INT. J. GEOGRAPHICAL INFORMATION SCIENCE, 1999, VOL. 13, NO. 2, 119–141

Research Article

A personal construct-based knowledge acquisition process for natural resource mapping

A-XING ZHU

Department of Geography, University of Wisconsin, Madison, Wisconsin 53706, USA
e-mail: axing@geography.wisc.edu

(*Received 26 November 1996; accepted 26 August 1998*)

Abstract. This paper presents an iterative, structured knowledge-acquisition process for extracting human understanding of relationships between a natural resource and its environment. This understanding can then be used to map natural resources as spatial continua. The knowledge acquisition process is based on personal construct theory and consists of several iterations. Each iteration has five structured interview sessions: preparation, key development, description, comparison, and quantification. The knowledge derived from each iteration is represented as a set of membership functions that describes the degree to which a given environmental condition impacts the status of the given resource. The final set of membership functions, which is the final version of knowledge, is derived through the comparison and 'fusion' of the membership functions from each iteration. The comparison of the membership functions among different iterations is also used to measure the consistency (integrity) of an expert's understanding of the relationships. In a soil mapping case study, knowledge on soil-environment relationships was acquired from a local soil scientist using the knowledge acquisition process. The case study shows that knowledge sets extracted a year apart were consistent with each other. The study also shows that the soil expert was more familiar with the relationships between soils and some environmental variables than with other environmental variables. The expert's understanding about soil-environmental relationships also differed among soil series. Although it was designed to extract expert knowledge for mapping natural resources as spatial continua under a GIS environment, this knowledge elicitation process can be easily adapted to extract expert knowledge for other knowledge-based applications.

1. Introduction

Environmental decision making requires knowledge about the spatial distribution, quantity and quality of environmental resources. Current mapping practice often employs the 'discrete spatial model' (as described in Bregt 1992), the 'object model' (as described in Goodchild 1989 and 1992), and the 'area-class map' concept (as in Mark and Csillag 1989). In each of these cases the spatial continuum of a given natural resource is discretized into discrete and distinct spatial units. This discretization limits the portrayal of details about the spatial variation of natural

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online
http://www.tandf.co.uk/JNLS/gis.htm
http://www.taylorandfrancis.com/JNLS/gis.htm

resources that are needed for many environmental management applications (Band and Moore 1995, Zhu 1997a).

Many researchers have investigated the use of fuzzy logic for representing spatial phenomena (Robinson 1988, Burrough 1989, Fisher and Pathirana 1990, Wang 1990, Burrough *et al.* 1992, McBratney and De Gruijter 1992, Odeh *et al.* 1992, McBratney and Odeh 1997). In particular, Zhu (1997a) developed a similarity model based on fuzzy logic for mapping soil as a spatial continuum. In general, the similarity model consists of two components: a similarity vector for representing soil in its parameter space and a raster scheme for representing the variation of these similarity vectors in geographical space. Using a similarity vector, the soil at a given point (i, j) is represented by an n-element similarity vector, $\boldsymbol{S}_{ij} = (S_{ij}^1, S_{ij}^2, ..., S_{ij}^k, ..., S_{ij}^n)$, where S_{ij}^k is the similarity value (membership value) of the soil at point (i, j) to soil type k; and n is the number of prescribed soil types. This representation allows the local soil at point (i, j) to be represented as intermediate to the prototypes of prescribed classes, and thus the local conditions of the soil can be preserved. Details on the spatial variation of soils can be accommodated by a raster data model which is capable of representing spatial phenomena at a very fine spatial resolution. Under the raster data model, soil over an area is presented as an array of pixels with the soil at each pixel being represented as a similarity vector. The spatial variation of soil is then represented as the variation of membership values in these similarity vectors over space (Zhu 1997a).

Mapping natural resources under this similarity model is a process of deriving membership values of the given resource at each location (pixel) to a set of prescribed resource categories (classes). Some natural resources (such as forests, or water bodies) are directly observable using remote sensing techniques and membership values for these resources over space can be derived through fuzzy classifications (Bezdek *et al.* 1984, Wang 1990, McBratney and De Gruijter 1992). Other natural resources (such as soils, and some wetlands) cannot be easily observed due to obscuring overstories and the high cost of collecting information about these resources at many locations across landscapes. However, the distribution of these resources may be inferred (or indirectly mapped) from other easily observable environmental conditions (Coulson *et al.* 1991, Skye and Naia 1993, Zhu and Band 1994, Mulder and Corns 1996, Skidmore *et al.* 1996, Zhu *et al.* 1996).

Indirect mapping of natural resources is based on the assumption that a relationship exists between a given resource and its environment, which can be expressed as

$$\boldsymbol{R}_{ij} = f(\boldsymbol{E}_{ij}) \tag{1}$$

According to the similarity model, $\boldsymbol{R}_{ij}$ in equation (1) is the similarity vector that represents the similarity of the resource at location (i, j) to a set of prescribed resource classes. The kth element in $\boldsymbol{R}_{ij}$, R_{ij}^k, is the similarity of the resource at (i, j) to resource category k. The collection of R_{ij}^k over the entire area forms a membership map, $\boldsymbol{R}^k$, representing the spatial distribution of similarity to the prescribed resource category k over the area (Zhu 1997a). The assemblage of these membership maps ($\boldsymbol{R}$) forms the similarity representation of the given resource over the area. $\boldsymbol{E}_{ij}$ in equation (1) is the environmental conditions at location (i, j), which can be derived from a GIS database. f is a set of membership functions each describing the relationship between a given resource category and its environmental conditions (Zhu *et al.* 1996).

The success of populating $\boldsymbol{R}_{ij}$ using equation (1) for a given resource depends on the accuracy of data on the environmental conditions ($\boldsymbol{E}_{ij}$) and the validity of membership functions (f). Although remote sensing and GIS techniques are often used to capture, derive, and represent the spatial distribution of environmental conditions (E_{ij}), detailed knowledge about the relationships between specific resource types and their environment conditions is often not stored in a database, particularly not in the form of membership functions. In order to define these membership functions, two types of information are needed. The first is the typical environment configurations under which resource category k exists. This knowledge is called *Type 1 knowledge* (Zhu *et al.* 1996), which may exist in forms such as global classification systems (e.g. soil taxonomy) and resource descriptions (e.g. soil descriptions). The second type of information needed to define the membership functions is the change of membership in resource category k with regard to the deviation of environmental conditions from the typical configuration. It is this second type of information, referred to as *Type 2 knowledge* (Zhu *et al.* 1996), which is needed to map natural resources as spatial continua under fuzzy logic. Type 2 knowledge may be approximated by general linear regression models. However, the general knowledge captured in these regression models does not adequately predict local conditions, particularly at the level of detail required by land managers and users interested in local conditions. Furthermore, the relationships between changes in membership and changes in environmental conditions may not be simply described using general regression models since the relationships can be highly nonlinear. In many cases, more detailed knowledge exists in the form of human expertise (Fisher 1989, Hudson 1990, 1992). It is likely that after many years of field work local resource experts (such as soil experts) understand well the typical environmental conditions for various resource classes and how the resource may change when one moves away from these typical environment configurations in a local area. It is therefore necessary to acquire the two types of knowledge from human experts so that indirect resource mapping can be possible and more detailed spatial information about resources can be obtained for that area.

Acquiring knowledge from domain experts has been considered as a bottleneck for the development of knowledge-based approaches and systems (Hayes-Roth *et al.* 1983, Robinson and Frank 1987, Molokova 1993, Weibel *et al.* 1995). The difficulty of knowledge acquisition is due to several factors. First, knowledge is often neither well formulated nor precisely defined. Second, there is a lack of understanding about how human beings acquire, organize and process domain knowledge. This lack of understanding leads to difficulty in communicating domain knowledge from the domain expert (the person whose knowledge is to be extracted) to the knowledge engineer during knowledge elicitation processes (Ford *et al.* 1991).

To overcome some of these difficulties, a knowledge acquisition process based on Kelly's personal construct theory was developed and is presented here. The next section of this paper presents a brief introduction to personal construct theory (Kelly 1955, Kelly 1970, Adams-Webber 1984, Ford *et al.* 1991, Shaw and Gaines 1993). The knowledge acquisition process itself is presented in §3 and further illustrated through a soil mapping case study in §4. Section 5 provides a discussion of the knowledge acquisition process and the experience gained through this case study. The conclusions are given in §6.

2. Personal construct theory

In general, personal construct theory assumes that people typically use cognitive dimensions (termed 'constructs') to learn and evaluate their experience (Kelly 1955).

Each construct, by definition, represents a single bipolar distinction. For example, a bird ecologist might use the construct 'hot/cold' to describe the temperature requirement of birds; a person may use the construct 'wet/dry', among others, to distinguish climate types. The underlying relation between the alternative poles of any construct is contrariety and no construct can be understood fully without considering the meaning of both poles (Husain 1983). A construct is a basis for making a distinction and is a dichotomous reference axis in a person's psychological space (Kelly 1970).

Kelly (1955, p. 68) also posits that 'Each construct is convenient for the anticipation of a finite range of events only' and has a specific *range of convenience*, which constitutes 'all those things to which the user would find its application useful'. The poles of a construct define the extremes of the range of convenience. This range can be scaled or divided into intervals that each represents a specific level of convenience. For example, students may use the construct 'excellent/poor' to summarize their evaluation of a professor's teaching performance. The range of convenience encapsulated by this construct could be scaled from excellent (4), above average (3), average (2), below average (1) to poor (0). Any particular construct may have a somewhat different 'context' or meaning for each person who uses it. In the teaching evaluation example, excellent teaching performance may have a different meaning to each student. As a result, different ratings about a professor's teaching may be reported by different students using the same construct.

Kelly (1955) introduces the notion of *psychological space* as a term for a region in which we may place and classify the elements of our experience. One's psychological space is made of different but overlapping (intersecting) constructs. It is the overlap (intersection) between the constructs' ranges of convenience that enables an event to be anticipated, placed, or classified. The psychological space comes into being through a process of construction; individuals create a space in which to place elements as they come to construe them. The process of construction is the process of forming relevant hypotheses that are tested against available evidence when one seeks to predict and control events. In other words, one's psychological space evolves (expands) when more constructs are added.

Under Kelly's personal construct theory, the accumulation of a resource expert's knowledge about the relationships between a resource and its environment can be considered as a process of constructing that person's psychological space about the relationships. For example, a soil scientist may accumulate knowledge about the relationship between a specific soil type and its environment by formulating constructs (such as 'steep/flat', 'south facing/north facing') and locating the intersections of these constructs to define the environmental niche or niches under which the soil exists. Under this notion, acquiring knowledge about relationships between a resource and its environment becomes a process of defining relevant (proper) constructs (both poles and intervals) and locating the intersections of these constructs for various classes of that resource.

3. The knowledge acquisition process

Based on the assumption that resource experts acquire and organize knowledge in a way theorized by personal construct theory, a knowledge acquisition process consisting of two phases was developed (figure 1). The first, or *iteration phase*, consists of many knowledge acquisition iterations. Each iteration consists of five sessions: (1) the preparation session, (2) the resource-environment key development session, (3) the resource-environment description session, (4) the key and description

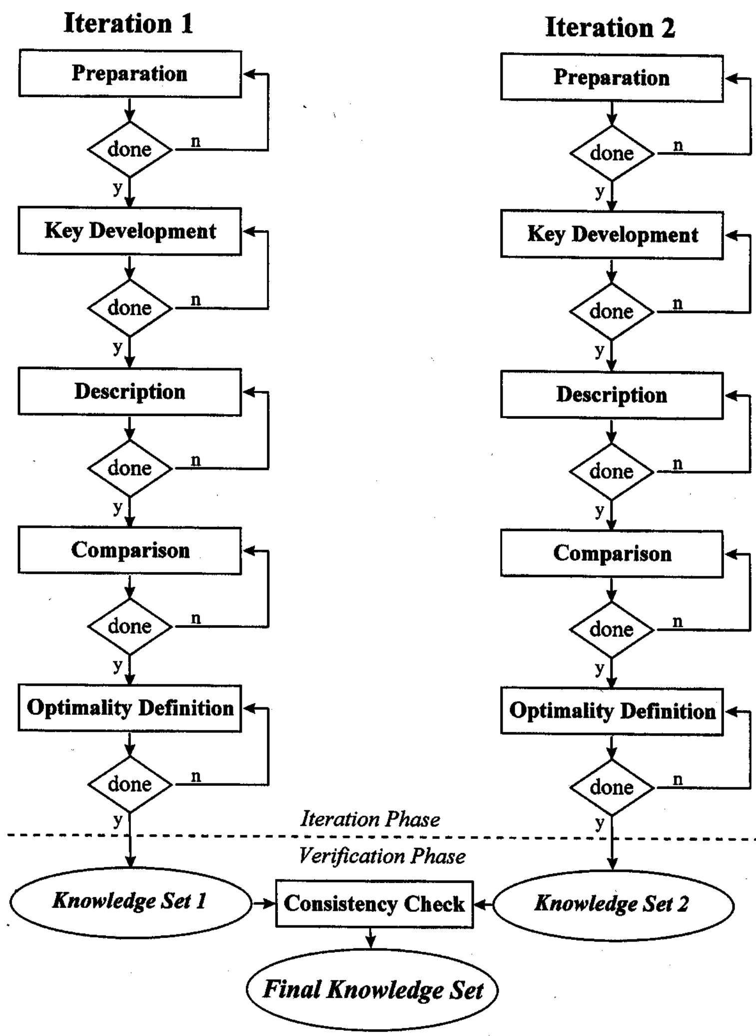

Figure 1 An iterative knowledge acquisition process.

comparison session, and (5) the optimality curve definition session. These five sessions are tightly connected and form a structured interview. In session 1, constructs are defined. Session 2, 3 and 4 are used to define the intersections (*Type 1 knowledge*) of these constructs. Session 5 is used to extract *Type 2 knowledge* and to integrate Type 1 knowledge with Type 2 knowledge to form the complete version of the knowledge for the iteration. Results from different iterations are then compared and analysed in the second phase called the *knowledge verification phase.*

3.1. *Iteration phase*

The five interview sessions should not be considered as separate meetings with the expert. An interview session can very well consist of many iterative meetings over a period of a few days. A check should be performed at the end of each interview session to improve the chance that the knowledge extracted is a 'good' representation of the expert's understanding. At the end of each session the knowledge engineer should always ask the expert to review the results and determine if changes need to be made. Before starting the next interview session, the knowledge engineer should again ask the expert to review the results produced in the previous interview session. This stepwise refinement technique allows the expert to examine over time the results produced in each interview session and to correct or update them. In this way, the knowledge extraction would not be limited to just a few interview sessions but becomes a long elicitation process for the expert.

Preparation Session: This preliminary interview session is designed to help a knowledge engineer and a resource expert formulate constructs and determine classes for a given resource. For this process, the term 'construct' will be replaced by 'variable' since domain experts are more familiar with that term. In fact, the term 'construct' should not be used in the communication between the knowledge engineer and the domain expert since it may confuse the expert. The application of personal construct theory is only to assist and guide the knowledge engineer during the knowledge elicitation process.

To begin, the knowledge engineer presents the following two questions to the domain expert(s): (1) How many different classes of the resource are there in the area to be mapped? (2) What environmental variables do you need to distinguish these different classes of resources? The first question establishes the events to be anticipated and the second question outlines the constructs of the resource environment (environmental constructs). Answering the first question may be relatively easier since existing classification systems can be used as a template. In some cases, project requirements may already specify the classes to be mapped or the classification system to be used. Knowledge engineers must have some basic understanding of the domain so that they can assist the expert in defining constructs. Knowledge engineers need to consider the following three aspects in this process: (*a*) the exhaustiveness of these constructs in describing the expert's psychological space; (*b*) the comfort of the expert in using the construct; (*c*) the feasibility of deriving spatial data about the construct from a GIS. One way to lead the domain expert is to start with common environmental variables (such as elevation, slope). The domain expert may then wish to add additional variables to the list.

Once constructs are determined, the poles and the intervals for each construct must be defined. The knowledge engineer may ask the domain expert to find the 'natural breaks' in the range of values for a given variable in the context of mapping the given resource so that the poles and the intervals for that construct may be defined. The breaks will not necessarily be numerical. Sometimes, the domain expert may prefer qualitative terms over quantitative intervals. The knowledge engineer may need to use examples to illustrate the meaning of natural breaks in the range. Defining the constructs is a process that enables a knowledge engineer to establish a coordinate system which is similar to that used to reference the expert's psychological space for 'placing' the knowledge. A domain expert may never consciously organize knowledge in this way but sub-consciously may use this system for placing and acquiring knowledge.

Key Development Session: Once the coordinate system is established, in the next step it is used to locate or extract the intersections of the constructs. The key development session was designed for the domain expert to develop a dichotomous key that differentiates different classes of resources using the constructs defined in session 1. For example, a forester may know that there are three types of forests in an area: one on south-facing slopes at low elevations (*Forest Type A*); one on south-facing, high elevation slopes and also at low elevation, north-facing slopes (*Forest Type B*); the third (*Forest Type C*) on north-facing slopes at high elevations. Given that we have two constructs: 'north-facing/south-facing' and 'high elevation/low elevation', the key for representing the forester's knowledge would then be:

North-facing
    High elevation .... Forest Type C
    Low elevation .... Forest Type B
South-facing
    High elevation .... Forest Type B
    Low elevation .... Forest Type A

In the key, the environmental constructs are listed on the left and the forest types (the events or classes) are listed on the right. The path leading to each forest type from the constructs identifies the intersection of these constructs. This intersection characterizes the environmental niche under which a given forest type can be found. For example, the path to Forest Type A is 'south-facing via low elevation' and the environmental niche for this forest type is 'areas with south-facing slopes at low elevation'. For Forest Type B, there are two paths to it, or two environmental configurations ('north-facing at low elevations' and 'south-facing at high elevation'), under which Forest Type B exists. The designation of different instances in the key compensates for multimodality which may destroy relationships that many statistical techniques were designed to seek.

The dichotomous key development allows both the engineer and the domain expert to traverse the psychological space using constructs. This traversing through the expert's psychological space helps the expert to organize knowledge and eases communication between the knowledge engineer and the domain expert. The end product of this session is a key that can be used to differentiate resource classes.

Environment Description Session: Unfortunately, knowledge acquisition is often complicated because of various psychological factors that limit the ability of an expert to express knowledge. Consequently, the environment description session is designed to extract Type 1 knowledge via other means. During this session, the expert is asked to describe the environmental niche(s) for a given resource class using defined constructs without reference to the key developed earlier. In the forest type example, a description for Forest Type A would be:

Forest Type A:
    Elevation: Low
    Aspect: South

This describes the environmental niche under which Forest Type A exists. For Forest Type B, there are two descriptions, each corresponding to one of the instances. The result from this session is a set of environmental descriptions for the resource classes. There is an apparent relationship between this description and the key from the previous session. If one travels in the reverse order through the key, which is from the right (the Forest Type) to the left (the constructs), and records the ratings for constructs encountered, an environmental description of a given resource class results. Since the expert does not have access to the key developed in the previous session, the description from this session is independent of the key.

Key and Description Comparison Session: The purpose of this session is to compare the results of the description session with those derived from the key and to correct inconsistencies between the two. Before the session starts, the knowledge engineer must convert the key into descriptions. Once the two sets of descriptions have been compiled, the knowledge engineer and the domain expert compare them and check for inconsistencies. If an inconsistency is found, the domain expert is asked to study this inconsistency and decide which version is correct or to provide a new version if both are incorrect. The knowledge engineer normally asks the domain expert to explain decisions in resolving inconsistencies so that the knowledge engineer can record which inconsistency is resolved after careful deliberation and which is resolved in an *ad hoc* fashion. The process continues until all inconsistencies between the key and the descriptions are resolved.

The final result after this comparison session is Type 1 knowledge that describes the typical environmental conditions under which different resource classes exist. Type 1 knowledge is a type of descriptive knowledge (declarative knowledge) that can be best represented (organized) using a knowledge frame (Bench-Capon 1990, Georgiyev 1993, Moody *et al.* 1996). A frame consists of an identification tag, an instance ID, and a set of slots. The number of slots in a frame depends on the number of constructs (environmental variables) used. Each slot has a slot ID and a filler. A slot ID would be the name of a construct and the filler is the rating on that construct. The structure of a frame and an example of Type 1 knowledge for Forest Type B are shown in figure 2.

Optimality Curve Definition Session: Type 2 knowledge defines quantitatively the relationship between a given resource and its environment. The question the domain expert needs to answer is 'how does the status of a given resource class change as environmental conditions deviate from the optimal configuration in the parameter space?' It is almost impossible for any expert to answer this question in a multi-dimensional space. However, it might be possible to have an expert answer this question one variable at a time. In the forest type example, the expert may be able to tell how Forest Type A might respond when one moves along the elevation gradient. The term 'optimality' is introduced to help experts measure a response.

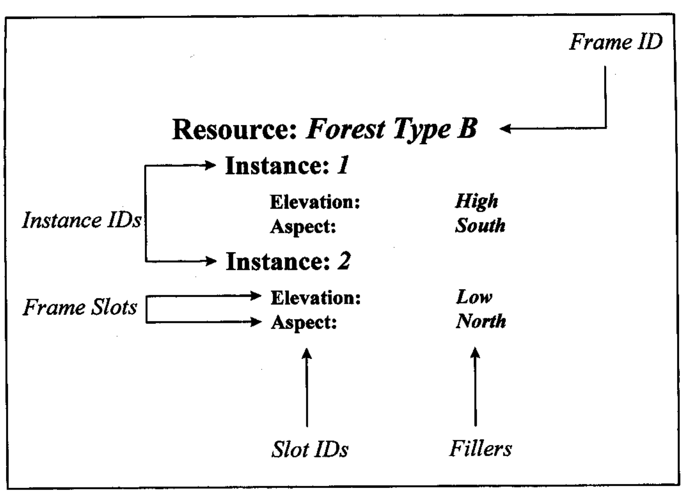

Figure 2. Frames used to represent *Type 1 knowledge.*

Optimality is defined as the degree to which an environmental condition favors the development of a given resource class assuming that all other environmental factors are most favourable. An optimality value of unity means that the given environmental condition is most favourable to the development of the given resource class, and an optimality value of zero means that the condition prohibits the development of the given resource class. For a given resource class, there should be an optimality value with respect to every condition of a given environmental variable. If one plots these optimality values with respect to the values of that environmental variable, one gets a curve (*optimality curve*, also referred to as *membership function*) that portrays the change of optimality for the given resource class with respect to the environmental variable. Since an optimality curve is defined with respect to a specific environmental variable and is specific to a given resource class, there would be *n* optimality curves for a given resource class when *n* environmental variables are involved.

Each of the optimality curves is defined by the domain expert with the use of a graphical user interface (GUI) (figure 3) which was developed to incorporate type 1 knowledge and to assist the domain expert in defining optimality curves. The GUI consists of three sections: the title area, the plotting area and the control area. The domain expert expresses his/her knowledge by specifying critical points for the optimality curve in the plotting area of the GUI. It is impossible for an expert to specify all points for the curve since the expert may not know the optimality for every environmental condition. However, it may be possible to ask the expert to specify critical points that are joined together to form a curve. The critical points are determined by performing the following tasks: (1) Indicate the environmental

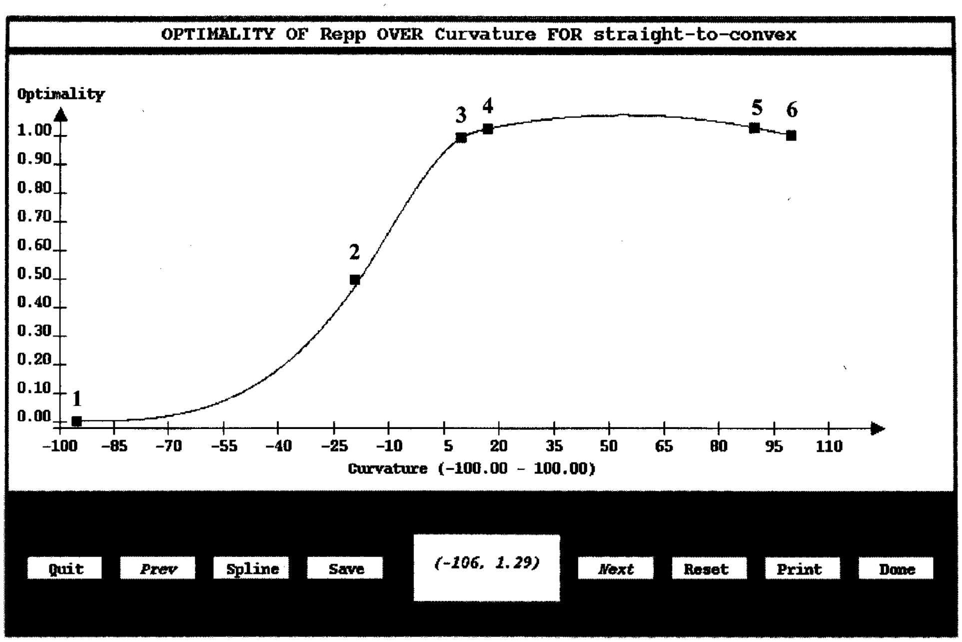

Figure 3. A graphical user interface (GUI) for optimality definition and an optimality curve for soil series Repp with respect to slope profile curvature. The profile curvature values are normalized to a range between −100 and 100 with −100 being concave, 0 being straight and 100 being convex. This curve shows the expert's understanding on the relationship between soil series Repp and the slope profile curvature.

conditions where the optimality values change to or move away from unity, (2) Indicate where the optimality values become zero or increase from zero, (3) Locate the points where the optimality values are at half of unity'. The formulation of these tasks is again based on personal construct theory which emphasizes that one organizes knowledge (distinguishes events) using distinctions and that the knowledge associated with extreme values is most likely well retained in human minds. For example, a forest ecologist may know the environmental conditions that most favour the growth of a given forest type and what environmental conditions would prohibit the development of that forest type. By associating the critical points with these extreme scenarios, the success of obtaining true knowledge from the expert would increase.

The critical points are joined using a spline function to form the optimality curve. There are two reasons for using a spline function. First, a spline curve passes through every critical point, which is very much desired for the optimality curve. Secondly, a spline passes through every critical point with a smooth transition from one side of a point to the other. After a spline has be fitted to the critical points, the expert inspects the curve and is allowed to change it by adding/deleting/moving the critical points and to refit a spline to these points. This process continues until the expert is satisfied with the curve. Figure 3 shows an optimality curve for soil series Repp with respect to environment variable ***Profile Curvature***. The optimality curve is drawn from a case study discussed later in this paper. Soil series Repp is a young soil with little horizon development and occurs on straight-to-convex slope segments where down-slope wash of parent materials prevents the development of soil horizons. The soil expert defined the optimality curve with respect to slope profile curvature by specifying four critical points (Points 1, 2, 3, and 6 in figure 3) where the optimality values are 0, 0.5, 1.0, and 1.0, respectively. The other points (Points 4 and 5) modify the shape of the curve to the expert's satisfaction.

During this interview session the knowledge engineer must explain clearly to the expert the meaning of optimality using the terms most familiar to the domain expert. Performing the three tasks for locating critical points must be explained and demonstrated to the expert so that the expert has an clear understanding of the meaning of the critical points. Once the expert understands the procedure the task of defining the optimality curve is much easier.

The outcome of this session is a set of optimality curves. These optimality curves are organized by resource class. In fact, the frame structure used for the Type 1 knowledge is used again to store these curves. Each filler in a frame for a resource class is now occupied by an optimality curve for that resource class with respect to the environmental variable designated to this slot (figure 4).

The entire iteration ends with the completion of the optimality definition session. Although there are two sets of frames for storing knowledge (one for the Type 1 knowledge and the other for the optimality curves), the knowledge frames that contain the optimality curves summarize all of the knowledge extracted from this iteration. The optimal environmental conditions extracted as Type 1 knowledge are now represented on the optimality curves as the environmental conditions where the optimality values are at unity. Other optimality values on the curves are Type 2 knowledge.

3.2. *Verification phase*

The extracted knowledge may contain substantial amounts of *ad hoc* information, which can be easily introduced in the optimality curve definition session. It is

Resource: ***Forest Type B***

Instance: ***1***

Elevation:	***Forest_B_Elevation_High.rel***
Aspect:	***Forest_B_Aspect_South.rel***

Instance: ***2***

Elevation:	***Forest_B_Elevation_Low.rel***
Aspect:	***Forest_B_Elevation_North.rel***

Figure 4. Frames for storing optimality curves.

therefore difficult to know if the knowledge extracted from an iteration is a true representation of the expert's understanding of the subject matter even though the extracted knowledge is derived from many well-structured interview sessions. There is a need to verify the significance of the extracted knowledge. It is worth pointing out that the term 'significance' here means how well the extracted knowledge approximates the expert's true understanding. It must not be confused with the term 'accuracy' which tends to imply the agreement between resource-environmental relationships in the extracted knowledge and those that exist in reality. The term 'verification' here means not only checking the significance of the extracted knowledge but also, more importantly, the process of improving its significance. There are two main tasks to be completed in the knowledge verification phase. The first is to remove as much of the *ad hoc* information as possible from the extracted knowledge (*knowledge refinement*). The second is to provide an index to measure the significance of the extracted knowledge.

Knowledge refinement: Recent literature on verification and validation (V & V) of knowledge-based systems (KBS) has suggested consistency checking as a means of refining and verifying knowledge bases (Mengshoel and Delab 1993, Craw 1996, Gamble and Baughman 1996, Plant 1996). Consistency checking looks for contradictions that occur in a knowledge set or inconsistency between two versions of knowledge provided by an expert. In this paper, consistency checking means the detection and correction of inconsistencies between different versions of knowledge derived from different iterations.

The knowledge acquisition iteration can be repeated at different times to obtain versions of knowledge from the same expert. There should be a sufficient time lag between successive iterations to allow the expert to forget what he did so that in the following iteration he must recall knowledge from real understanding, not from an impression of what was done in the last iteration. Once the iterations are completed, the results can be used to check for inconsistencies and the expert can then correct these inconsistencies. For the convenience of this discussion two iterations are assumed to have been conducted. Since the results from each iteration are stored as optimality curves, the consistency check is performed by comparing optimality curves from the first iteration with corresponding curves from the second iteration. The comparison begins by displaying two corresponding optimality curves in a GUI (figure 5). The expert is then asked to compare these two curves. If there are inconsistencies between the two curves, the expert is asked to study the inconsistencies and to provide a final optimality curve by choosing one of the two or defining a new one (figure 5). The final version of the extracted knowledge is expected to approximate the expert's understanding better than either one of the original versions since it is derived by comparing the two original versions and by studying the inconsistencies between them.

Measuring consistency: The consistency between two versions of knowledge extracted during different iterations is described by *consistency measure (CM)*:

$$CM = \frac{2(A_{1\cap 2})}{(A_1 + A_2)} \tag{2}$$

where A_1 is the area under optimality curve 1, A_2 is the area under optimality curve 2, and $A_{1\cap 2}$ is the area under both optimality curve 1 and optimality curve 2 (figure 6(*a*)). It is apparent that the more coincident the two optimality curves are, the larger $A_{1\cap 2}$ is, and therefore the higher CM is. CM ranges from 0 to 1. When CM equals 1, there is a perfect match between the two optimality definitions (figure 6(*b*), Top). If CM is 0, it means that there is no agreement between the two optimality definitions (figure 6(*b*), Bottom). CM is sensitive to the spreads of two corresponding curves. If the distance between the centres (such as modes) of the two curves is fixed, the wider the spread, the bigger the overlap, and the higher is the

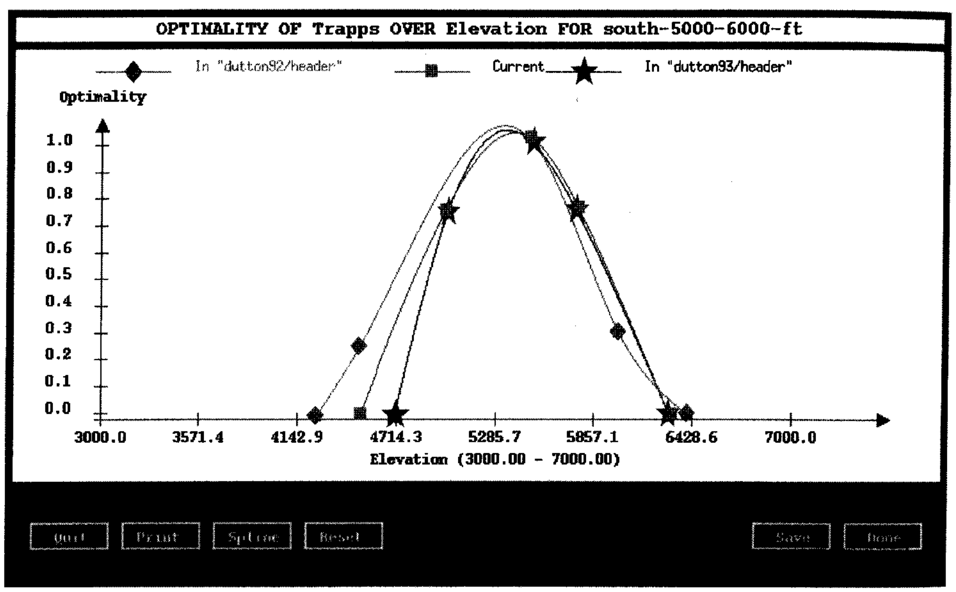

Figure 5. A graphical user interface for comparing optimality curves defined during different knowledge acquisition iterations. The curve with diamond nodes (dutton92/header) was defined in 1992 and the curve with star nodes (dutton93/header) was defined in 1993. The final version of the curve has nodes labelled as squares (current). Data were drawn from the case study discussed later in the text.

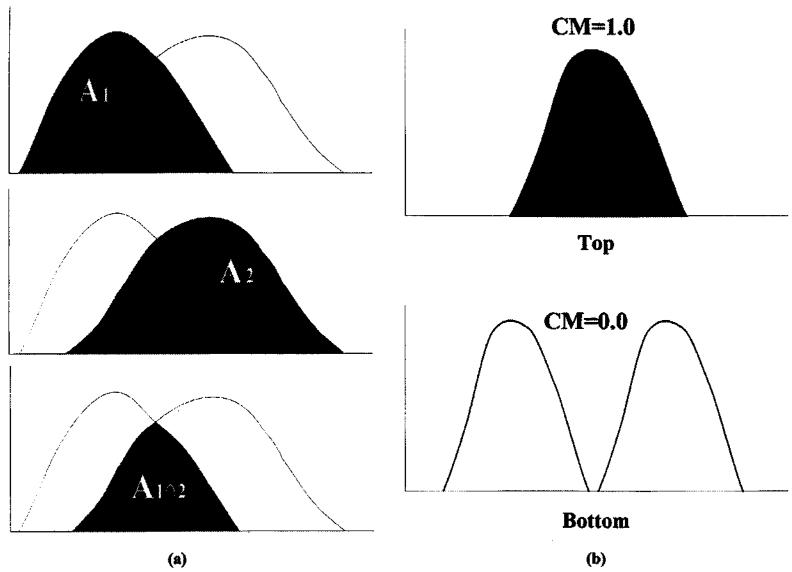

Figure 6. Definition of consistency measure (*CM*).

CM. Figure 7 shows three different degrees of matches between optimality curves along with their respective *CM* values (taken from the case study described below). The match in figure 7(*a*) is very good but that in figure 7(*c*) is poor. The match in figure 7(*b*) lies between the two. It is difficult to obtain a perfect match between two corresponding curves since a subtle difference between the two curves will result in a *CM* value less than 1. In practice, the consistency between corresponding optimality curves from two different iterations is good if the *CM* value between the two curves is greater than 0.8.

The consistency measure is not designed to remove or identify bias in the expert's knowledge but serves as an indirect measure of the significance of the final version of knowledge. It is assumed that if the different versions of knowledge extracted from different iterations are consistent then the final version is a good representation of the expert's understanding of the relationships.

4. Case study and results

A case study was conducted to illustrate the use and demonstrate the potential of the above knowledge acquisition process for resource mapping under the similarity model. The resource to be mapped is soil. The study area is the Lubrecht Experimental Forest which was established in 1937 to foster research on natural resources (Nimlos 1986). The area is about 50 km north-east of Missoula, Montana, USA and is in the mountainous terrain of western Montana with a moderate to strong relief. The climate is considered to be semi-arid to semi-humid. The soil expert was Barry Dutton, a certified soil scientist and the president of Land and Water Consulting, Inc. Mr. Dutton has conducted several soil field camps over a number of years in the Lubrecht Forest and is regarded as the expert in the soils of the area. The knowledge engineer was the author of this paper. The extracted knowledge was used to map the soil resources in this area under the similarity model using fuzzy inference (Zhu *et al.* 1996).

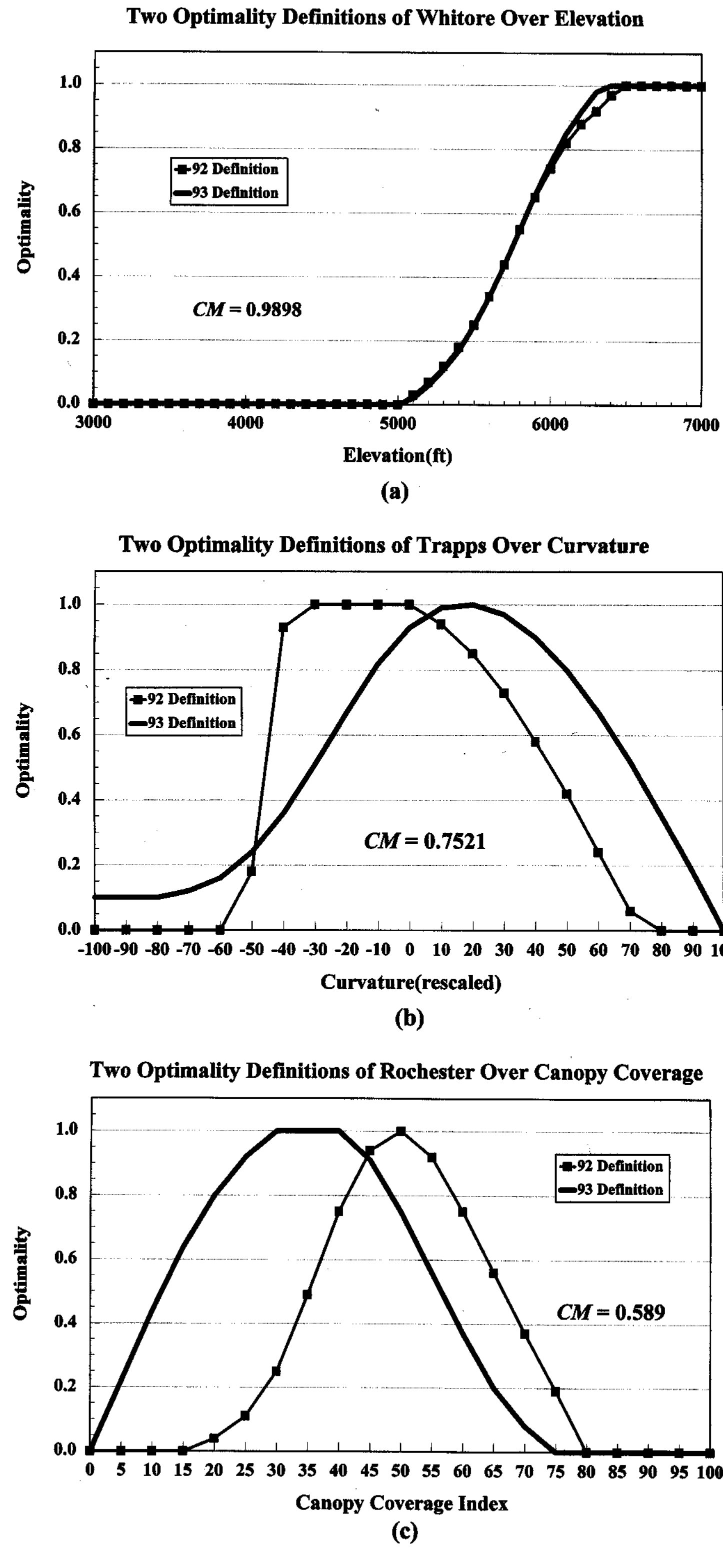

Figure 7. Different degrees of matches between two corresponding optimality definitions: (*a*) good; (*b*) average; (*c*) poor.

4.1. *Knowledge acquisition*

Two knowledge acquisition iterations were conducted a year apart. During the second iteration the expert did not have access to the information derived from the first iteration. During the preparation session, the soil category used for mapping was determined to be soil series since most soil maps are mapped at the soil series level and the expert was most comfortable working with soils at that level. Twelve soil series were identified for the area, and these are listed in table 1. Descriptions for these soil series can be found in *Soil Series Description* (MCSS 1983). These series were separated into three groups based on parent materials (table 1). The list of environment variables used in this case study resulted from a discussion between the expert and the engineer. Five environment variables were identified and used to distinguish the soil series within each of the three parent materials and served as the constructs for this knowledge acquisition exercise. These variables are: elevation, slope gradient, slope aspect, canopy coverage and profile curvature. The discussion also included measurement units for the environmental data and the standard units (e.g. feet) were chosen since they made more sense than the metric units to the soil expert. The unit for elevation was feet and the slope gradient was measured in percent. The range of curvature values was scaled to -100 to 100 with -100 being concave, 0 being straight, and 100 being convex.

After the determination of these environmental variables as constructs, the expert and the knowledge engineer defined the poles and the intermediate ratings (intervals) for each construct. The expert used the extreme values of a variable over this study area as the opposing poles of the corresponding construct. For example, if the elevation range is 3000 ft to 6000 ft, then one pole of the elevation construct would be 3000 ft and the other would be 6000 ft. The natural breaks in the range of values of the variable were used to define the intermediate ratings. Table 2 lists the constructs and their respective poles with intermediate ratings.

In the key development session, the expert used the defined ratings for each of the constructs to develop three keys (one for each of the three parent materials). To assist the expert in developing the keys, the knowledge engineer suggested that the expert use the most important environmental variable to first divide the environment into two types, use the second most important variable to divide each of these two types and, so on until the expert believed each of the resultant environmental types

Table 1. Soil series in the Lubrecht study area.

Soil series	Parent material	Soil subgroup
Evaro	Belt	Typic cryochrepts
Sharrott	Belt	Lithic ustochrepts
Tevis	Belt	Dystric eutrochrepts
Winkler	Belt	Udic ustochrepts
Winkler (cool)	Belt	Udic ustochrepts
Ambrant	Granite	Udic ustochrepts
Elkner	Granite	Typic cryochrepts
Ovando	Granite	Typic cryorthents
Rochester	Granite	Typic ustorthents
Repp	Limestone	Typic ustochrepts
Trapps	Limestone	Typic Eutroboralfs
Whitore	Limestone	Typic Cryochrepts

Table 2. Constructs and their poles and intermediate ratings for the Lubrecht area.

Constructs	Left pole	Intermediate ratings	Right pole
Elevation (ft)	3700	4000, 4500, 5000, 5500, 6000	6500
Aspect	North		South
Gradient (%)	Flat	5, 10, 15, 30, 60	Steep
Canopy coverage	Sparse	Medium	Dense
Profile curvature	Concave	Straight	Convex

to be occupied just by one soil series. A key to the environments of the soil series on the granite parent material is shown in figure 8.

During the environmental description session, the expert was asked to use the defined ratings to describe the typical environmental conditions under which a given soil series would exist. The soil expert was free to use descriptive terms compounded from the basic ratings (such as gradient between 15–60%, elevation <4500 ft). The typical environmental conditions for soil series Ambrant is listed in figure 9 as an example. Once the description session was completed, the results from this session were compared with the key developed earlier to check for consistency.

The optimality curves for parent materials were not defined by the expert but rather approximated by step functions. The optimality was assumed to be 1.0 for parent materials on which the series occurs and 0.0 for parent materials on which the series does not occur. During the optimality definition session, the soil expert was asked to define the optimality curves for the other five environmental variables. A total of 80 optimality curves were defined by the expert for each iteration. A total of 80 pairs of curves had been defined by the end of the second iteration. Two weeks after the completion of the second iteration, the expert was asked to compare these 80 pairs and derive a final version of optimality curves. The knowledge acquisition was complete when the final version was derived a week later.

Construct and Ratings	**Soil Series**
North facing	
>4,500 ft (1370 m)	
Gradient > 60%	**Ovando**
Gradient < 60%	**Elkner**
<4,500 ft (1370 m)	
Gradient > 60%	**Rochester**
Gradient < 60%	**Ambrant**
South facing	
>6,000 ft (1820 m)	
Gradient > 60%	**Ovando**
Gradient < 60%	**Elkner**
<6,000 ft (1820 m)	
Gradient > 60%	**Rochester**
Gradient < 60%	**Ambrant**

Figure 8. A key to the soil series on the granite parent material.

Soil Series: Ambrant		
Instance: 1		
	Parent Material:	**Granite**
	Elevation:	**3,700 - 4,500 ft**
	Aspect:	**North Facing**
	Gradient:	**15-60%**
	Canopy:	**Medium Coverage**
	Curvature:	**Straight to Convex**
Instance: 2		
	Parent Material:	**Granite**
	Elevation:	**3,700 - 6,000 ft**
	Aspect:	**South Facing**
	Gradient:	**15-60%**
	Canopy:	**Medium Coverage**
	Curvature:	**Straight to Convex**

Figure 9. Environmental descriptions for soil series Ambrant in the Lubrecht study area.

4.2. *Knowledge consistency analysis*

Measures of consistency (CM) between the two sets of optimality curves derived from the two iterations were calculated and analysed to assess the significance of the extracted knowledge. The mean of all 80 CM values is 0.836 with a standard deviation of 0.112. The frequency distribution of these 80 CM values is shown in figure 10 which shows that the distribution of CM values is heavily negatively skewed and the CM values are heavily clustered at the high value end. These general statistics for the CM values indicate that the two sets of optimality curves are generally consistent with each other and therefore the final version of the extracted knowledge is significant. However, this does not necessarily mean that the accuracy of the extracted knowledge is high since it is possible that an expert can provide a consistent answer that is incorrect or inaccurate.

Zhu (1997a,b) and Zhu *et al.* (1997) evaluated the quality of the final version of the extracted knowledge by examining the quality of soil spatial information derived from it. The extracted knowledge was used to populate the soil similarity model (Zhu *et al.* 1996). The populated model was then used to produce a soil map and a

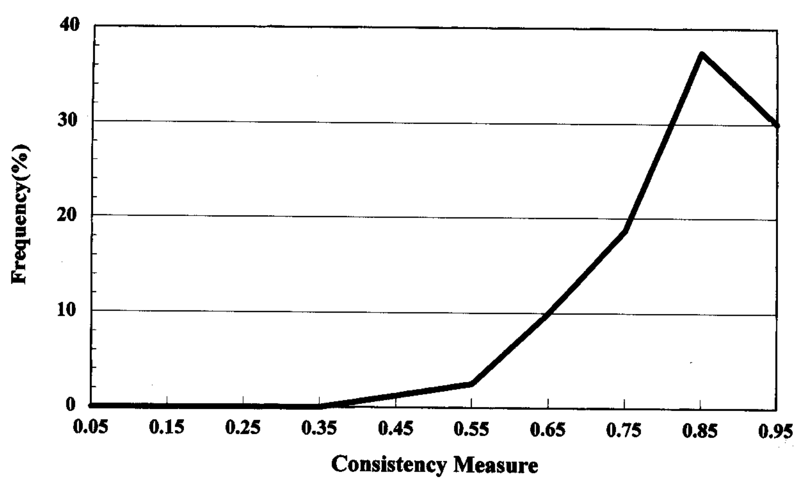

Figure 10. The distribution of CM values.

soil property map (Zhu *et al.* 1997). It was found that the soil map (inferred) contained much more spatial detail and was of higher quality than the conventional soil map (Zhu 1997a) and the derived soil property (*A*-horizon depth) information was more accurate than that from the conventional soil map (Zhu *et al.* 1997). It can be concluded from these study reports that the final version of the extracted knowledge is of good quality.

The *CM* values can be summarized by variables to determine if there is any relationship between the variables and the values of *CM*. Table 3 lists the statistics for the *CM* values grouped by environmental variables; that is, the consistency measures of all pairs of optimality curves related to a particular environmental variable are grouped together for computing these statistics. Both ***Aspect*** and ***Elevation*** have more pairs of optimality curves than the other variables because different instances of a soil series are differentiated only on the basis of aspect and elevation (figure 9). The average *CM* related to ***Elevation*** is high (0.89) and the associated standard deviation is small (0.098). This high *CM* average associated with a small standard deviation indicates a good overall consistency between the two sets of optimality curves related to ***Elevation***. The same can be said with regard to the consistency between the two sets of optimality curves related to ***Aspect***. However, the consistency between the two sets of optimality curves for ***Curvature*** and that for ***Canopy Coverage*** are both low as indicated by their respective low *CM* averages. The *CM* average for ***Gradient*** lies between these two general groups. ***Aspect***, ***Elevation***, and ***Gradient*** are the common variables used extensively to describe the environment of soils in soil surveys, particularly in mountainous terrains. Therefore, soil scientists often know more about the relationships between soils and these environmental variables. Although ***Canopy Coverage*** is also often used in soil surveys to describe the soil formative environment, the value for canopy coverage is normally estimated in the field and varies a great deal from person to person. Knowledge about soil-canopy coverage relationships is not as well defined as those for ***Aspect*** and ***Elevation***. It is clear that the soil expert was least confident about the relationships between soils and profile curvature since profile curvature is hardly used in conventional soil surveys.

The statistics about the *CM* values with respect to each soil series are listed in table 4. The soil expert was able to provide consistent answers about soil series that occur at high elevations (such as Evaro, Tevis, Ovando and Whitore) in terms of their relationships to the environment (relative high *CM* averages with small standard deviations within their respective groups). On the other hand, the expert had difficulty in singling out the soil-environmental relationships for soil series occurring at low elevations (such as Rochester, Repp and Winkler). It is true that at high elevations the environment is less heterogeneous than at low elevations in this semi-arid to

Table 3. Statistics about the *CM* values with respect to environmental variables.

Environment variables	Mean	Minimum	Maximum	Standard Deviation	No. of pairs
Aspect	0.8785	0.7471	0.9354	0.04243	22
Canopy	0.7649	0.4063	0.9641	0.16154	12
Curvature	0.7451	0.5777	0.8383	0.07336	12
Elevation	0.8928	0.6683	0.9899	0.09817	22
Gradient	0.8163	0.6432	0.9633	0.11103	12

Table 4. Statistics about the *CM* values with respect to soil series.

Soil series	Mean	Minimum	Maximum	Standard deviation
Belt soils				
Evaro	0.8642	0.7471	0.9748	0.09022
Sharrott	0.8658	0.7404	0.9774	0.09361
Tevis	0.887	0.7907	0.9641	0.07016
Winkler	0.8009	0.4063	0.9766	0.23215
Winkler cool	0.8618	0.6705	0.9649	0.10192
Granite soils				
Ambrant	0.823	0.6767	0.9003	0.09158
Elkner	0.8053	0.5777	0.9676	0.14119
Ovando	0.8293	0.6768	0.935	0.09038
Rochester	0.7374	0.589	0.8864	0.11852
Limestone soils				
Repp	0.7932	0.6768	0.9496	0.12412
Trapps	0.8649	0.7521	0.9336	0.06212
Whitore	0.8773	0.7361	0.9899	0.09611

semi-humid region. In particular, the soil series at high elevations tend to be spatially contiguous and the soil series at low elevations tend to be spatially intermittent. Understanding the relationships between spatially contiguous soil series and their environments would be easier than understanding the relationships between intermittently distributed soil series and their environments. Also, the Rochester soil series often occurs around small rock outcrops or along small spurious ridges and divides which could not be described well with the environmental variables employed. Therefore, it was difficult for the expert to establish the relationships between Rochester and these environmental variables. This may help to explain the low consistency measure between two sets of optimality curves related to the Rochester soil series (table 4).

5. Discussion

It can be concluded through this case study that the knowledge acquisition process was able to extract a good representation of a soil scientist's understanding about soil-environmental relationships in the Lubrecht area. The consistency measures between the knowledge sets obtained from different iterations also revealed the characteristics of the soil expert's understanding of soil-environmental relationships. The expert tended to have a better understanding of the relationships for some soil series than for others. The method also revealed that the expert was more knowledgeable about the relationships of soils to some environmental variables than those to other environmental variables.

The case study has demonstrated the use and potential of a personal construct theory-based knowledge acquisition process. This section is intended to discuss its scope of application and the key issues of the knowledge acquisition process.

This knowledge acquisition process was developed to extract knowledge about relationships between natural resources and their environment. Even though this knowledge is described as a set of membership functions, it is still considered as declarative as opposed to procedural (McGraw and Harbison-Briggs 1989, Armstrong 1991, Byrd *et al.* 1992). The knowledge acquisition process presented in this paper may have limited use for extracting procedural knowledge since the techniques needed for extracting procedural knowledge may be very different from those used to extract declarative knowledge (Moody *et al.* 1996). However, the knowledge acquisition process can be easily adapted to applications that require only descriptive knowledge from domain experts. In these cases, there would be no need to define optimality curves since they are needed only to compute similarity values under fuzzy logic.

Experience derived from the case study suggests that the knowledge engineer should have a working knowledge of the subject domain to achieve maximum efficiency in knowledge acquisition. This domain knowledge would expedite knowledge acquisition in several ways. First, it would help to clearly specify what knowledge would be required for mapping resources and help to clarify specifications for the expert so that he/she understands exactly what is needed. In the soil mapping example, the knowledge engineer used many examples in soil science to illustrate and clarify the definitions of the two types of knowledge. Second, the knowledge engineer's basic understanding of the subject domain would help to translate the personal construct concepts into the terminology of the subject domain and would ensure the correct use of these concepts during the knowledge acquisition process. Finally, the communication between a knowledge engineer and a domain expert in an interview is a scientific engagement. Conversations should keep the experts intellectually challenged and help them to reveal much of their understanding about the subject. The knowledge engineer's working knowledge of the domain would enhance this engagement.

Knowledge acquisition interviews can be very tense. Retrieving knowledge is not easy for any expert and experts are often frustrated by a knowledge engineer's questions. Agnew (1994) cites Marvin Minksy, one of the fathers of AI, as saying that 'we teach our children not to ask too many whys in a row ... not to even think of asking more than a few whys. And for good reason ... how many why's in a row can even a smart person handle?' Particularly, when the knowledge engineer points out inconsistencies in knowledge, the expert could feel embarrassed or even humiliated. Persistence is important in knowledge acquisition, but the knowledge engineer must be resourceful and use different ways to communicate with the expert and make the expert feel relaxed and comfortable during the entire knowledge acquisition process.

6. Conclusions

This paper presents a process based on personal construct theory for acquiring knowledge about resource-environment relationships and also presents a case study that demonstrates the use and potential of this process for extracting expertise from specialists. The case study shows that the process was successful not only in eliciting knowledge from a soil expert but also in revealing the characteristics of the expert's understanding about the relationships between soils and their environment. Using consistency analysis, it was found that the soil expert was more familiar with the relationships between soils and some environmental variables than with other environmental variables. The expert's knowledge on soil-environment relationships also differed by soil series.

The implication is that this knowledge acquisition process would be useful in acquiring knowledge about resource-environment relationships for mapping natural resources using the similarity model. It is recommended that when it is used to acquire knowledge from resource experts, this process should be performed by a knowledge engineer who has a working knowledge of the subject domain so that personal construct concepts can be translated into the domain terminology, and the communication between the knowledge engineer and the resource expert can be as smooth as possible. It is also recommended that knowledge engineers use the concepts of personal construct theory to guide themselves in the knowledge elicitation process, such as in the design of questions and in the organization of the interviews. They must not convey these concepts in their original forms to the expert since these concepts may confuse the expert. In fact, these terms or concepts should be translated into the terminology of the expert's domain so that the expert can use these concepts without being confused and the engineer can achieve maximum efficiency of knowledge acquisition.

Acknowledgments

The author wishes to thank Dr Lawrence Band for his encouragement and support in this research. Barry Dutton's cooperation in this research as a soil expert is greatly appreciated. The author also wishes to express his gratitude to the GIS Laboratory, School of Forestry, University of Montana for providing the DEM and remote sensing imagery of the study area. The start-up fund provided to the author by the Graduate School, University of Wisconsin-Madison made it possible to complete the work reported in this paper. I would also like to express my thanks to Professor Robert Sack for reading and commenting on an earlier draft of this paper. The author greatly appreciates the generous effort of Professor Marc Armstrong in editing this paper.

References

Adams-Webber, J. R., 1984, Personal construct theory. In *Encyclopedia of Psychology*, edited by R. Corsini (New York: Wiley-Interscience), pp. 9–12.

Agnew, N., 1994, Guest Editor's introduction: concepts of expertise. *International Journal of Expert Systems*, **7**, v–vii.

Armstrong, M. P., 1991, Knowledge classification and organization. In *Map Generalization: Making Rules for Knowledge Representation*, edited by B. P. Buttenfield and R. B. McMaster (England: Longman), pp. 86–102.

Band, L. E., and Moore, I. D., 1995, Scale: landscape attributes and geographical information systems. *Hydrological Processes*, **9**, 401–422.

Bench-Capon, T. J. M., 1990, *Knowledge Representation: An Approach to Artificial Intelligence* (New York: Academic Press).

Bezdek, J. C., Ehrlich, R., and Full, W., 1984, FCM: The fuzzy *c*-means clustering algorithm. *Computers & Geosciences*, **10**, 191–203.

Bregt, A. K., 1992, Processing of soil survey data, Doctoral Thesis (Wageningen, The Netherlands: Wageningen Agricultural University).

Burrough, P. A., 1989, Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science*, **40**, 477–492.

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

Byrd, T. A., Cossick, K. L., and Zmud, R. W., 1992, A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS Quarterly*, **16**, 117–138.

Coulson, R. N., Lovelady, C. N., Flamm, R. O., Spradling, S. L., and Saunders, M. C., 1991, Intelligent geographic information systems for natural resource management. In *Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity*, edited by M. Turner and R. H. Gardner (New York: Springer-Verlag), pp. 153–172.

CRAW, S., 1996, Refinement complements verification and validation. *International Journal of Human-Computer Studies*, **44**, 245–256.

FISHER, P., 1989, Knowledge-based approaches to determining and correcting areas of unreliability in geographic databases. In *Accuracy of Spatial Databases*, edited by M. Goodchild and S. Gopal (London: Taylor & Francis), pp. 45–54.

FISHER, P., and PATHIRANA, S., 1990, The evaluation of fuzzy membership of land cover classes in the suburban zone. *Remote Sensing of Environment*, **34**, 121–132.

FORD, K. M., PETRY, F. E., ADAMS-WEBBER, J. R., and CHANG, P. J., 1991, An approach to knowledge acquisition based on the structure of personal construct systems. *IEEE Transactions on Knowledge and Data Engineering*, **3**, 78–87.

GAMBLE, R. F., and BAUGHMAN, D. M., 1996, A methodology to incorporate formal methods in hybrid KBS verification. *International Journal of Human-Computer Studies*, **44**, 213–244.

GEORGIYEV, V. O., 1993, Subject domain knowledge representation models for interactive systems (a survey). *Journal of Computer and System Sciences International*, **31**, 1–18.

GOODCHILD, M. F., 1989, Modeling error in objects and fields. In *Accuracy of Spatial Databases*, edited by M. Goodchild and S. Gopal (London: Taylor & Francis), pp. 107–113.

GOODCHILD, M. F., 1992, Geographical data modeling. *Computers & Geosciences*, **18**, 401–408.

HAYES-ROTH, F., WATERMAN, D. A., and LENAT, D. B., 1983, *Building Expert Systems* (Reading, MA: Addison-Wesley).

HUDSON, B. D., 1990, Concepts of soil mapping and interpretation. *Soil Survey Horizon*, **31**, 63–73.

HUDSON, B. D., 1992, The soil survey as paradigm-based science. *Soil Science Society of America Journal*, **56**, 836–841.

HUSAIN, M., 1983, To what can one apply a construct. In *Applications of Personal Construct Theory*, edited by J. R. Adams-Webber and J. C. Mancuso (New York: Academic), pp. 11–28.

KELLY, G. A., 1955, *The Psychology of Personal Constructs* (New York: Norton).

KELLY, G. A., 1970, A brief introduction to personal construct theory. In *Perspectives in Personal Construct Theory*, edited by D. Bannister (London: Academic Press), pp. 1–29.

MARK, D. M., and CSILLAG, F., 1989, The nature of boundaries on 'Area-Class' maps. *Cartographica*, **26**, 65–78.

MCBRATNEY, A. B., and DE GRUIJTER, J. J., 1992, A continuum approach to soil classification by modified fuzzy k-means with extragrades. *Journal of Soil Science*, **43**, 159–175.

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

MCGRAW, K. L., and HARBISON-BRIGGS, K., 1989, *Knowledge Acquisition: Principles and Guidelines* (Englewood Cliffs, N.J.: Prentice Hall).

MENGSHOEL, O. J., and DELAB, S., 1993, Knowledge validation: principles and practice. *IEEE Expert*, **8**, 62–68.

MCSS (MISSOULA COUNTY SOIL SURVEY), 1983, *Soil Series Description* (Missoula, Montana: MCSS).

MOLOKOVA, O. S., 1993, Methodology for the acquisition of knowledge for expert systems: part 1. Basic concepts and definitions. *Journal of Computer and Systems Sciences International*, **31**, pp. 1–5.

MOODY, J. W., WILL, R. P., and BLANTON, J. E., 1996, Enhancing knowledge elicitation using the cognitive interview. *Expert Systems With Applications*, **10**, 127–133.

MULDER, J. A., and CORNS, I. G. W., 1996, Knowledge based ecosystem prediction: field testing and validation. In *GIS Applications in Natural Resources 2*, edited by M. Heit, H. D. Parker and A. Shortreid (Fort Collins: GIS World, Inc.), pp. 392–398.

NIMLOS, J. T., 1986, Soils of Lubrecht Experimental Forest. Miscellaneous Publication No. 44, Montana Forest and Conservation Experiment Station, School of Forestry, University of Montana, Missoula, Montana.

ODEH, I. O. A., MCBRATNEY, A. B., and CHITTLEBOROUGH, D. J., 1992, Soil pattern recognition with fuzzy c-means: application to classification and soil-landform interrelationships. *Soil Science Society of America Journal*, **56**, 505–516.

PLANT, R., 1996, Editorial: special issue on verification and validation. *International Journal of Human-Computer Studies*, **44**, 123–125.

Robinson, V., 1988, Implications of fuzzy set theory for geographic databases. *Computers, Environment and Urban Systems*, **12**, 89–98.

Robinson, V., and Frank, A., 1987, Expert systems for geographic information systems. *Photogrammetric Engineering & Remote Sensing*, **53**, 1435–1441.

Shaw, M. L. G., and Gaines, B. R., 1993, Personal construct psychology foundations for knowledge acquisition and representation. In *Knowledge Acquisition for Knowledge-Based Systems: Proceedings of 7th European Workshop, EKAW'93*, edited by N. A. G. Boy, B. R. Gaines, M. Linster, J. G. Ganascia and Y. Kodratoff (London: Springer-Verlag), pp. 255–276.

Skidmore, A. K., Watford, F., Luckananurug, P., and Ryan, P. J., 1996, An operational GIS expert system for mapping forest soils. *Photogrammetric Engineering & Remote Sensing*, **62**, pp. 501–511.

Skye, D., and Naia, ?., 1993, Ecologically-oriented spatial and knowledge-based framework to support forest and land resource management. In *Proceedings of GIS-93: 7th Annual Symposium on Geographic Information Systems in Forestry, Environment and Natural Resources Management* (Vancouver, B.C.: Ministry of Supply and Services Canada), pp. 161–166.

Wang, F., 1990, Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing*, **56**, 1163–1169.

Weibel, R., Keller, S., and Reichenbacher, T., 1995, Overcoming the knowledge acquisition bottleneck in map generalization: the role of interactive systems and computational intelligence. In *Spatial Information Theory: A Theoretical Basis for GIS*, edited by A. Frank and W. Kuhn (London: Springer-Verlag), pp. 139–156.

Zhu, A. X., 1997a, A similarity model for representing soil spatial information. *Geoderma*, **77**, 217–242.

Zhu, A. X., 1997b, Measuring uncertainty in class assignment for natural resource maps under a similarity model. *Photogrammetric Engineering & Remote Sensing*, **63**, 1195–1202.

Zhu, A. X., and 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., Dutton, B., and Nimlos, T., 1996, Automated soil inference under fuzzy logic. *Ecological Modelling*, **90**, 123–145.

Zhu, A. X., Band, L. E., Vertessy, R., and Dutton, B., 1997, Derivation of soil properties using a soil land inference model (SoLIM). *Soil Science Society of America Journal*, **61**, 523–533.