18, 97–136.
Ehlschlaeger, C.R., Goodchild, M.F., 1994. Dealing with uncertainty in categorical coverage maps: defining, visualizing, and managing errors. In: Proceedings,

Workshop on Geographical Information Systems at the Conference on Information and Knowledge Management, Gaithersburg, MD, pp. 86–91.
Ehlschlaeger, C.R., Shortridge, A.M., Goodchild, M.F., 1997. Visualizing spatial data uncertainty using animation. Computers and Geosciences 23, 387–395.
Ferrari, G., 1996. Boundaries, concepts, language. In: Burrough, P.A., Frank, A.U. (Eds.), Geographic Objects with Indeterminate Boundaries. Francis and Taylor, London, pp. 99–108.
Fillmore, C., 1985. Frames and semantics of understanding. Quaderni di Semantica 6, 222–253.
Fisher, P.F., 1994. Visualizing uncertainty in soil maps by animation. Cartographica 30, 20–27.
Fisher, P.F., Comber, A.J., Wadsworth, R., 2005. Approaches to uncertainty in spatial data. In: Devillers, R., Jeansoulin, R. (Eds.), Fundamentals of Spatial Data Quality. ISTE, London, pp. 43–59.
Fisher, P.F., 2010. Uncertainty and error. In: Encyclopedia of Geographic Information Science. SAGE Publications. Accessed at: 〈http://sage-ereference.com/geoinfoscience/Article_n221.html〉.
Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Saitta, L. (Ed.), Proceedings of the 13th International Conference on Machine Learning, Morgan Kaufmann, Bari, Italy, pp. 148–156.
Goodchild, M.F., Sun, G., Yang, S., 1992. Development and test of an error model for categorical data. International Journal of Geographical Information Systems 6, 87–104.
Goodchild, M.F., Chin-Chang, L., Leung, Y., 1994. Visualizing fuzzy maps. In: Hearnshaw, H.M., Unwin, D.J. (Eds.), Visualization in Geographical Information Systems. John Wiley & Sons, New York, pp. 158–167.
Greve, M.H., Greve, M.B., 2004. Determining and representing width of soil boundaries using electrical conductivity and multiGrid. Computers & Geosciences 30, 569–578.
Harrower, M., 2003. Representing uncertainty: does it help people make better decisions? UCGIS Workshop: Geospatial Visualization and Knowledge Discovery Workshop, National Conference Center, Landsdowne, VA., November 18–20.
Hodza, P., 2010. Fuzzy logic and differences between interpretive soil maps. Geoderma 156, 189–199.
Lakoff, G., 1987. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. University of Chicago Press, Chicago 632 pp.
Leugn, Y., Yan, J., 1998. A locational error model for spatial features. International Journal of Geographical Information Systems 12, 607–620.
Markman, A.B., 1999. Knowledge Representation. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.
Miller, H.J., Han, J., 2001. Geographic data mining and knowledge discovery: an overview. In: Miller, H.J., Han, J. (Eds.), Geographic Data Mining and Knowledge Discovery. Taylor & Francis, New York, pp. 3–32.
Minda, J.P., Smith, J.D., 2001. Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. Journal of Experimental Psychology: Learning, Memory, and Cognition 27, 775–799.
Moran, C.J., Bui, E.N., 2002. Spatial data mining for enhanced soil map modeling. International Journal of Geographical Information Science 16, 533–549.
Ola, A., Johannes, K., Oukbir, K., 2000. Rough classification and accuracy assessment. International Journal of Geographical Information Science 14, 475–496.
Plewe, B., 2002. The nature of uncertainty in historical geographic information. Transactions in GIS 6, 431–456.
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., 2004. Knowledge discovery from area-class resource maps: data preprocessing for noise reduction. Transaction in GIS 8, 297–308.
Qi, F., Zhu, A.X., Harrower, M., 2006. Fuzzy soil mapping based on prototype category theory. Geoderma 136, 774–787.
Qi, F., Zhu, A.X., 2006. Modeling uncertainty in knowledge discovery for classifying geographic entities with fuzzy boundaries. In: Riedl, A., Kainz, W., Elmes, G. (Eds.), Progress in Spatial Data Handling: 12th International Symposium on Spatial Data Handling. Springer Verlag, Berlin.
Qi, F., Zhu, A.X., Pei, T., Qin, C., Burt, J.E., 2008. Knowledge discovery from area-class resource maps: capturing prototype effects. Cartography and Geographic Information Science 35, 223–237.
Rosch, E.H., 1973. Natural categories. Cognitive Psychology 4, 328–350.
Rosch, E.H., 1978. Principles of categorization. In: Rosch, E.H., Lloyd, B.B. (Eds.), Cognition and Categorization. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 27–48.
Rosch, E.H., Mervis, C., 1975. Family resemblances: studies in the internal structure of categories. Cognitive Psychology 7, 573–605.
Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography 27, 171–197.
Shi, W., 1998. A generic statistical approach for modelling error of geometric features in GIS. International Journal of Geographical Information Systems 12, 131–143.
Shi, X., Zhu, A.-X., Burt, J.E., Simonson, D., Qi, F., 2004. A case-based reasoning approach to fuzzy soil mapping. Soil Science Society of America Journal 68, 885–894.
Smith, E.E., Medin, D.L., 1981. Categories and Concepts. Harvard University Press, Cambridge, MA.
Tversky, B., Hemenway, L., 1984. Objects, parts, and categories. Journal of Experimental Psychology: General 113, 169–193.
Van Der Wel, F.J.M., Van Der Gaag, L.C., Gorte, B.G.H., 1998. Visual exploration of uncertainty in remote-sensing classification. Computers & Geosciences 24, 335–343.
Willmott, C.J., 1984. On the evaluation of model performances in physical geography. In: Gaile, G.L., Willmott, C.J., Reidel, D. (Eds.), Spatial Statistics and Models. Dordrecht, Holland, pp. 443–460.
Wittgenstein, L., 1953. Philosophical Investigations. Macmillan, New York.
Worboys, M.F., 2001. Nearness relations in environmental space. International Journal of Geographical Information Science 15, 633–651.
Zhang, B., Valentine, I., Kemp, P., 2005. Modeling the productivity of naturalized pasture in the North Island, New Zealand: a decision tree approach. Ecological Modeling 186, 299–311.
Zadeh, L., 1965. Fuzzy sets. Information and Control 8, 338–353.
Zhu, A.X., 1997a. Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. Photogrammetric Engineering & Remote Sensing 63, 1195–1202.
Zhu, A.X., 1997b. 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.
Zimmermann, H.J., 2000. An application-oriented view of modeling uncertainty. European Journal of Operational Research 122, 190–198.