The Role of Attribute Selection in GIS Representations of the Biophysical Environment

Y. X. Deng* and J. P. Wilson†

*Department of Geography, Western Illinois University, Macomb, Illinois
†Department of Geography, University of Southern California, Los Angeles

This article questions the arbitrary selection of input attributes for the definition of landform classes and other fiat objects that are used to represent the biophysical environment in geographic information science. It suggests that attribute selection influences the characterization of both geographic and attribute space in these applications. Hence digital elevation model-based fuzzy *c*-means landform classification relies on sensible selection of terrain attributes to generate fuzzy landform classes (memberships) with biophysical meanings. A case study employed several sets of sensitivity tests and evaluated how selections of terrain attributes may affect the outputs of fuzzy *c*-means landform classifications. The results showed an average classification difference of 37 percent when different numbers of attributes are used and 18 percent when similar terrain attributes are swapped. Effects of attribute selection also show obvious dependence on spatial resolution and number of classes. These results indicate that the current approach of selecting terrain attributes (not only for landform classifications but also for other applications) according to tacit expert knowledge needs to be better justified. Because the fuzzy *c*-means classification method is essentially data-driven, the adoption of an exploratory approach as a part of this method is crucial. Such an approach may help to identify membership distributions (and corresponding classifications) that summarize the correspondence between landforms and specific biophysical patterns. *Key Words: attribute selection, environmental modeling, fuzzy c-means landform classification.*

Human decisions are inevitable in the process of summarizing and classifying landforms, and the one regarding how to select landform properties—or terrain attributes—is vital. Ideally, selection decisions would be based on the purpose of characterizing/classifying landforms and consequently determining the biophysical meanings of defined landform classes or membership patterns, although these meanings may not be obvious or easily interpretable. At present, landform properties are usually selected according to tacit expert knowledge (e.g., MacMillan et al. 2000; Burrough et al. 2001) with the underlying implication that the chosen combination of attributes is based on the best available knowledge. In many cases, however, this strategy of attribute selection is not well justified. Pike (1988) indicated that the parametric characterization and classification of landforms are synthetic, multivariate, and statistical, and he demonstrated diverse roles of terrain attributes in landslide studies using a discriminant analysis. Nevertheless, few landform studies to date have adopted discriminant analysis or some other type of sensitivity test to support attribute selection.

Numerous landform properties can be derived from a digital elevation model (DEM) to describe local terrain shape, surface location, or landscape context (Moore, Grayson, and Ladson 1991; Florinsky 1998; Wilson and Gallant 2000a; Pfeffer, Pebesma, and Burrough 2003; MacMillan, Jones, and McNabb 2004). Past efforts in modeling topographic controls on meteorological, hydrologic, pedologic, and ecological processes have established strong correlations between terrain attributes and the biophysical environment. For example, Park, McSweeney, and Lowery (2001) reviewed terrain attributes as key predictors of soil properties based on evidence from (1) statistical correlation between terrain attributes and field-measured soil attributes (McKenzie and Austin 1993; Moore, Gessler, et al. 1993; Gessler, et al. 1995), (2) classification of terrain attributes based on predefined criteria (Pennock, Zebarth, and de Jong 1987; Skidmore et al. 1991; Cook et al. 1996; Zhu et al. 1997), and (3) statistical clustering of terrain indices (Irvin, Ventura, and Slater 1997; de Bruin and Stein 1998; Lark 1999).

The existence of many useful terrain attributes and the fact that two or more attributes may affect the same biophysical process with different contributions (e.g., Pennock, Zebarth, and de Jong 1987) imply potential uncertainty in knowledge-based selection of terrain attributes. Multiple attribute combinations may seem

Annals of the Association of American Geographers, 96(1), 2006, pp. 47–63
Initial submission, May 2004; revised submission, April 2005; final acceptance, May 2005
Published by Blackwell Publishing, 350 Main Street, Malden, MA 02148, and 9600 Garsington Road, Oxford OX4 2DQ, U.K.

equally eligible in the eyes of an expert. This problem is particularly serious when the purpose of landform representation is general or vague. In this context, this article answers the following research question: How sensitive are the results of fuzzy *c*-means (as originally named by Bezdek 1974; also called *k*-means) landform classifications to varying selections (combinations) of commonly used terrain attributes? Using fuzzy *c*-means landform classification as an example is preferred because, as we describe in more detail later, this is a data-driven approach to the modeling of biophysical patterns. Consequently, the meanings of the resultant fuzzy classes (memberships) eventually depend on the way input data are organized.

In the next section we outline the theoretical background of the attribute selection issue by describing its role in conceptualization, definition, and representation of geographic objects and classes in GIScience. In that section we also review fuzzy *c*-means landform classification and provide a novel evaluation of this method as a synthesizing tool that can be used to explore the multivariate terrain attribute space. Then, in the following section, we report a case study that includes several sets of experiments to define the sensitivity of fuzzy *c*-means landform classification to (1) including or excluding specific terrain attributes, (2) swapping certain primary and secondary terrain attributes, and (3) the number of incorporated terrain attributes. The final section summarizes key findings and interprets their broader importance in the GIScience literature.

Background and Theory

The sharp contrast between precise individual properties and imprecise, comprehensive (e.g., synthesized or summarized) definitions of places, categories, and fiat objects (e.g., Fonte and Lodwick 2004) often confounds our cognition of the biophysical environment. Traditional soil and landform classifications offer typical examples—it is difficult to precisely define variability of soil attributes while respecting the authority of soil boundaries drawn in field-based soil surveys (Burrough 1993), and it is also difficult to identify a threshold slope gradient to distinguish mountains from no-mountains (Mark and Smith 2003, 2004), even though mountains imply higher slope gradients. Other related problems include (1) the disagreement between people on definitions of fiat objects and classes, and their boundaries (Smith and Mark 1998; see Omernik 1995 for another example); and (2) the distortion caused by applying categorical knowledge (about what is) to accidental predictions (about how it is, or the properties) of places (Smith and Mark 1998).

Most of the aforementioned problems in GIScience applications may be related to the fact that definitions of geographic objects are unsatisfactory (Fisher 2000; Robinson 2003). For example, much attention has been paid to the fact that crisp representations in GIS do not match indeterminate or nonexistent geographic boundaries in the real world (Burrough and Frank 1996), thereby revealing the inadequacy of traditional Boolean logic for the design of spatial databases for GIS (e.g., McBratney and Odeh 1997). Other reasons cited for poor definitions of geographic objects (Robinson 2003) are ontological confusion (Mark and Smith 2003) and insufficient knowledge of data quality and accuracy (Goodchild and Gopal 1989; Guptill and Morrison 1995; Unwin 1995). However, the issue regarding how attributes (i.e., properties) should be selected to define and characterize a spatial object has attracted relatively little attention.

We recognize attribute selection as a fundamental step in GIScience conceptualizations and representations of the biophysical environment. It identifies dimensions (variables) that are used to describe or characterize places or objects in attribute space, and is essential because how objects are delineated in geographic space depends on how they are represented in attribute space and on which dimensions are adopted to specify the attribute space. For example, whether soil boundaries are fuzzy or crisp depends on (1) whether continuous soil properties such as soil water content, instead of crisp ones such as soil class names, are used to describe soil distributions; and (2) whether gradual transitions of individual soil properties between data points are allowed in attribute space. By way of summary, the accuracy of geographic objects defined in GIScience may need to be measured not only in terms of spatial or attribute accuracy, but also in terms of relevancy of the selected properties or considered variables/dimensions of the attribute space.

The theory of fuzzy sets, which was first developed by Zadeh (1965), is useful in the context outlined above. Within the Earth sciences, this theory has been used to classify climatic data (McBratney and Moore 1985), geologic data (Bezdek, Ehrlich, and Full 1984), remote sensing images (Robinson and Thongs 1986; Fisher and Pathirana 1990), soil data (McBratney and de Gruijter 1992; Odeh, McBratney, and Chittleborough 1992; Zhu et al. 1997), and terrain data (Irvin, Ventura, and Slater 1997; Burrough, van Gaans, and MacMillan 2000; Gorsevski, Gessler, and Jankowski 2003). Fuzzy sets allow individual data points to have partial belongings (memberships) to multiple classes, and therefore allow for the existence of overlapped classes that can vary gradually from one to another. In this way, the spatial continuity existing in the biophysical environment can be represented without relying on crisp boundaries in geographic or attribute space.

Fuzzy *c*-means algorithms represent a more data-driven approach to the generation of fuzzy membership functions. This method was originally developed by Bezdek (1974, 1981) and Dunn (1973), then generalized by Bezdek, Ehrlich, and Full (1984). de Gruijter and McBratney (1988) and McBratney and de Gruijter (1992) later developed a modified objective function that accommodates outliers (i.e., extragrades) during classification. The method starts by randomly allocating data points into k clusters. The center of each cluster is calculated as the average (so-called *c*-mean) of point attribute values in the cluster. Next, k similarity indices are calculated for each data point based on the attribute distances of the data point to k cluster centers. According to these indices, each data point is reallocated to the most similar cluster. New cluster centers and new similarity indices can subsequently be calculated. This process continues until a stable solution (or a threshold) is reached and k stable cluster centers are established (Burrough and McDonnell 1998). Instead of calculating the maximum likelihood to determine whether a data point belongs to a cluster, the membership μ_{ik} of each data point i to class center k is calculated as a scalar value according to its similarity (or inverse attribute distance) to the class center. The most commonly used fuzzy *c*-means membership function for the calculation of μ_{ik} is written as

$$\mu_{ik} = \frac{[(d_{ik})^2]^{-1/(q-1)}}{\sum_{k'=1}^{c} [(d_{ik'})^2]^{-1/(q-1)}}, \qquad (1)$$

where d is the distance measure (calculated with a selected distance function) that is used to represent the similarity of a data point i to a class center k, c is the selected number of fuzzy classes, and q is the exponent of fuzziness (Burrough and McDonnell 1998). The parameter q falls between 1 and ∞; the larger q is, the fuzzier the classification is, and $q = 1$ produces a classification that is equivalent to a crisp classification, although it is not solvable with Equation (1). As a result, each kth class center (represented by a vector of attribute values)—instead of a class boundary—signifies the existence of a class, and μ_{ik} necessarily falls *between* zero (i has no similarity to k) and one (i is exactly the same as k), meaning that each data point can have partial belongings (memberships) to more than one class.

Considerable attention has been paid to the fact that q and c pose significant influence on the classification results (McBratney and Moore 1985; Robinson and Thongs 1986; Odeh, McBratney, and Chittleborough 1992; Burrough, van Gaans, and MacMillan 2000; Gorsevski, Gessler, and Jankowski 2003). In most cases q is determined empirically, which may involve a process of trial and error. This parameter is particularly important when we have identified a meaningful class center but need to delineate the way in which similarities of other data points to this class center decay across the space. In landform classification, for instance, q will influence how well the spatial continuity can be ideally reproduced when spatial covariations of cluster centers between landform and another biophysical pattern (e.g., soil, vegetation) have been detected. Indices such as the fuzzy performance index and normalized classification entropy (Roubens 1982; Burrough, van Gaans, and MacMillan 2000) can help in identifying the optimal partitioning scheme (e.g., number of classes) for the input data.

Burrough and McDonnell (1998, 270–91), in their discussion of fuzzy classification methods, contrasted the semantic import (SI) approach with fuzzy *c*-means classifications. SI is useful when the user has "a very good, qualitative idea" of possible class centers and the degree of fuzziness present in the data. As a result, the user can identify class centers and then fit a prespecified membership function to represent spatial continuity (fuzziness) and to draw fuzzy boundaries for the data, which would appear as crisp boundaries in traditional classifications. This approach was adopted in fuzzy soil classification by Zhu et al. (1997), who first identified typical soils (central soil concepts) based on expert knowledge or existing soil maps, and then mapped the soil spatial continuity according to the spatial continuity of topographic and other landscape properties.

The fuzzy *c*-means classification may be more useful for landform classification than the SI approach because human definitions of the central concepts of landforms are so diverse and inaccurate (e.g., hills and valleys; Mark and Smith 2003), making it impractical to identify agreeable central concepts prior to classification. Additionally, it is difficult to fit an existing membership function to often-heterogeneous landform patterns. The strengths of the fuzzy *c*-means method are therefore that (1) it can synthesize input terrain attributes and generate more "natural" cluster centers of landforms based on specific landscape and application situations, and (2) the spatial continuity (fuzziness) of landforms is represented by membership distributions calculated according to attribute distances between data points. Hence the classification method is more "objective" and data-driven. This also implies that the strengths of fuzzy *c*-means landform classifications rely primarily on the selection of terrain attributes, or organization of input data. Classification parameters (e.g., *q* and *c*) become important only when input attributes have been selected in a meaningful way.

An important characteristic of fuzzy *c*-means classification is that the continuous membership distributions of class centers are the key output. However, not necessarily all class centers of a classification are equally useful for a particular classification purpose, and the optimal partitioning scheme of data does not necessarily contain the class center that fits well with the application purpose. For example, if delineating "wetness" of places is the main purpose of classification, only two class centers need to be identified, "dry" and "wet", even though the optimal number of classes may be more than two. It is often the case in environmental modeling that individual biophysical properties, instead of the overall environmental structure, are of major concern (e.g., Beven and Kirkby 1979; Moore, Gessler, et al. 1993; Chuvieco et al. 2004). In this case, how to define one landform class center ("central" concept) and its membership distribution that could be used to map a biophysical property—instead of identifying the optimal partitioning scheme for the entire input data—may become the primary goal of landform classification. This again indicates that the parameter *c* in fuzzy *c*-means landform classification may not be critical in these situations.

There are two related uncertainty issues in fuzzy *c*-means landform classification. First, how to organize data or select terrain attributes in an optimal way is generally unknown, although candidate attributes can be listed based on the terrain–environment relations identified in previous work. Second, the meanings of landform classes (memberships) produced with the fuzzy *c*-means method need to be post-interpreted, instead of predefined. Knowledge of landform–environment relations is again useful in this interpretation process, but this knowledge can only support a qualitative, imprecise interpretation.

A "ground truth" pattern (e.g., obtained from field work and/or remote sensing data) that corresponds to the classification purpose will be needed to evaluate the classification accurately. When the result is unsatisfactory, the classification may need to be redone, perhaps with new attributes and parameters. In this case, an iterative "trial and error" procedure and, accordingly, an exploratory attitude toward landform classification, will be needed to produce a desired landform membership distribution (with appropriate class centers) that "matches" the ground truth pattern of interest.

A complete fuzzy *c*-means landform classification procedure may thereby require (1) a clearly defined and reasonably precise classification purpose to which this method is applicable; (2) availability of elevation and environmental data; and (3) an iterative evaluation procedure that can identify the most useful cluster center and membership distribution, as well as the corresponding attribute combination and classification parameters, from a series of competing classifications. At a time when many environmental properties still cannot be precisely measured across space, this procedure has the potential of defining application-oriented classification schemes that can help to extrapolate known biophysical patterns to other places (Burrough et al. 2001), or to interpolate these patterns to finer spatial resolutions (e.g., close to DEM resolution).

Concerns about the effects of spatial scale or resolution are ubiquitous in terrain analysis (Moore, Lewis, and Gallant 1993; J. Wood 1996; Wilson et al. 1998; Evans 2003; Schneider 2001) and terrain-based environmental modeling (Band 1986; E. C. Wood et al. 1988; Beven 1989, 1997; Moore, Lewis, and Gallant 1993; Zhang and Montgomery 1994; Band and Moore 1995; Quinn, Beven, and Lamb 1995; Florinsky and Kuryakova 2000; Kienzle 2004). For example, a common concern is that calculated terrain attributes (such as slope and topographic wetness index) are scale-dependent, and environmental models supported by these attributes may only be valid across a limited range of spatial scales (Zhang and Montgomery 1994; Band and Moore 1995; Florinsky and Kuryakova 2000; Kienzle 2004). The scale dependency of input terrain attributes implies that fuzzy *c*-means landform classification may also be scale-dependent. Consequently, the above-described exploratory classification procedure needs to be able to handle complications introduced by spatial scales and/or be applied at multiple scales, since the classification outputs may differ with varying spatial scales.

The fuzzy *c*-means landform classification method assumes the existence of spatial continuity and covariation of the spatial continuity between landforms and targeted biophysical properties. As a result, it may not be suitable for natural boundaries (e.g., cliffs, geologic faults, etc.) and discrete biophysical objects (e.g., volcanic craters, deep narrow gorges, etc.). It should also be used with caution when landforms do not play a significant influencing role (e.g., in spatial variation of managed or nutrition-stressed vegetation) or when the spatial variation of a biophysical pattern is controlled by a threshold process (e.g., at timberline). Another limitation of fuzzy *c*-means landform classification is that its data-driven nature implies heavy reliance on not only elevation data but also accurate biophysical data (ground truth).

The above literature review demonstrates that the analysis of uncertainty in attribute selection leads to the need to alter the existing expectations of fuzzy *c*-means landform classification. As a result, this classification method could be more than a landscape-partitioning tool and can be used to explore the data. Instead of producing optimal, deterministic, one-step-for-all classifications, it could serve as a multiscale iterative procedure that produces classification outputs for specific application purposes. Some additional work will be required to support such an approach. A continuing study by the current authors has automated the fuzzy *c*-means landform classification and subsequent pattern comparison procedure, and the case study reported in the next section uses a series of sensitivity tests to provide some primary support for this approach.

Case Study

Experimental Design

The case study reported in this section tests the sensitivity of fuzzy *c*-means landform classification to the selection/combination of terrain attributes, so as to provide evidence and support for an exploratory approach toward attribute selection. We believe this approach is necessary in not only fuzzy *c*-means landform classification but also in other efforts of delineating spatial objects/classes. The case study incorporated six commonly used primary terrain attributes: 1, elevation; 2, slope; 3, aspect (transformed value); 4, plan curvature; 5, profile curvature; and 6, upslope contributing area, as well as four secondary attributes: 7, distance to nearest ridgeline; 8, incoming solar radiation index; 9, topographic wetness index; and 10, sediment transport capacity index (Wilson and Gallant 2000b). Eight of these attributes (1, 2, 4, 5, 7, 8, 9, and 10) were also used by Burrough et al. (2001) in a fuzzy *c*-means landform classification of the Yellowstone National Park area. They selected these attributes based on expert knowledge and used them with a coarse-resolution DEM (100 m) to produce an "optimal" classification scheme of the data based on the fuzzy performance index and normalized classification entropy metric. After "defuzzifying" membership distributions into crisp landform classes, they demonstrated that the resultant class distributions corresponded well with remotely-sensed land cover patterns.

Table 1 uses the same attribute codes as presented above and lists eight pairs of terrain attribute groups. Fuzzy *c*-means landform classifications of these attribute groups were compared to evaluate: (1) the effects of including or excluding individual terrain attributes (comparisons a, b, and c); (2) the effects of swapping primary and secondary terrain attributes (comparisons d, e, and f); and (3) the effects of choosing different numbers of terrain attributes as inputs (comparisons g and h). In a later section we describe methods used in this study to measure differences between classifications. These differences were respectively summarized for each pair of terrain attribute groups; for three-class, four-class, and seven-class classification schemes; and at 10-, 20-, 100-, and 200-m resolutions to describe the sensitivity of fuzzy *c*-means landform classification to input attribute selection.

It was argued earlier that the classification parameters (e.g., q and c) are less important than input data organization. A practical difficulty in our case study is that the optimal number of classes may shift with changes in the input data. In this situation, the difference between two optimal classifications is the result of both input data change and variation in the number of classes. Since

Table 1. Comparison scheme for the case study of attribute selection

Comparison	First attribute group*	Second attribute group*	Comparison characteristic
a	1, 2, 4, 5, 7, 8, 9, and 10	1, 2, 5, 7, 8, 9, and 10	4 : in or out
b	1, 2, 4, 5, 7, 8, 9, and 10	1, 2, 4, 7, 8, 9, and 10	5 : in or out
c	1, 2, 4, 5, 7, 8, 9, and 10	1, 2, 4, 5, 8, 9, and 10	7 : in or out
d	1, 2, 4, 7, 8, 9, and 10	1, 2, 5, 7, 8, 9, and 10	swapping 4 and 5
e	1, 2, 4, 5, 7, 8, 9, and 10	1, 2, 4, 5, 6, 7, 8, and 10	swapping 9 and 6
f	1, 2, 4, 5, 7, 8, 9, and 10	1, 2, 3, 4, 5, 7, 9, and 10	swapping 8 and 3
g	1, 2, 4, 5, 7, 8, 9, and 10	1, 2, 4, and 5	secondary attributes: in or out
h	1, 2, 3, 4, 5, and 6	1, 2, 4, and 5	reducing primary attribute number

*1 = elevation, 2 = slope, 3 = aspect (transformed value), 4 = plan curvature, 5 = profile curvature, 6 = upslope contributing area, 7 = distance to nearest ridgeline, 8 = incoming solar radiation index, 9 = topographic wetness index, 10 = sediment transport capacity index.

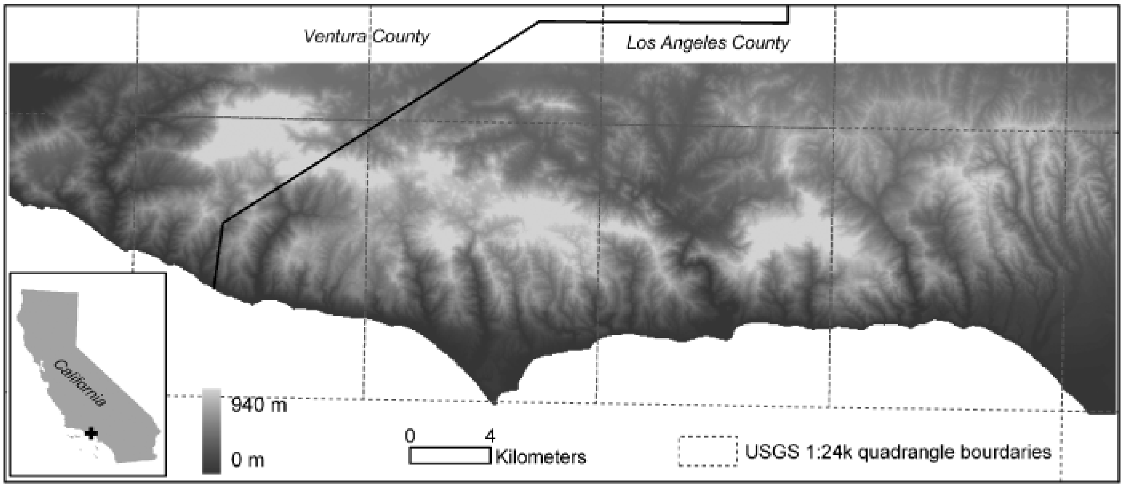

Figure 1. Santa Monica Mountains case study area.

useful outputs (e.g., cluster centers and membership distributions) of fuzzy *c*-means landform classification may not correspond to the optimal data partitioning scheme (see the previous section), we assume classification schemes with nonoptimal numbers of class centers can be meaningful too. This allows us to simplify the problem by (1) ignoring the question of whether the compared classifications represent the optimal data partitioning scheme or not, and (2) comparing classifications with varying input data but the same number of classes. As a consequence, each of the different groups of terrain attributes was respectively used to produce classifications containing three, four, and seven classes. And all comparisons (Table 1) were conducted between classifications with the same number of classes. We also adopted four spatial resolutions (10, 20, 100, and 200 m) for attribute calculation and classification to demonstrate how the effects of attribute selection behave across various spatial scales. The influence of parameter c on the comparison results was demonstrated by summarizing attribute selection effects across the three adopted numbers of classes. A uniform q of 1.5 was selected for all classifications. This choice was based on a visual comparison between terrain attribute patterns (elevation, slope, etc.) and membership distributions generated after choosing different q (e.g., 1.3, 1.5, 1.6, and 2). It was found that $q = 1.5$ produced spatial continuity of memberships that best matched the spatial continuity of terrain attributes.

Study Area

The case study area of about 670 km^2 covers the majority of the Santa Monica Mountains, California (Figure 1). The Santa Monica Mountains extend east–west along the Pacific coast and are characterized by medium to steep slopes and an elevation range of 0–940 m. Numerous short, narrow, parallel stream channels stretch from the central ridgeline north into inland valleys or south into the Pacific Ocean. Many short streams produced by the Mediterranean climate run seasonally along these channels, constitute the major erosive force, and cut deeply into the mountains. The high density, short length, and deep-cutting character of the channels mean that the terrain surface is relatively rugged (Table 2), and the stream network and landform structure hierarchies are reasonably well developed. The dominant vegetation is medium-high to tall chaparral shrub of varying densities (Stephenson and Calcarone 1999). The natural terrain surface is well preserved in most of the study area, but human disturbance is substantial in some locations. This study area contains coastal plains and various intact coastal mountain profiles to the central ridgeline (and peaks), and then to the inland valleys and plains. Its diverse, erosion-dominant landforms shaped by water flow and locally modified by various slope processes (e.g., landslides of various sizes

Table 2. Terrain characteristics of study area, summarized at 10 m

	Elevation (m)	Slope (%)	Roughness (m)*
Mean	325.7	41.34	17.34
Standard deviation	164.78	22.94	7.93
Maximum	948.2	331.59	60.53
Minimum	0.1	0	0

*Roughness was calculated as the local standard deviation of elevations with a circular moving window that has a uniform radius of 110 m.

and depths) are common in mountainous areas of the western United States. In consequence, the results of sensitivity tests obtained from the case study are likely to apply to other mountainous landscapes in the western United States.

Data and Tools

United States Geologic Survey (USGS) 10 m DEMs for roughly five 1:24 K USGS map quadrangle areas were downloaded from the GIS Data Depot and resampled in ArcGIS to 20-, 100-, and 200-m spatial resolutions using the nearest neighbor method (with a subsequent grid shift), so that both the original elevation values and their grid point locations could be preserved. Three sets of tools were utilized. ArcGIS was used for data preprocessing, map calculations, and spatial comparisons. PCRaster (Karssenberg et al. 2001) was used to calculate terrain attributes. The FNX730 program written by Simon Vriend and Pauline van Gaans of Utrecht University in the Netherlands, based on the fuzzy *c*-means algorithms of Bezdek, Ehrlich, and Full (1984) and McBratney and de Gruijter (1992), was used to produce fuzzy landform classes.

Calculation of Terrain Attributes

The ten terrain attributes listed earlier were calculated using PCRaster at 10-, 20-, 100-, and 200-m spatial resolutions. The same algorithms were used for the calculation of these attributes throughout the study although the details of these algorithms are not described in this article (for additional details, see PCRaster Version 2 Manual). However, the following details need to be clarified for this research:

1. Elevation values at all resolutions are the original (grid) point elevation records from the USGS 10-m DEMs.
2. The method suggested by Copland (1998) was used to convert measurements of aspect into a new variable A, where

$$A = \cos(\text{aspect}), \tag{2}$$

   such that values of A range from -1 to 1 and represent the extent to which a slope faces north ($A = 1$) or south ($A = -1$). The A values were used in place of aspect to represent the amount of incoming solar radiation in some comparisons. This conversion is necessary because aspect is calculated as circular degrees clockwise from 0° to 360° and is difficult to incorporate directly into classifications and comparisons.
3. Plan and profile curvatures were both multiplied by one hundred because their values have a very small magnitude. However, this does not influence the classification results since variances of the attributes are changed correspondingly.
4. Upslope contributing area and distance to the nearest ridgeline were calculated with the D8 flow-routing algorithm (O'Callaghan and Mark 1984).
5. The topographic wetness and sediment transport capacity indices were calculated from slope and upslope contributing area using simplified methods described by Wilson and Gallant (2000b). The natural logarithm of the sediment transport capacity index was used for the classifications and comparisons.
6. The incoming solar radiation was calculated in a pair of PCRaster programs that first calculated the relief and topographic shading effects from elevation, slope gradient, aspect, and sun height (according to latitude), and then used this information to calculate and accumulate daily incoming solar radiation for each grid point.

Correlation analysis between terrain attributes was performed at the four resolutions using 699 random sample points (described in the next section). This analysis was designed to define the extent of correlation between the adopted attributes, and to reveal possible reasons for differences between classifications performed.

Spatial Sampling and Fuzzy *c*-Means Landform Classification

A total of 699 points were randomly sampled and used to develop each of the landform classifications. The locations of these sample points were identified and matched at each of the tested resolutions. These sample points were very sparse (less than one sample per km^2), but they may represent as many as 699 situations (potential cluster centers) in attribute space (i.e., 699 samples in a 940-m elevation range, or more than one sample for every 2-m relief). Therefore, use of these 699 samples may sufficiently capture strong attribute clusters (and corresponding landform patterns) that repetitively occur in attribute and geographic space, which is an outcome that in most cases well serves the classification purpose.

Terrain attributes at the sampling points were extracted and saved in a dBase table as input for the fuzzy classifier. The weights of terrain attributes were not varied in the classification, and the diagonal norm distance measure, which transforms all classified terrain attributes to the same magnitude using their sample variances, was used. For all of the classifications we used the fuzzy *c*-means classification algorithm by McBratney and de Gruijter (1992), which calculates an extragrade class to diminish the influence of aberrant values. At the last step, the classification obtained from the sample points was assigned to the entire study area in PCRaster. This was accomplished by calculating attribute distances between each data point and each of the class centers using the diagonal norm distance function. Consequently, each cell has a membership value (ranging from zero to one), corresponding to the inverse attribute distance, for each class in the classification, representing its similarity to the class center. The entire procedure was repeated at four spatial resolutions to represent attribute effects at different resolutions. A total of ninety-six classifications were generated for comparison.

Comparison Methods and Procedures

The possible landscape meaning of the derived landform classifications was demonstrated (visualized) with an example landform class. First, attributes of the landform class center were interpreted with reference to the other class centers and the landform pattern present in the DEM. The same example was used to visualize how different attribute selections may influence the membership distributions and class centers.

A two-step method was designed to compare the spatial patterns of the classification results. The first step identified similar class-pairs with the diagonal norm distance function, which calculates attribute distances between class centers of two classifications i and j as

$$d_{ij} = \sum_k \left[(a_{ik} - a_{jk})^2 \times \frac{1}{\sigma_k^2} \right], \tag{3}$$

where d_{ij} is the attribute distance between two respective class centers in two classifications (i and j), k signifies the kth common attribute used by the two classifications, and σ_k^2 is the sample variance of the kth common attribute. σ_k^2 was used to convert various attribute values to the same magnitude so that all attributes have the same weight in the calculation of attribute distance. After this first step, a class in one classification could be matched to a certain class in another classification to form a pair of most similar classes that have the minimum d_{ij}, or the minimum accumulated attribute distance, between the two class centers in attribute space.

The second step incorporates three tasks to measure the similarity between both similar classes and pairs of classifications. First, a relative membership difference surface was generated for similar class-pairs in two classifications (i and j) by calculating

$$\mu_{ij} = \frac{|\mu_i - \mu_j|}{\mu_i + \mu_j} \times 100\%, \tag{4}$$

where cell-by-cell values of μ_{ij} are the relative difference between two corresponding memberships μ_i and μ_j of two class centers i and j. As a result, the value of μ_{ij} ranges from 0 to 100 percent, indicating the percentage difference between these two memberships. Second, the mean and standard deviation of this difference surface were calculated over the entire study area to represent how statistically different two membership surfaces are. Lastly, the means of the membership difference surfaces for all similar class-pairs of the two classifications were averaged to represent the overall difference between two classifications.

Results

Attribute Correlations. Pearson correlation analyses between all terrain attributes used in this study were performed at the 699 sample points and correlations between most attributes proved to be weak. This result is important to the classification because strong correlations imply overlapped representations of certain correlated attributes in the classification, which may consequently cause and hide biased (i.e., nonuniform) assignments of weights for these attributes even though uniform weights for all attributes are implicitly assumed. The correlation analysis results obtained at the 10-m spatial resolution are listed in Table 3. At this and all the other tested spatial resolution(s), the most significant correlation (−0.81 at a 10-m spatial resolution) is identified between transformed aspect A and incoming solar radiation, and the classifications obtained by swapping these two attributes have been incorporated in the subsequent comparisons.

The comparatively strong correlations observed at the 10-m spatial resolution (bold numbers in Table 3) did not show obvious change, or trend of change, over other spatial resolutions. From finer resolutions (10 and 20 m) to coarser resolutions (100 and 200 m), nevertheless, a small yet consistent increase of correlations was observed between some other pairs of terrain attributes (Table 4). These changing correlations indicate that (1) spatial scales may impact the computation and interpretation of classes incorporating the same group of attributes, and (2) the presumed uniform attribute weights at one scale may be shifted to nonuniform weights at another scale.

Table 3. Pearson correlation coefficients between terrain attributes calculated at 10-m spatial resolution

Attributes*	1	2	3	4	5	6	7	8	9
2	0.242								
3	0.125	0.044							
4	– 0.030	0.020	– 0.021						
5	0.041	– 0.027	0.031	– 0.437					
6	– 0.108	0.086	0.059	0.335	– 0.199				
7	– 0.049	– 0.094	– 0.028	0.028	– 0.089	0.025			
8	0.002	– 0.239	**– 0.807**	– 0.049	0.052	– 0.095	0.006		
9	0.198	**0.591**	0.025	0.342	– 0.222	0.210	0.162	– 0.186	
10	– 0.259	**– 0.514**	– 0.055	0.414	– 0.246	0.361	0.374	0.071	0.015

Notes: Bold face indicates comparatively strong correlations.
*1 = elevation, 2 = slope, 3 = aspect (transformed value), 4 = plan curvature, 5 = profile curvature, 6 = upslope contributing area, 7 = distance to nearest ridgeline, 8 = incoming solar radiation index, 9 = topographic wetness index, 10 = sediment transport capacity index.

Fuzzy Landform Classes. Tables 5 and 6 present class centers for seven-class classifications at the 10-m spatial resolution, using eight and four terrain attributes, respectively. In both cases, the number of classes (seven) did not correspond to the optimal partitioning scheme of the input data. The central concept represented by Class Center Four (bold numbers) in Table 5 is valley bottoms of various sizes where water flow converges. This is evident because (1) the plan curvature is the highest, indicating strong concave contour patterns of narrow valleys in this class; (2) the proximity to the nearest ridgeline is comparatively far; (3) the mean topographic wetness index is the highest; and (4) the mean sediment transport capacity index is the highest.

Figure 2B can be used to verify the above interpretation. A hard class map for Class Four in Table 5 was derived by selecting cells whose highest membership belongs to Class Center Four. This hard class map was draped on top of the 10-m DEM that had been used to generate the classification. Strong correspondence is obvious between DEM-displayed valley locations and the hard class map for Class Center Four.

The most useful information produced by this classification is presented in Figure 2A, which is the membership distribution for Class Center Four in Table 5. If the interpretation of this class center is correct, this map shows precisely how similar each point over the landscape is to channels or valley bottoms, from the exact same (membership = 1) to a total difference (membership = 0). To go further, this information would be more useful in the situation where we can identify certain biophysical phenomena that only (or mostly) occur in channels or valley bottoms. Then the membership map can be used to describe the likelihood that these phenomena occur at each location (e.g., point) across the landscape.

Table 6 presents the result of the four-attribute, seven-class classification. Because secondary terrain attributes that directly describe some biophysical processes (Wilson and Gallant 2000b) were not used for this classification, the class centers were more difficult to interpret than in the case of Table 5. However, the highest mean plan curvature, a typical indicator of channel or valley bottom shape, was present in Class Center Three, indicating that this class center most likely represents a similar class to valley bottoms. This conclusion was verified by the hard class map of Class Center Three shown in Figure 2D, which is also draped on top of 10-m DEM.

A visual comparison between Class Four in the eight-attribute classification (Figures 2A and 2B) and Class Three in the four-attribute classification (Figures 2C and 2D) indicates these two classes are a pair of similar classes belonging to two different classifications. However, the change in attribute selection between these two

Table 4. Change of Pearson correlation coefficients over different spatial resolutions between pairs of terrain attributes

Attributes*	10 m	20 m	100 m	200 m
4 vs. 5	–0.437	–0.503	–0.577	–0.627
4 vs. 7	0.335	0.412	0.542	0.598
5 vs. 7	–0.199	–0.321	–0.498	–0.570
5 vs. 9	–0.222	–0.338	–0.391	–0.428
5 vs. 10	–0.246	–0.276	–0.375	–0.416
6 vs. 9	0.162	0.214	0.378	0.464
7 vs. 9	0.361	0.444	0.558	0.593

*1 = elevation, 2 = slope, 3 = aspect (transformed value), 4 = plan curvature, 5 = profile curvature, 6 = upslope contributing area, 7 = distance to nearest ridgeline, 8 = incoming solar radiation index, 9 = topographic wetness index, 10 = sediment transport capacity index.

Table 5. Landform class centers for an eight-attribute, seven-center classification at 10 m. Class Center 4 represents a similar class to Class 3 in Table 6

Attributes*	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6	Center 7
1	158.8	289.7	309.3	**312.1**	369.8	385.2	404.8
2	7.4%	30.3%	49.0%	**35.0%**	52.7%	41.4%	54.7%
4	–0.11	–0.02	0.54	**2.26**	–0.24	–1.46	–0.32
5	–0.04	–0.29	–0.67	**–1.52**	0.26	0.67	0.03
7	10.02	15.15	26.34	**20.18**	12.24	3.76	13.17
8	1075.6	1027.5	982.6	**1029.6**	1111.6	1038.0	724.8
9	9.37	7.74	7.56	**10.31**	6.64	5.98	6.57
10	0.77	3.25	4.06	**4.98**	3.73	2.85	3.77

Notes: Bold face indicates the central concept represented by Class Center Four.
*1 = elevation, 2 = slope, 3 = aspect (transformed value), 4 = plan curvature, 5 = profile curvature, 6 = upslope contributing area, 7 = distance to nearest ridgeline, 8 = incoming solar radiation index, 9 = topographic wetness index, 10 = sediment transport capacity index.

classifications generated noticeable changes in both the membership distribution and hard class maps. A more complete and spatially consistent depiction of channel and valley patterns is provided by the eight-attribute classification, because the two secondary terrain attributes it used—the topographic wetness and sediment transport capacity indices—incorporated the effects of water flow accumulation. This process represents a typical case of nonneighbor relationships across the landscape. In contrast, the four-attribute classification depicted more "patchy" channel/valley patterns because the terrain attributes used in this classification only describe local morphology (slope, and plan and profile curvatures) or point characteristics (elevation) of the terrain surface. However, it is encouraging to observe that the four-attribute classification still produced generally similar channel/valley patterns to those depicted by the eight-attribute classification, indicating that landform patterns of this area show strong self-organization, so that clusters of various terrain attributes would co-occur over the landscape.

Evaluation of comparison methods. Table 7 gives an example of the results that were obtained with the first step adopted to compare the classifications. Pairs of class centers with the shortest attribute distances were easily detected (bold numbers in Table 7). This proved to be an accurate way of measuring differences between class centers in attribute space. Using this method, pairs of similar classes can be identified between classifications even though the classifications are vastly different in general. This capacity is also useful for identifying all interesting cluster centers for an exploratory study of landforms. However, this step is essentially nonspatial and a further step is required for accurate comparison of spatial patterns.

Table 8 gives an example of class-to-class membership differences between two classifications, which can be readily linked to membership maps of compared classes (e.g., Figures 2A and 2C). This step gives the most informative comparison because its cell-by-cell comparisons can accurately report differences between spatially continuous patterns. However, although membership maps may represent the most accurate way of visualizing the results of fuzzy *c*-means classifications, the overall comparison of the entire classifications is very difficult to visualize with this final method.

Summary of comparison results. Table 9 lists the results of the eight comparisons averaged over four spatial

Table 6. Landform class centers for a four-attribute, seven-center classification at 10 m. Class Center 3 (as indicated by bold numbers) represents a similar class to Class 4 in Table 5

Attributes*	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6	Center 7
1	94.0	270.7	**302.3**	302.7	345.2	453.7	533.3
2	13.5%	58.1%	**51.1%**	19.2%	57.7%	42.7%	54.2%
4	–0.06	–0.11	**2.87**	–0.03	–1.63	–0.56	1.04
5	–0.17	–0.35	**–3.23**	–0.19	1.62	0.50	–0.71

*1 = elevation, 2 = slope, 3 = aspect (transformed value), 4 = plan curvature, 5 = profile curvature, 6 = upslope contributing area, 7 = distance to nearest ridgeline, 8 = incoming solar radiation index, 9 = topographic wetness index, 10 = sediment transport capacity index.

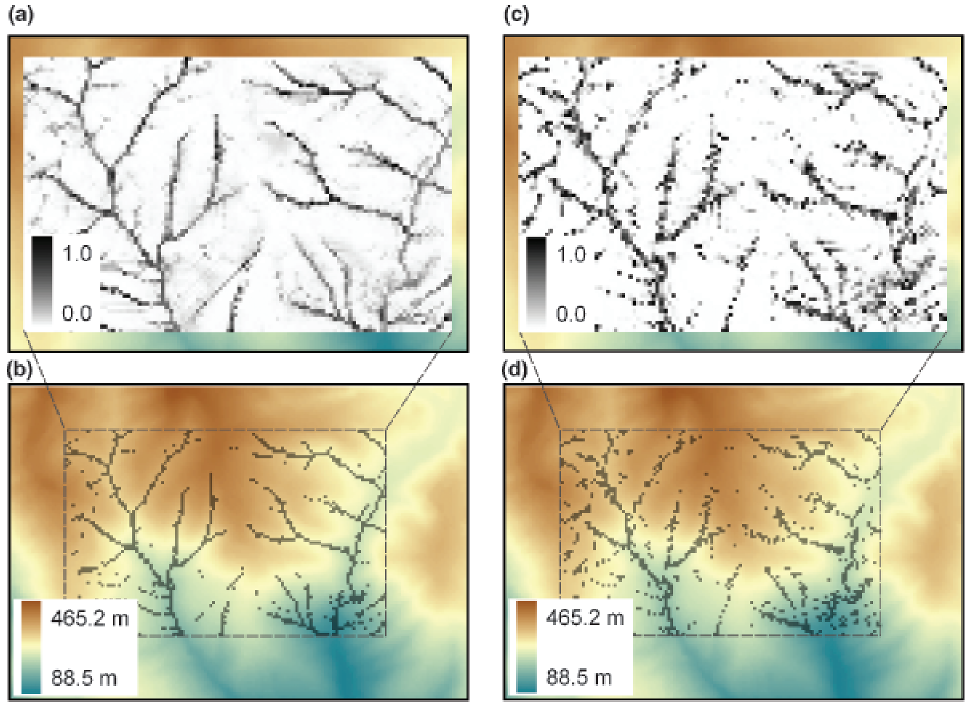

Figure 2. Classification results (and differences) in a small area, corresponding to Class Centers 4 and 3 in Tables 5 and 6. The maps show: (a) membership distribution of Class Center 4 in Table 5, (b) hard class map for Class 4 in Table 5 draped on 10-m DEM, (c) membership distribution of Class Center 3 in Table 6, and (d) hard class map for Class 3 in Table 6 draped on 10 m DEM.

resolutions (10, 20, 100, and 200 m) and three classification schemes (three-class, four-class, and seven-class classifications). Average membership differences (defined previously in the "Comparison Methods and Procedures" section) of 15 to 39 percent were generated in the fuzzy *c*-means landform classification results when attribute selection was varied for the classification. Including/excluding terrain attributes or swapping terrain attributes produced relatively small impacts on the classification (an average membership difference of 18 percent). In contrast, the number of input terrain attributes had a much stronger influence (an average membership difference of 37 percent).

Among the tested terrain attributes, profile curvature seems to have the weakest impact on the classification results (with a membership difference of roughly 16 percent, see comparisons b and d), whereas distance to nearest ridgeline shows the strongest influence (with a membership difference of 23 percent). Although incoming solar radiation is significantly correlated with transformed aspect A (Table 3), swapping these two attributes still produced obvious differences (with a membership difference of 15 percent on average), presumably because the solar radiation index incorporated the effects of topographic shading.

Overall, membership differences are characterized by low mean values and relatively high standard deviation values, and the same characteristic is observed between most of the individual comparisons of classification pairs. This indicates the coexistence of (1) a significant number of cells with aberrant, large difference values, indicating these cells are very sensitive to the adjustment of

Table 7. Attribute distances between class centers of two 4-class classifications at 20-m spatial resolution: four-attribute classification (C_{41}–C_{44}) versus eight-attribute classification (C_{81}–C_{84})

	C_{41}	C_{42}	C_{43}	C_{44}
C_{81}	**0.005**	4.599	4.583	4.629
C_{82}	3.123	**0.172**	6.479	2.904
C_{83}	3.854	2.599	1.578	**0.320**
C_{84}	3.615	6.807	**0.248**	1.305

Notes: Bold face indicates the pairs of class centers with the shortest attribute distances.

Table 8. Membership differences between two 4-class classifications at 20-m spatial resolution: four-attribute classification (C_{41}–C_{44}) versus eight-attribute classification (C_{81}–C_{84})

	C_{41}	C_{42}	C_{43}	C_{44}
C_{81}	**26.96%**	59.78%	61.77%	61.01%
C_{82}	58.87%	**33.73%**	65.28%	58.72%
C_{83}	64.07%	56.80%	53.56%	**38.09%**
C_{84}	61.25%	68.53%	**36.52%**	49.10%

Notes: Bold face indicates the pairs of class centers with the smallest membership differences.

Table 9. Membership differences caused by varying the attribute selections in a fuzzy *c*-means landform classification. The results presented are averaged over four spatial resolutions (10, 20, 100, and 200 m) and three classification schemes (3, 4, and 7 classes)

		Membership difference	
	Comparisons*	Mean	Standard deviation
Inclusion/exclusion	a	18.39%	16.35%
	b	16.13%	15.64%
	c	22.68%	19.65%
Exchange	d	16.50%	17.39%
	e	18.25%	15.61%
	f	14.80%	14.72%
Number of attributes	g	34.76%	26.86%
	h	39.20%	27.60%

*See Table 1 for comparison schemes.

terrain attributes (also see Figures 2A and 2C); and (2) a large number of cells with very low membership differences. This interpretation indicates that the effects of attribute selection vary over landscapes.

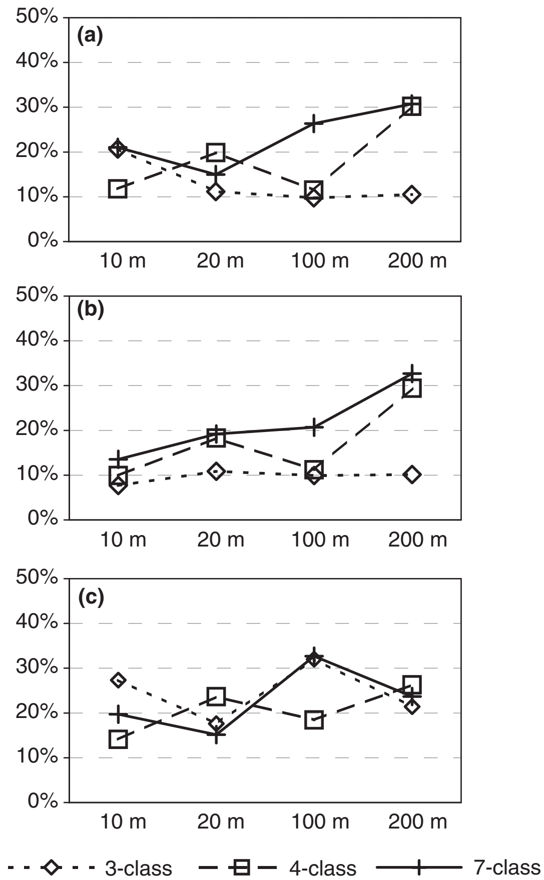

Figure 3. Percent membership differences over four resolutions and three classification schemes caused by including or excluding individual terrain attributes: (a) plan curvature, (b) profile curvature, (c) distance to nearest ridgeline.

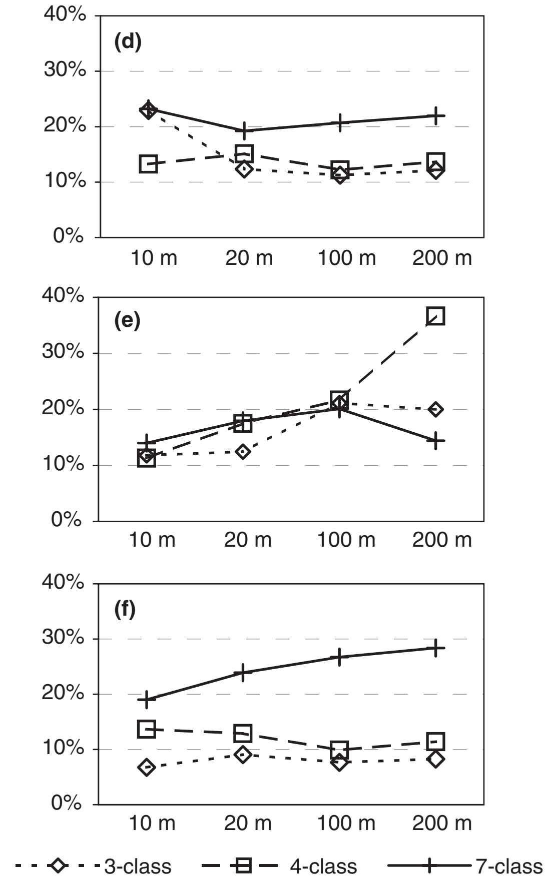

Figure 4. Percent membership differences over four resolutions and three classification schemes caused by swapping terrain attributes: (d) plan curvature versus profile curvature, (e) topographic wetness index versus upslope contributing area, (f) incoming solar radiation versus transformed aspect A.

Impact of spatial resolution and number of classes. Figures 3, 4, and 5 group the results of the eight comparisons based on the three types of comparisons: for individual attribute inclusion/exclusion, for attribute exchanges, and for the number of attributes. Each point in the graphs represents the comparison of one particular pair of classifications. The variation of spatial resolutions caused membership differences of up to 15 percent. This result indicates that the same adjustment of attributes could cause very different classification effects at different spatial resolutions; hence the need to consider spatial resolutions when adjusting or selecting terrain attributes for fuzzy *c*-means landform classifications. Figures 3, 4, and 5 also show differences of up to 25 percent between different predefined numbers of classes in terms of the effect of attribute selection on classification results. For example, including or excluding plan curvature at 200 m produced a 10 percent change in membership differences for three-class classifications, but a 31 percent difference for four- and seven-class classifications.

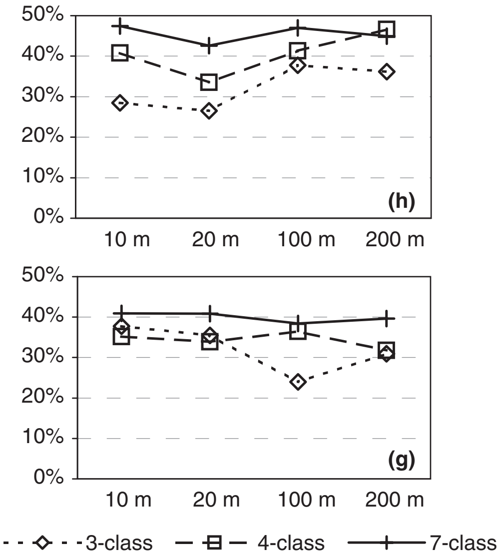

Figure 5. Percent membership differences over four resolutions and three classification schemes caused by number of terrain attributes: (g) four attributes versus eight attributes, (h) four attributes versus six attributes.

Discussions and Conclusions

Discussions

An essential argument presented in this article is that the continuous terrain surface can be conceptualized and classified in numerous ways (Pike 1988), especially when the classified terrain attributes are reorganized or regrouped. However, we are by no means the first to realize the importance of attribute selection. Burrough et al. (2001), for example, acknowledged that the eight attributes they used might not produce the best classification result and their attribute list could easily be extended, even though successful prediction of vegetation cover was accomplished by their landform classification. Pfeffer, Pebesma, and Burrough (2003) later used scatterplots, correlation analyses, and stepwise regression between terrain attributes and a series of vegetation scores to improve the process of attribute selection based on the same eight attributes. The regression residuals were then examined using variogram analysis to assist a universal kriging interpolation of the vegetation scores. A 50 to 60 percent success rate was achieved for the classification.

MacMillan et al. (2000) recognized selection and computation of terrain attributes as a key challenge for the representation and interpretation of landform elements. Their landform classification incorporated three groups of terrain attributes to represent topographic control on soil properties and precision agriculture: (1) slope, aspect, and plan and profile curvatures for their influence on the movement of water and materials; (2) topographic wetness index to capture the influence of soil moisture on crop growth; and (3) a series of elevation percentile and relief measures to capture the role of landform position in the formation of soil properties. However, significant discrepancies were observed when the classification results were compared with field-sampled soil properties and crop yield measurements of canola and wheat.

Table 10 lists some representative landform classifications that integrate various terrain attributes for general or specific classification purposes. Based on the

Table 10. A comparison of attribute selection and classification purposes of selected landform classifications

Author(s)	Attributes used for classification	Classification purpose(s)
Hammond (1964)	Percentage of gentle slopes, relief, general profile character, surface materials	Regional landform description
Pennock et al. (1987)	Slope, plan curvature, profile curvature	Soil morphological properties
Pike (1988)	Elevation, elevation variance spectrum, slope of slope features, slope calculated at regular intervals	Landslide mapping
Dikau (1989)	Slope gradiant, aspect, plan curvature, profile curvature, size order (scale)	Land component mapping
Graff and Usery (1993)	Elevation, slope, critical points	Identification of mounts
Dymond et al. (1995)	Threshold elevation, slope, aspect	Land component mapping
Brabyn (1997)	Relief, slope, profile type	Regional landform description
Irvin et al. (1997)	Elevation, slope, tangent curvature, profile curvature, topographic wetness index, incident solar radiation	Soil-landscape analysis
MacMillan et al. (2000)	Four elevation percentiles, local relief, local maximum relief, slope, plan curvature, profile curvature, topographic wetness index	Precision farming
Burrough et al. (2001)	Elevation, slope, plan curvature, profile curvature, proximity to nearest ridgeline, topographic wetness index, sediment transport capacity index, incident solar radiation	Forest cover mapping
Gorsevski et al. (2003)	Elevation, slope, tangent curvature, profile curvature, topographic wetness index, incident solar radiation	Landslide mapping

literature, we summarize a few factors as key contributors for the difficulty of selecting terrain attributes in these and other landform classification efforts. First of all, meaningful attribute selection needs to be guided by a specific, precisely defined classification purpose. This factor explains the success of Graff and Usery (1993) in delineating landform mounts, or locally elevated areas, from nonmounts using three simple attributes: slope, elevation, and critical points. In this case, a clearly (narrowly) defined classification purpose allows a prototype of mounts to be defined. In contrast, clear guidance for attribute selection and agreeable "prototypes" of landform classes are less available for general-purpose landform classifications (e.g., Pennock, Zebarth, and de Jong 1987; Dikau 1989; Dymond, Derose, and Harmsworth 1995).

Second, there are numerous terrain attributes that can be computed, but a common practice in landform classification is to use expert knowledge and select/adopt a short list of attributes as potential candidates (Dymond, Derose, and Harmsworth 1995; Irvin, Ventura, and Slater 1997). MacMillan et al. (2000) emphasized the importance of including contextual terrain attributes that represent the topographic position of each classified point. Pike (1988) and Dikau (1989) indicated the need to consider not only cell- or point-based terrain attributes, but also terrain attributes of landform facets (objects) that are composed of groups of cells/points. The spatial scale, particularly resolution and size, is a fundamental component of these terrain attributes. However, because topographic position is directly related to biophysical processes such as soil erosion, and landform objects often exhibit internal uniformity of certain biophysical characteristics (Pike 1988; Dikau 1989), these two types of attributes are more than outcomes of coarsening DEM resolutions.

Third, terrain attributes are often correlated to each other (e.g., the overlap between topographic wetness index and sediment transport capacity index). Pike (1988) thereby strove to define a series of complementary attributes of topographic forms to ensure that the adopted terrain attributes are relatively independent of each other. However, a more significant issue may be that overlaps between terrain attributes (or differentiated attribute weights) need to correspond to the variable effects of terrain attributes on the biophysical property of interest.

Attribute selection is still an uncertain activity in cases in which we had a clearly defined classification purpose, a long ("complete") list of terrain attributes, and the desire to consider and incorporate variable attribute weights. This is because presently (if not permanently) we cannot define the cause-effect relationships between landform properties and biophysical properties. A similar predicament may exist in other geographic conceptualization, representation, and classification activities as well. All these require an up-front enthusiasm to deal with uncertainty in our knowledge and to adopt a flexible, exploratory attitude in our analysis. In fuzzy *c*-means landform classifications, this enthusiasm can be corralled and implemented with the iterative, exploratory procedure suggested earlier.

Conclusions

Attribute selection determines what properties will be used to characterize and define places, objects, and classes. Although its importance has been insufficiently addressed in the literature, it is vital to our geographic analyses. The case study showed that fuzzy *c*-means landform classification outputs depend not only on how many terrain attributes are used, but also on which attributes are used. The classification is very sensitive to the number of attributes, which caused a 37 percent difference on average, but swapping similar terrain attributes also caused an 18 percent difference on average. To make matters more complicated, various tested attributes showed different effects on the classification results such that the effects of terrain attribute selection are not uniform over different spatial resolutions or over different numbers of classes. All these results indicate that the current approach used to select terrain attributes according to tacit expert knowledge (Irvin, Ventura, and Slater 1997; Burrough et al. 2001) needs to be better justified.

The attribute selection issue has particularly important implications for data-driven (or DEM-based) landscape delineation approaches like the fuzzy *c*-means landform classification method. The types of adjustments in attribute selection examined in this study may disturb class membership distributions or shift definitions of landform class centers, and it will generally not be possible to determine in advance which selected attributes give the best classification result. This uncertainty of attribute selection implies the need for an exploratory approach when we adopt this classification method. The importance of such an approach is twofold: it will influence our ability to organize data for meaningful classification, and it will influence our ability to interpret the classification results. These two outcomes complement one other and, taken as a whole, indicate the need for an iterative procedure.

Consequently, an innovative view of fuzzy *c*-means landform classification is suggested. First, the membership distribution and corresponding class center of a particular landform class are identified, using an iterative "trial and error" procedure, as the "best match" to a particular biophysical pattern, which may correspond to a specific application purpose. Second, this membership distribution and its landform class center are emphasized as the key outputs of the landform classification. In a spatially continuous fashion, this output can potentially be used to delineate the correlated biophysical pattern in other places or at finer spatial resolutions (e.g., DEM resolution).

In all, the genetic control of attribute selection on the definitions of classes (or objects) in attribute and geographic space is articulated. It is demonstrated that uncertainty or difficulty pertaining to attribute selection is entangled with issues such as spatial continuity (or fuzziness of boundaries), spatial scale, and diversity of human purposes (Mark and Smith 2003, 2004). The analysis and case study presented in this article suggest that (1) optimal attribute selection should be based on well-defined classification (or object-definition) aims; and (2) an exploratory approach that challenges the traditional once-for-all, knowledge-based selection of attributes is vital for accurate descriptions of patterns and relationships existing in the biophysical environment.

Acknowledgments

Professors Peter Burrough and Simon Vriend from the Utrecht University of the Netherlands provided the fuzzy *k*-means classification programs. Conversations with Professor Michael Goodchild from the University of California, Santa Barbara, and Professor Bernard Bauer from Okanagan University College, Canada, stimulated the research ideas that are developed in this article. Comments from five anonymous reviewers helped tremendously in improving the first draft. This work was partly funded by Grant No. BCS#0425273 from the National Science Foundation's Geography and Regional Science Program.

References

ArcGIS, 9.0. Redlands, CA: Environmental Systems Research Institute.

Band, L. E. 1986. Topographic partition of watersheds with Digital Elevation Models. *Water Resources Research* 22: 15–24.

Band, L. E., and I. D. Moore. 1995. Scale: Landscape attributes and geographical information systems. *Hydrological Processes* 9:401–22.

Beven, K. J. 1989. Changing ideas in hydrology. The case of physically based models. *Journal of Hydrology* 105:157–72.

———. 1997. TOPMODEL: A critique. *Hydrological Processes* 11:1069–85.

Beven, K. J., and M. J. Kirkby. 1979. A physically based variable contributing area model of basin hydrology. *Hydrology Science Bulletin* 24:43–69.

Bezdek, J. C. 1974. Cluster validity with fuzzy sets. *Journal of Cybernetics* 3:58–73.

———. 1981. *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

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

Brabyn, L. 1997. Classification of macro landforms using GIS. *ITC Journal* 14:26–40.

Burrough, P. A. 1993. Soil variability: A late 20th century view. *Soils and Fertilizers* 56:531–61.

Burrough, P. A., and A. U. Frank, eds. 1996. *Geographic objects with indeterminate boundaries*. London: Taylor & Francis.

Burrough, P. A., and R. A. McDonnell. 1998. *Principles of geographical information systems*, 265–91. Oxford, U.K.: Oxford University Press.

Burrough, P. A., P. F. M. van Gaans, and R. A. MacMillan. 2000. High-resolution landform classification using fuzzy *k*-means. *Fuzzy Sets and Systems* 113:37–52.

Burrough, P. A., J. P. Wilson, P. F. M. van Gaans, and A. J. Hanson. 2001. Fuzzy *k*-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. *Landscape Ecology* 16:523–46.

Chuvieco, E., D. Cocero, D. Riaño, P. Martin, J. Martínez-Vega, J. de la Riva, and F. Pérez. 2004. Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sensing of Environment* 92:322–31.

Cook, S. E., R. J. Corner, G. Grealish, P. E. Gessler, and C. J. Chatres. 1996. A rule-based system to map soil properties. *Soil Science Society of America Journal* 60:1893–900.

Copland, L. 1998. The use of terrain analysis in the evaluation of snow cover over an alpine glacier. In *Landform monitoring, modeling and analysis*, ed. S. Lane, K. Richards, and J. Chandler, 385–404. New York: Wiley.

de Bruin, S., and A. Stein. 1998. Soil-landscape modeling using fuzzy *c*-means clustering of attribute data derived from a Digital Elevation Model (DEM). *Geoderma* 83:17–33.

de Gruijter, J. J., and A. B. McBratney. 1988. A modified fuzzy *k*-means for predictive classification. In *Classification and related methods of data analysis*, ed. H. H. Bock, 97–104. Amsterdam: Elsevier.

Dikau, R. 1989. The application of a digital relief model to landform analysis in geomorphology. In *Three dimensional applications in geographic information systems*, ed. J. Raper, 51–77. London: Taylor & Francis.

Dunn, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3:32–57.

Dymond, J. R., R. C. Derose, and G. R. Harmsworth. 1995. Automated mapping of land components from digital elevation data. *Earth Surface Processes and Landforms* 20: 131–37.

Evans, I. S. 2003. Scale-specific landforms and aspects of the land surface. In *Concepts and modelling in geomorphology: International perspectives*, ed. I. S. Evans, R. Dikau, E.

Tokunaga, H. Ohmori, and M. Hirano, 61–84. Tokyo: Terrapub.

Fisher, P. F. 2000. Sorites paradox and vague geographies. *Fuzzy Sets and Systems* 113:7–18.

Fisher, P. F., and S. Pathirana. 1990. The evaluation of fuzzy membership of land cover classes in the suburban zone. *Remote Sensing of Environment* 34:121–32.

Florinsky, I. V. 1998. Combined analysis of digital terrain models and remotely sensed data in landscape investigations. *Progress in Physical Geography* 22:33–60.

Florinsky, I. V., and G. A. Kuryakova. 2000. Determination of grid size for digital terrain modelling in landscape investigations – exemplified by soil moisture distribution at a micro-scale. *International Journal of Geographical Information Science* 14:795–814.

Fonte, C. C., and W. A. Lodwick. 2004. Areas of fuzzy geographical entities. *International Journal of Geographical Information Science* 18:127–50.

Gessler, P. E., I. D. Moore, N. J. McKenzie, and P. J. Ryan. 1995. Soil-landscape modeling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* 9:421–32.

GIS Data Depot. http://data.geocomm.com/(last accessed 24 February 2005).

Goodchild, M. F., and S. Gopal, eds. 1989. *Accuracy of spatial databases*. London: Taylor & Francis.

Gorsevski, P. V., P. E. Gessler, and P. Jankowski. 2003. Integrating a fuzzy *k*-means classification and a Bayesian approach for spatial prediction of landslide hazard. *Journal of Geographical Systems* 5:223–51.

Graff, L. H., and E. L. Usery. 1993. Automated classification of generic terrain features in digital elevation models. *Photogrammetric Engineering and Remote Sensing* 59:1409–17.

Guptill, S. C., and J. L. Morrison, eds. 1995. *Elements of spatial data quality*. Oxford: Pergamon.

Hammond, E. H. 1964. Analysis of properties in landform geography: An application to broadscale land form mapping. *Annals of the Association of American Geographers* 54: 11–19.

Irvin, B. J., S. J. Ventura, and B. K. Slater. 1997. Fuzzy and isodata classification of landform elements from digital elevation data in Pleasant Valley, Wisconsin. *Geoderma* 77: 137–54.

Karssenberg, D., P. A. Burrough, R. Sluiter, and K. de Jong. 2001. The PCRaster software and course material for teaching numerical modeling in the environmental sciences. *Transactions in GIS* 5:99–110.

Kienzle, S. 2004. The effect of DEM raster resolution on first order, second order and compound terrain derivatives. *Transactions in GIS* 8:83–112.

Lark, R. M. 1999. Soil-landform relationships at within-field scales: An investigation using continuous classification. *Geoderma* 92:141–65.

MacMillan, R. A., R. K. Jones, and D. H. McNabb. 2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). *Computers, Environment and Urban Systems* 28:175–200.

MacMillan, R. A., W. W. Pettapiece, S. C. Nolan, and T. W. Goddard. 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems* 113:81–109.

Mark, D. M., and B. Smith. 2003. Do mountains exist? Ontology of landforms and topography. *Environment & Planning B: Planning and Design* 30:411–27.

———. 2004. A science of topography: Bridging the qualitative-quantitative divide. In *Geographic information science and mountain geomorphology*, ed. M. P. Bishop and J. Shroder, 75–100. Chichester, U.K.: Springer-Praxis.

McBratney, A. B., and J. J. de Gruijter. 1992. A continuum approach to soil classification by modified fuzzy *k*-means with extragrades. *Journal of Soil Science* 43:159–75.

McBratney, A. B., and A. W. Moore. 1985. Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology* 35:165–85.

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

McKenzie, N. J., and M. P. Austin. 1993. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma* 57:329–55.

Moore, I. D., P. E. Gessler, G. A. Nielsen, and G. A. Peterson. 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57:443–52.

Moore, I. D., R. B. Grayson, and A. R. Ladson. 1991. Digital terrain modeling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes* 5: 3–30.

Moore, I. D., A. Lewis, and J. C. Gallant. 1993. Terrain attributes: Estimation methods and scale effects. In *Modelling change in environmental systems*, ed. A. J. Jakeman, M. B. Beck, and M. J. McAleer, 189–214. New York: Wiley.

O'Callaghan, J. F., and D. M. Mark. 1984. The extraction of drainage networks from digital elevation data. *Computer Graphics and Image Processing* 28:323–44.

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough. 1992. Soil pattern recognition with fuzzy *k*-means: Application to classification and soil-landform interrelationships. *Soil Science Society of America Journal* 56:505–16.

Omernik, J. M. 1995. Ecoregions: A spatial framework for environmental management. In *Biological assessment and criteria: Tools for water resource planning and decision making*, ed. W. S. Davis and T. P. Simon, 49–63. Boca Raton, FL: Lewis Publishers.

Park, S. J., K. McSweeney, and B. Lowery. 2001. Identification of the spatial distribution of soils using process-based terrain characterization. *Geoderma* 103:249–72.

PCRaster. Version 2 Manual at http://pcraster.geog.uu.nl/documentation/pcrman/book1.htm (last accessed 18 May 2005).

Pennock, D. J., B. J. Zebarth, and E. de Jong. 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. *Geoderma* 40:297–315.

Pfeffer, K., E. J. Pebesma, and P. A. Burrough. 2003. Mapping alpine vegetation using vegetation observation and topographic attributes. *Landscape Ecology* 18:759–76.

Pike, R. J. 1988. The geometric signature: Quantifying landslide-terrain types from digital elevation models. *Mathematical Geology* 20:491–511.

Quinn, P., K. J. Beven, and R. Lamb. 1995. The $\ln(\alpha/\tan\beta)$ index: How to calculate it and how to use it within the TOPMODEL framework. *Hydrological Processes* 9: 162–82.

Robinson, V. B. 2003. A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS* 7:3–30.

Robinson, V. B., and D. Thongs. 1986. Fuzzy set theory applied to the mixed pixel problem of multispectral landcover databases. In *Geographic information systems in government*, ed. B. Opitz, 871–85. Hampton, VA: A. Deepak Publishing.

Roubens, M. 1982. Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research* 10: 294–301.

Schneider, B. 2001. Phenomenon-based specification of the digital representation of terrain surfaces. *Transactions in GIS* 5:39–52.

Skidmore, A. K., P. J. Ryan, W. Dawes, D. Short, and E. O'Loughlin. 1991. Use of an expert system to map forest soils from a geographical information system. *International Journal of Geographical Information Systems* 5:421–45.

Smith, B., and D. M. Mark. 1998. Ontology and geographic kinds. In *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)*, ed. T. K. Poiker and N. Chrisman, 308–20. Vancouver, B.C.: International Geographical Union.

Stephenson, J. R., and G. M. Calcarone. 1999. *Southern California mountains and foothills assessment: Habitat and species conservation issues*. General Technical Report GTR-PSX-172. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture: 1–61.

Unwin, D. J. 1995. Geographical information systems and problems of error and uncertainty. *Progress in Human Geography* 19:549–58.

Wilson, J. P., and J. C. Gallant. 2000a. Digital terrain analysis. In *Terrain analysis: Principles and applications*, ed. J. P. Wilson and J. C. Gallant, 1–28. New York: Wiley.

———. 2000b. Secondary terrain attributes. In *Terrain analysis: Principles and applications*, ed. J. P. Wilson and J. C. Gallant, 87–132. New York: Wiley.

Wilson, J. P., D. J. Spangrud, G. A. Nielsen, J. S. Jacobsen, and D. A. Tyler. 1998. Global positioning system sampling intensity and pattern effects on computed topographic attributes. *Soil Science Society of America Journal* 62:1410–17.

Wood, E. C., M. Sivapalan, K. Beven, and L. Band. 1988. Effects of spatial variability and scale with implications to hydrologic modelling. *Journal of Hydrology* 102:29–47.

Wood, J. 1996. Scale-based characterization of digital elevation models. In *Innovations in GIS 3*, ed. D. Parker, 163–75. London: Taylor & Francis.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Theory* 8: 338–53.

Zhang, W., and D. R. Montgomery. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulation. *Water Resources Research* 30:1019–28.

Zhu, A. X., L. Band, R. Vertessy, and B. Dutton. 1997. Derivation of soil properties using a Soil Land Inference Model (SoLIM). *Soil Science Society of America Journal* 61:523–33.

Correspondence: Department of Geography, Western Illinois University, Macomb, IL 61455, e-mail: Y-Deng2@wiu.edu (Deng); Department of Geography, University of Southern California, Los Angeles, CA 90089-0255, e-mail: jpwilson@usc.edu (Wilson).