**Earth Surface Processes and Landforms**
*Earth Surf. Process. Landforms* **31**, 1452–1462 (2006)
Published online 3 August 2006 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/esp.1401

# Effects of Variable Attribute Weights on Landform Classification

Y. X. Deng,[1]* J. P. Wilson[2] and J. Sheng[2]
[1] *Department of Geography, Western Illinois University, Macomb, IL, USA*
[2] *Department of Geography, University of Southern California, Los Angeles, CA, USA*

*Correspondence to: Yongxin Deng, Department of Geography, Western Illinois University, Macomb, IL 61455-1390, USA. E-mail: Y-Deng2@wiu.edu*

## Abstract

**This paper focuses on the attribute weight issue and advocates use of modifiable attribute weights in terrain-based environmental analysis and classification. A question was asked: 'How much will the result of a terrain-based environmental analysis be affected if the weights of used terrain attributes are changed?' The literature on landform classification and the fuzzy *k*-means method was reviewed in particular to help clarify the background and importance of this weight assignment issue. As an example, the effects of modifying attribute weights were evaluated for fuzzy *k*-means landform classification in a case study area. A total of 102 classifications were compared with each other and with a soil map, and comparison methods were specifically designed to evaluate the differences between these classifications. The results show that fuzzy *k*-means landform classification is sensitive to weight adjustments of adopted terrain attributes. The sensitivity is particularly high when the attribute weights started to be tuned away from the standard (i.e. uniform) weight of one. Better matching between landform classification and a soil map may be produced when attribute weights are tuned. In all, we advocate the widespread adoption of an exploratory attitude in assigning attribute weights for environmental analysis and classification. Copyright © 2006 John Wiley & Sons, Ltd.**

Received 8 June 2005; Revised 6 February 2006; Accepted 10 April 2006

**Keywords: attribute; weights; sensitivity; landform classification; fuzzy *k*-means**

## Introduction

Scientists strive to construct linkages between topography and other biophysical components, so as to represent environmental patterns and processes based on terrain information (Carrara, 1983; Moore *et al.*, 1991; Franklin, 1995; Quinn *et al.*, 1995; Park *et al.*, 2001; Pfeffer *et al.*, 2003). In so doing, they usually adopt a uniform weight for involved attributes, even though multiple topographic properties may carry different degrees of significance for a particular biophysical process or pattern (Pennock *et al.*, 1987; Pike, 1988; MacMillan *et al.*, 2000; Deng and Wilson, in press). For example, the relief, steepness, orientation and surface curvature of a hillslope may all influence the potential landslide hazard in direct or indirect ways, but these factors do not play the same role (Carrara, 1983; Gorsevski *et al.*, 2003). A subsequent question is 'How much will the result of a terrain-based environmental analysis be affected if the weights of used terrain attributes are changed?'.

We use fuzzy *k*-means landform classification as an example to answer the above question. As will be described later, this classification method is a data-driven and pattern-based approach that summarizes the collective role of topography in differentiating the environment. Based on calculated terrain attributes (Moore *et al.*, 1991; Florinsky, 1998; Wilson and Gallant, 2000a), the classification follows an iterative procedure to construct continuous spatial patterns of fuzzy landform class memberships. The spatial co-variation between these memberships and other biophysical properties, which are often continuous in space as well, can then be identified. This step facilitates the interpretation of resultant landform classes. In practice, however, the limited availability of environmental data often frustrates this last step, and forces it to be replaced with comparisons between boundary-dominant environmental data and 'defuzzified' landform classes (Irvin *et al.*, 1997; Burrough *et al.*, 2001). As shown by the case study reported in

Copyright © 2006 John Wiley & Sons, Ltd.

this paper, spatial aggregations of both environmental properties and landform class memberships eventually influence comparison results.

We advocate the introduction of uneven attribute weights and, subsequently the need of exploratory methods to identify the optimal weight combinations in fuzzy *k*-means landform classification and other terrain-based environmental analyses (e.g. Pike, 1988). This paper aims to provide evidence for this suggestion by demonstrating the sensitivity of landform classification to attribute weights. Hence, the next section outlines the literature for landform classification and the fuzzy *k*-means method to envision the role of attribute weights in environmental analysis. The third section explains two experiments that defined (1) weight-caused change of fuzzy *k*-means landform classification results and (2) variation of the spatial overlaps between classification results and a soil map. The last section briefly summarizes lessons we learned from these two experiments.

## Background and Theory

As a long-standing tradition of physical geography, landform classification aims to delineate topographic classes in space, in the hope they are indicative of biophysical patterns (Fenneman, 1928; Hammond, 1964; Brabyn, 1997; Giles, 1998; Burrough *et al.*, 2001). The attraction of doing this lies in the contrast between the relative ease of representing landform patterns and the extreme complexity of measuring the biophysical environment across space. This contrast is accompanied by the underlying belief that the biophysical environment exists as a whole with 'interlinkages . . . between spheres of environment' at various spatial scales (Gregory, 2000). We hereby interpret the linkages between landforms and the biophysical environment as identifiable relationships between topographic properties and other biophysical properties in attribute space. Ultimately, these linkages are explicitly expressed in geographic space as spatial correlations or co-variations between landform patterns and other biophysical patterns. Knowledge of these linkages is valuable because it allows us to infer biophysical information from landform information in various ways.

When Fenneman (1928) traced the major physiographic boundaries of the U.S. by dividing the country into eight major divisions that contain 25 subdivisions, or physiographic provinces, two rules were established for landform classification. First, 'form' (landform pattern) should be viewed as the 'primary basis of subdivision' (Fenneman, 1928). Second, accuracy of boundary lines and, correspondingly, homogeneity or similarity within boundaries, are essential for the quality of classification results. These rules were established and practiced at a very coarse spatial resolution at that time, and the usefulness of Fenneman's topographic delineation depended on whether the combination of geographic boundaries and the homogeneity within boundaries can effectively differentiate the biophysical environment at a corresponding spatial extent (scale) – the entire United States.

Hammond (1964) introduced three topographic properties (i.e. terrain attributes) – slope inclination, relief or vertical dimension, and general profile character – into his landform classification. He summarized these properties across the United States for square-shaped areal units measuring six miles on each side, and then classified these units into 45 significant classes (Hammond, 1964). Compared to Fenneman's classification, this approach signified a tremendous improvement in spatial resolution, and it is also an early implementation of what has become known to be the 'field-based' conception of geographic space (Couclelis, 1992).

The introduction of landform properties, which can be generalized with simple geometric calculations (Evans, 1980; Zevenbergen and Thorne, 1987; Moore *et al.*, 1991; Wood, 1996; Florinsky, 1998; Wilson and Gallant, 2000a), proved to be crucial to later developments of landform classification. It assisted the abstraction of landform-environment linkages based on correlations between individual properties; it allowed landforms to be quantitatively characterized and classified at a very fine spatial resolution – down to a data point or cell; and it supported the repetition of landform classes over space, given the repetition of properties. However, use of multiple properties in classification does not guarantee these properties to co-vary exactly or exhibit perfectly correlated thresholds (Plewe, 1997; Cheng, 2002). As a result, terrain attributes can be summarized in vastly different ways, which makes landform classification results a function of the selected classification method and parameters. Indeterminate geographic boundaries may also emerge in consequence (Burrough and Frank, 1996; Plewe, 1997).

Computers brought new opportunities for landform classification because of their strong capacity and high speed in processing spatial data. At fine spatial resolutions (e.g. 12·5–50 m grid cells), for example, Dikau (1989) automated a similar (but more sophisticated) classification procedure to Hammond in a GIS environment using a longer list of terrain attributes calculated for both DEM cells and delineated relief objects. These resolutions imply the differentiation and classification of landform units within a small watershed or even within a hillslope (e.g. 100 m to 1000 m in length scale). Two levels (hierarchies) of basic 'relief units' were identified and then classified 'based on the concept of homogeneity of geometrical attributes' as well as on the 'recent functionality and historical genesis' of landforms

Copyright © 2006 John Wiley & Sons, Ltd.

DOI: 10.1002/esp

(Dikau, 1989). Nevertheless, the classification results are inevitably dependent on the method used to summarize the large number of attributes.

Working at an even smaller scale (finer resolution), Irvin *et al.* (1997) classified landforms of a 50-ha area in Wisconsin, USA, using ISODATA (Iterative Self-Organizing Data Analysis Technique), an unsupervised clustering method that incorporates a three-step iterative process (see also Ventura and Irvin, 2000). First, data points and arbitrary means were randomly assigned to a user-defined number of clusters in attribute space. Second, a procedure was repeated to evaluate new cluster means and to reassign data points to the nearest clusters, until the change of clusters reached a threshold. Third, a signature file that contained cluster means and their covariance matrices (calculated from the data points contained in the clusters) was used to calculate the statistical probability (likelihood) that each data point belonged to a specific cluster. The highest probability was used to assign a data point to a cluster. ISODATA can thereby identify natural clusters in a reproducible way no matter how many attributes are adopted.

The above efforts fully developed the 'crisp' landform classification paradigm established by Fenneman (1928), Hammond (1964), and others. Within this paradigm, landforms of large areas can be classified rapidly into small, repetitive, and 'areally organized' (Hammond, 1964) units at various spatial scales. However, the spatial continuum that has long been recognized as an essential characteristic of landforms (Hammond, 1964; Mark and Smith, 2003) is poorly represented in this paradigm.

The theory of fuzzy sets (Zadeh, 1965; Robinson, 2003) provided a possible solution to the 'crispness' problem in traditional landform classifications. Fuzzy sets allow individual data points to have partial belongings (memberships) to multiple classes, and therefore allow for the existence of overlapped classes that transition 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. Within earth sciences, this theory has been applied to classify climate data (McBratney and Moore, 1985), geologic data (Bezdek *et al.*, 1984), remote sensing images (Robinson and Thongs, 1986; Fisher and Pathirana, 1990), soil data (McBratney and de Gruijter, 1992; McBratney *et al.*, 1992; Odeh *et al.*, 1992; Zhu *et al.*, 1997) and terrain data (Irvin *et al.*, 1997; Burrough *et al.*, 2000; Gorsevski *et al.*, 2003).

Burrough and McDonnell (1998, pp. 265–291) contrasted two ways of summarizing fuzziness in spatial distributions: the SI (semantic import) approach imports pre-specified class centers and membership functions for membership assignments; whereas the fuzzy $k$-means (clustering) approach identifies natural class centers existing in data and computes class memberships according to attribute distances. The SI approach is more suitable when the central concepts of landforms – classes or objects (e.g. mountain peaks) – are well-known or easily definable. In these cases, we have a very good idea of how to classify data or how to extract terrain objects based on representative landform geometries and properties (Herrington and Pellegrini, 2000; Mark and Smith, 2004). However, the complications involved in organizing or classifying topographic data in multivariate space tend to be overwhelming for the SI approach. For example, the fact that topographic and other biophysical properties do not co-vary exactly (Plewe, 1997; Cheng, 2002) may force the user to choose from many competing semantic concepts or predefined membership functions (Burrough and McDonnell, 1998, pp. 270–273) to classify the same group of terrain attributes.

The fuzzy $k$-means algorithms provide a data-driven, reproducible approach to the detection of cluster centers (similar to the ISODATA procedure) and to the generation of fuzzy membership functions. These algorithms were originally developed by Dunn (1973), and then generalized by Bezdek (1974, 1981). McBratney and de Gruijter (1992) later developed a modified membership function that considers the extragrades – data points dissimilar to any cluster centers – during classification. In a multivariate attribute space, this method first allocates data points randomly into $k$ clusters. The center of each cluster is calculated as the average of point attributes in the cluster. Next, $k$ similarity indices are calculated for each data point to represent its similarity to the $k$ cluster centers. This calculation is accomplished with the assistance of a selected attribute distance function. According to these similarity indices, each data point is reallocated to the most similar cluster. New cluster centers and new similarity indices can subsequently be calculated. This process goes on until a stable solution (or a threshold) is reached and $k$ stable cluster centers are established (Burrough and McDonnell, 1998).

A prominent issue in fuzzy $k$-means classification, therefore, is to evaluate the similarity of a data point to each of the cluster centers, instead of determining whether a data point belongs to a cluster (as in the ISODATA case). The membership $\mu_{ik}$ of each data point $i$ to class center $k$ is calculated according to the attribute distance between $i$ and $k$. The most commonly used fuzzy $k$-means membership ($\mu_{ik}$) function is written as:

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

Copyright © 2006 John Wiley & Sons, Ltd.


DOI: 10.1002/esp

where $d_{ik}$ is the attribute distance between $i$ and $k$, $c$ is the number of fuzzy classes, and $q(q > 1)$ is the exponent of fuzziness (Burrough and McDonnell, 1998), indicating the extent of fuzziness – the larger $q$ is, the fuzzier the classification is. As a result, $\mu_{ik}$ necessarily falls between 0 ($i$ has no similarity to $k$) and 1 ($i$ is exactly the same as $k$).

Terrain attribute weights affect fuzzy $k$-means landform classification in multiple ways. First, they determine conceptual significances of each topographic property in the classification. Second, they influence attribute distance calculations, a critical step in fuzzy $k$-means classification (e.g. Equation (1)). For example, the following equation shows the diagonal norm distance function that is often used to calculate the attribute distance $d_{ik}$ in Equation (1):

$$d_{ik} = \sqrt{\sum_{j=1}^{n} \frac{1}{\sigma_j^2} \times (a_{ij} - a_{kj})^2} \tag{2}$$

where $\sigma_j$ is the standard deviation of the attribute $a_j$, $a_{ij}$ and $a_{kj}$ are values of $a_j$ for the $i$th point and the $k$th class center respectively and $n$ is the number of attributes used. The same weight is applied to considered attributes in this equation, implying all attributes were treated in the same way. Equation (2) can thus be re-written by adding in a weight term $w_j$ for attribute $a_j$ as:

$$d_{ik} = \sqrt{\sum_{j=1}^{n} \frac{w_j^2}{\sigma_j^2} \times (a_{ij} - a_{kj})^2} \tag{3}$$

Equation (3) allows use of differentiable weights for various attributes, or the existence of variable significances of topographic properties. It may consequently produce variable results for fuzzy $k$-means landform classification. Hence the third effect of differentiable attribute weights is on calculated class centers and membership distributions.

There is a key challenge for fuzzy $k$-means landform classification: it results in a stable solution that can optimally partition data, but the biophysical (real-world) meanings of produced classes – class centers and memberships – need further interpretation. There is no guarantee in this process that the uniform weight assignment will correspond to the most meaningful classification, in part because the optimal weight assignment is generally unknown. This is to say, attribute weights may need to be tuned and evaluated in an iterative way to produce interpretable landform classes that best fit a particular application or best match a specific biophysical pattern. Hence another effect of terrain attribute weights is on meanings that can be interpreted for landform classes.

The crisp options of either including or not including a terrain attribute are equivalent to assigning crisp weights of either *1 (included)* or *0 (not included)*, instead of a value ranging between *0* and *1*, to all adopted attributes. This situation is not limited in the literature of landform classification, but is prevalent in most biophysical analyses that incorporate multiple terrain attributes. The underlying assumptions are that: (1) all included attributes are equally important to the analysis, (2) other attributes are equally unimportant, and (3) the effects of included attributes do not overlap. The above assumptions persist because we cannot confidently define the significance of one-to-one correlations – or closeness of linkages – between topographic properties and other biophysical properties. Although it is not likely that this problem can be solved in a foreseeable future, a rigid attitude towards attribute weight assignment should still be avoided. In this context, the next section of this paper presents a sensitivity test illustrating the variations in fuzzy $k$-means landform classification results that can be caused by variable terrain attribute weights.

# Case Study

## Experimental design

*Data sources.* We conducted a case study in a 157 km$^2$ area located in the southeast part of Ventura County, California (Figure 1(a)). Undisturbed mountainous landscape (Santa Monica Mountains) covers the majority of this study area. Steep coastal slopes, mountaintops, and deep valleys formed by short, incisive stream channels are common landform types. The elevations range from the sea level to 948 m, and the average slope gradient for the entire area is 40 per cent (calculated from 10 m DEMs).

Four USGS 10 m DEMs were used as the source data for landform classifications. The Digital Soil Survey Geographic Database (SSURGO) of the study area was summarized to produce a five-class soil map that can be compared with landform classifications. SSURGO presented a total of 210 polygons belonging to 61 soil map units for the study area. We chose 171 polygons of 46 soil map units for our analysis, and the remainder was excluded for two reasons. First, they constitute

Copyright © 2006 John Wiley & Sons, Ltd.

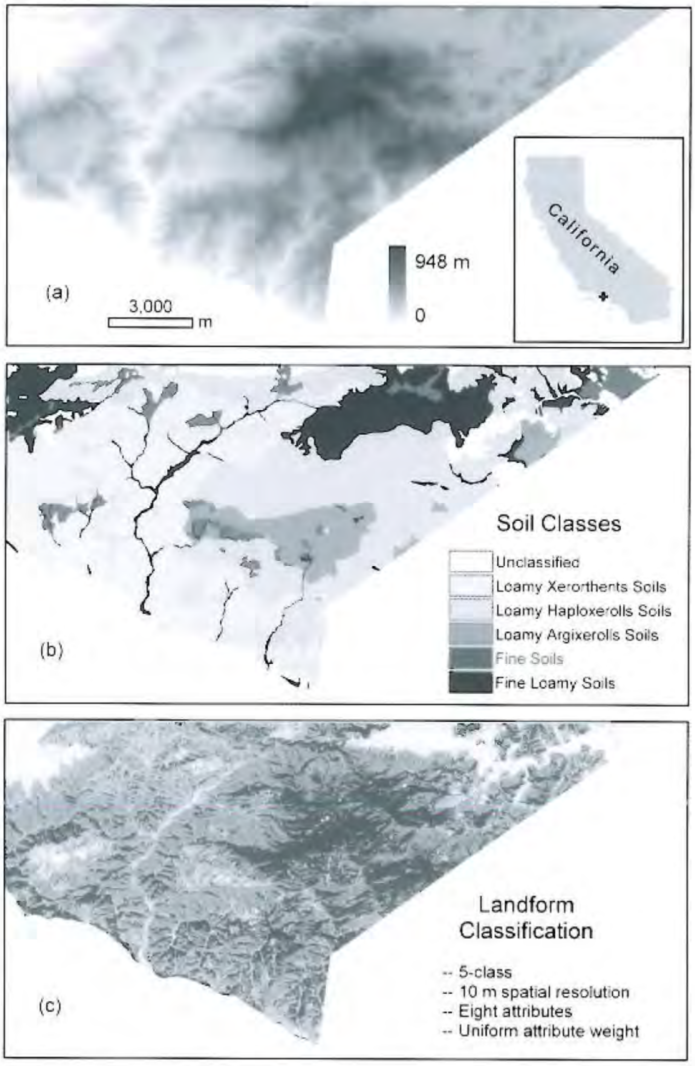


**Figure 1.** The case study area DEM (a), soil map (b) and map of landform classification with uniform weight (c).

water bodies, bare rocks, urban areas, or landscaped areas (10 soil map units, 33 polygons); and second, they exist in this area as 'extragrades' (five soil map units, six polygons), thus being difficult to aggregate into other soil classes.

The selected soil map units were first summarized into nine categories according to their great group names and soil texture classes (Soil Survey Staff, 1999). The nine categories were further grouped into five classes based on soil texture classes and/or soil profile development (Figure 1(b)). These classes include: (1) loamy Haploxerrolls characterized by loamy soil textures and poor soil horizon development; (2) loamy Argixerolls characterized by loamy soil texture and high sub-surface clay contents; (3) loamy Xerothents characterized by loamy soil textures and soil profiles with relatively low moisture content; (4) fine loamy soils (including fine loamy Haploxerolls, Argixerolls, Xerorthents and Xerofluvents) characterized by fine loamy soil textures; and (5) fine soils (including fine Haploxererts and Argixerolls) characterized by fine soil textures.

*Terrain attributes.* Eight terrain attributes – elevation, slope, plan curvature, profile curvature, proximity to nearest ridgeline, incoming solar radiation, topographic wetness index, and sediment transport capacity index (Gallant and Wilson, 2000; Wilson and Gallant, 2000b) – were extracted/calculated from the DEM. Elevation was directly read from the DEM. Slope was calculated using the third-order finite difference method of Horn (1981). Plan and profile curvatures were calculated according to the algorithms by Zevenbergen and Thorne (1987). Upslope contributing area and distance to nearest ridgeline were calculated with the assistance of the D8 flow-routing algorithm (O'Callaghan and Mark 1984). Topographic wetness and sediment transport capacity indexes (Moore and Wilson, 1992; Quinn

Copyright © 2006 John Wiley & Sons, Ltd.

DOI: 10.1002/esp

*et al.*, 1995) were calculated from slope and upslope contributing area using the simplified methods described in Wilson and Gallant (2000b). Natural logarithm of sediment transport capacity index was used. Incoming solar radiation was calculated following two steps. Firstly, information of relief and topographic shading effects was extracted from spatial patterns of elevation, slope, aspect, and sun height (corresponding to latitude). Secondly, daily incoming solar radiation was calculated and accumulated for each grid point based on information obtained from the first step.

*Relative sensitivity test.* We gave all the other attributes the same weight (of unity) throughout the experiments and adjusted the weights of slope and topographic wetness index between 0·25 and 5 (following the scheme of 0·25, 0·5, 1, 1·5, 2, 3, 4 and 5) in one of five ways:

(1) give a weight of 1 to topographic wetness index and change the weight of slope;
(2) give a weight of 5 to slope and change the weight of topographic wetness index;
(3) give a weight of 1 to slope and change the weight of topographic wetness index;
(4) give a weight of 5 to topographic wetness index and change the weight of slope;
(5) simultaneously change weights of slope and topographic wetness index from 0·25 to 5.

Combining (1) and (2), or (3) and (4), in the above list provides continuous weight adjustments of the two attributes, one at a time. Using three-, four- and five-class classification schemes and the above weight assignments, we produced and compared a total of 102 fuzzy $k$-means landform classifications for the sensitivity test. The algorithms presented in the background and theory section were used for fuzzy $k$-means landform classification. The weighted diagonal norm distance function (Equation (3)) was used to identify class centers and assign class memberships.

*Comparisons of landform and soil classifications.* All five-class fuzzy $k$-means landform classifications were compared with the five-class soil map described above to investigate how their spatial overlaps vary with the adjustment of slope and topographic wetness index weights. These two terrain attributes are often cited as key factors for the explanation of soil variability (see, e.g., Moore *et al.*, 1993; Park *et al.*, 2001).

*Comparison methods.*

Comparison method for relative sensitivity test. In this section, we redefine the meaning of $a_{ij}$ in Equation (2) of §2 as the value of attribute $a_j$ for the $i$th class center, given that class centers $i$ and $k$ belong to two classifications. Equation (2) was then used to compute (and then compare) attribute distances between pairs of class centers of two classifications, so as to identify similar class pairs (with one-to-one correspondence) as those that have the smallest attribute distances (Table I).

A surface of the relative membership difference for each similar class pair of two classifications ($i$ and $k$) was obtained by performing the following calculation for each grid point:

$$\mu_{ik} = \frac{|\mu_i - \mu_k|}{\mu_i + \mu_k} \times 100\% \tag{4}$$

where $\mu_i$ and $\mu_k$ are memberships of the same grid point to two similar classes in two classifications, and $\mu_{ik}$ is the relative difference between $\mu_i$ and $\mu_k$. The value of $\mu_{ik}$ ranges from 0 to 100 per cet. The mean of $\mu_{ik}$ for all grid points was then used to represent the difference between a pair of similar classes. Lastly, the calculated differences between each pair of similar classes were averaged to represent the overall difference between two classifications. As a result, both the most similar classes and the most similar classifications can be explicitly identified (Deng and Wilson, in press).

Method for comparisons with soil map. We used five-class fuzzy $k$-means landform classifications for this test, and assigned each grid point to the landform class of the largest membership. This step 'defuzzified' the classification results (Burrough *et al.*, 2000) so that they could be compared with the crisp (i.e. boundary-based) soil map. Next, we conducted a class-to-class overlay between maps of soil and landform classifications, and identified the two landform

**Table I.** Pair-wise attribute distances between class centers of two 4-class classifications $i$ and $k$. Bold numbers indicate pairs of most similar classes

| | | Classification $i$ | | | |
|---|---|---|---|---|---|
| | | **Class 1** | **Class 2** | **Class 3** | **Class 4** |
| Classification $k$ | Class 1 | **0·005** | 4·599 | 4·583 | 4·629 |
| | Class 2 | 3·123 | **0·172** | 6·479 | 2·904 |
| | Class 3 | 3·854 | 2·599 | 1·578 | **0.320** |
| | Class 4 | 3·615 | 6·807 | **0·248** | 1·305 |

Copyright © 2006 John Wiley & Sons, Ltd.

DOI: 10.1002/esp

classes that have the maximum and second-largest overlaps with the soil class. This produced two percentage values for the soil class that correspond to the maximum overlap and the sum of top two overlaps respectively. These percentage values were then averaged over the five soil classes using the area of each soil class as the weight, and the result was used to describe the strength of the spatial correlations between soil and landform classifications.

## Results

*Sensitivity test.* Figure 2 shows overall membership differences between the landform classification produced with the uniform attribute weight and classifications with variable attribute weights. A maximum membership variation of

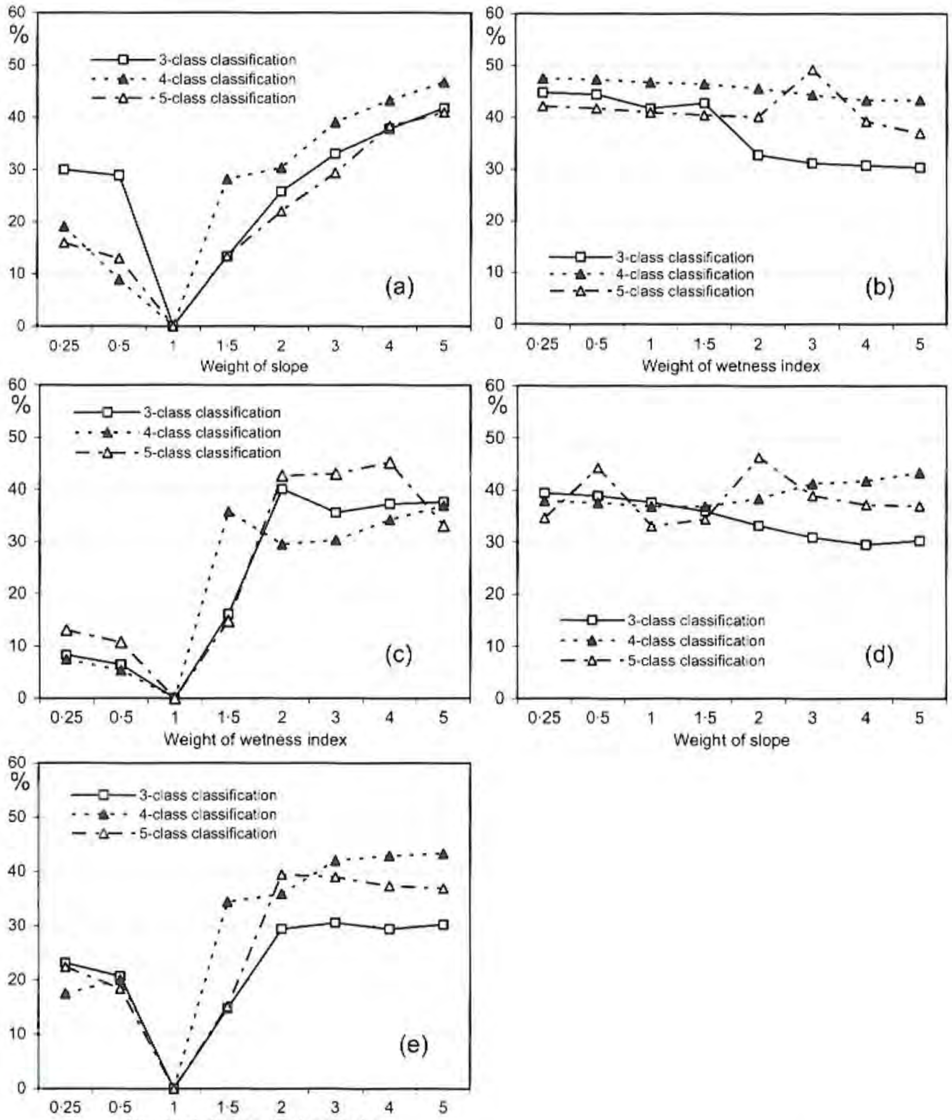


**Figure 2.** Results of relative sensitivity test – membership differences (in %, see §3.1.5.1) between the classification with uniform weight assignment and classifications where: (a) topographic wetness index has the same weight (equal to 1) as other attributes but the weights of slope are adjusted; (b) slope has a high weight (equal to 5) and the weights of topographic wetness index are adjusted; (c) slope has the same weight (equal to 1) as other attributes but the weights of topographic wetness index are adjusted; (d) topographic wetness index has a high weight (equal to 5) and the weights of slope are adjusted; (e) weights of both slope and topographic wetness index are adjusted.

49 per cent (Figure 2b), corresponding to a weight of 5 for slope and 3 for wetness index, was observed. This indicates that fuzzy $k$-means landform classification is sensitive to the weights used for the terrain attributes. The biggest variation rate occurred when the attribute weights started to drift away from the uniform weight. There seems to be a threshold weight of 1·5 or 2 for topographic wetness index, but less so for weights of slope. In most cases, increasing weights of the two attributes caused more severe impacts than decreasing them. Membership differences changed relatively little and did not show obvious trends when one of the two attributes retained a high weight, which again suggested the possible existence of threshold weight(s).

*Soil map comparisons.* A visual comparison of the soil map (Figure 1(b)) and one of the landform classification maps (Figure 1(c)) indicates that there are low levels of correspondence between these two sets of classifications. This result corresponds with the observation of Band and Moore (1995, p. 402): 'Typically, soil information is known with

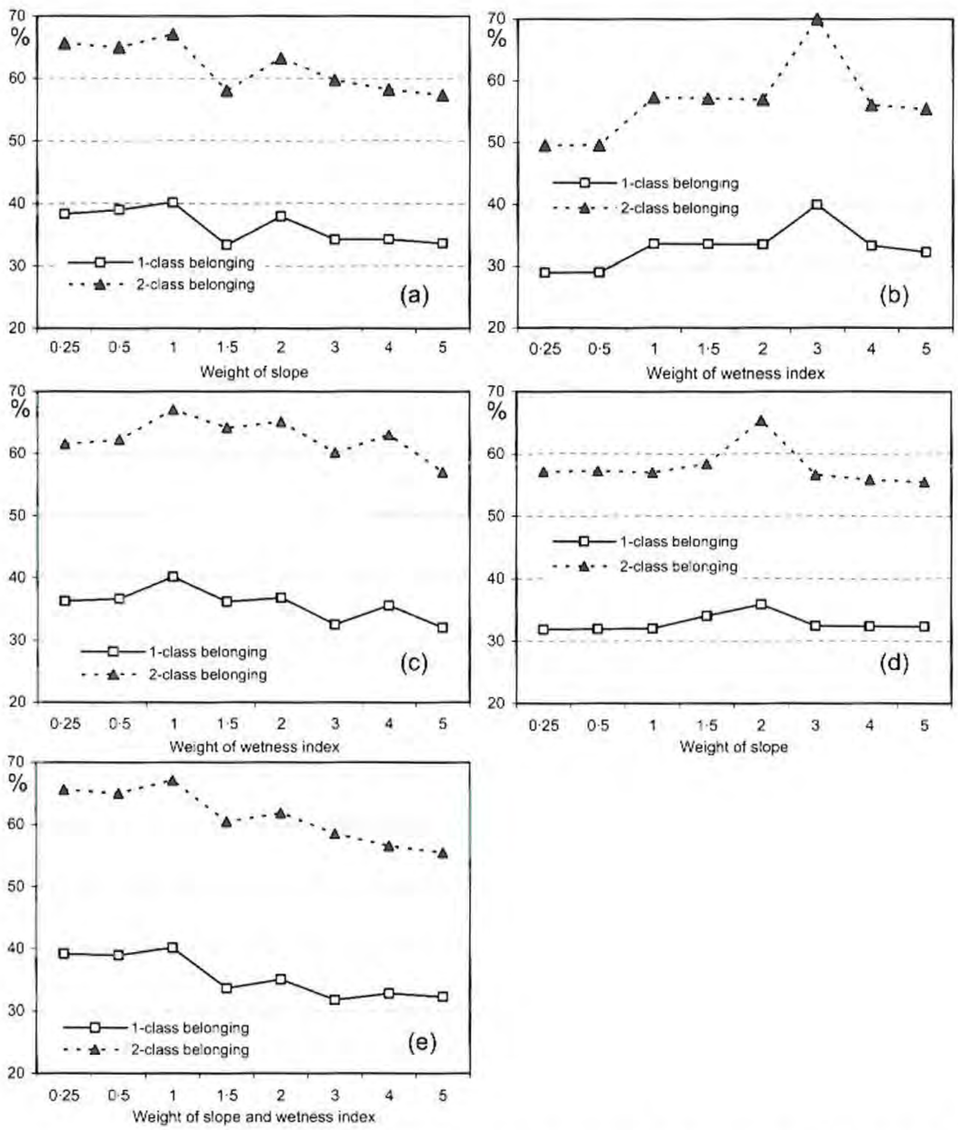


**Figure 3.** Results comparing a soil map and landform classifications – in particular, the maximum area of each soil class falling into one or two landform classes (in %, see §3.1.5.2) when landform classes are derived with the following weight adjustments: (a) topographic wetness index has the same weight (equal to 1) as other attributes but the weights of slope are adjusted; (b) slope has a high weight (equal to 5) and the weights of topographic wetness index are adjusted; (c) slope has the same weight (equal to 1) as other attributes but the weights of topographic wetness index are adjusted; (d) topographic wetness index has a high weight (equal to 5) and the weights of slope are adjusted; (e) weights of both slope and topographic wetness index are adjusted.

Copyright © 2006 John Wiley & Sons, Ltd.

the least certainty and at the greatest level of generalization due to the difficulty of sampling, high-field variability and the general inability to remotely image the required soil properties. Simple GIS overlay of digital terrain data or remotely sensed vegetation data with digital soil maps often results in poor estimates of the co-occurrence of these variables.' However, it is still reasonable to assume the existence of inherent linkages, as well as certain spatial correlations, between landform classification and soil maps, because soil map production relied heavily on the traditional soil–landscape model (Soil Survey Staff, 1951, 1999; Hudson, 1990, 1992).

Figure 3 shows the results of comparing the soil map (Figure 1(b)) and landform classification maps (Figure 1(c)). The maximum overlap between the two classifications, as defined in 3.1.5.2, was generally low (30–40 per cent) and varied little with change of attribute weights (mostly <10 per cent). There seemed to be weak but consistent decrease of overlaps between the two sets of maps when the weights of slope and topographic wetness index were tuned away from the uniform weight (Figures 3a, c, and e), and increasing weights of slope seemed to have produced heavier influences than decreasing weights of slope. Obvious peak overlaps appeared when weights of the two attributes were high: the highest at (5, 3) and the second highest at (2, 5) for slope and wetness index respectively.

## Conclusions

This paper sought to reject the traditional Boolean logic, or an 'in or out' strategy, in selecting terrain attributes for biophysical analysis. We proposed use of modifiable attribute weights, and suggest that tuning attribute weights may produce classifications that better fit for particular applications. The reported experiments indicate that results of fuzzy $k$-means landform classification are sensitive to weight adjustments of adopted terrain attributes. The sensitivity is particularly high when the attribute weights started to be tuned away from the uniform one, which is the most commonly used weight assignment so far. Better matching between landform classification and a soil map may be produced when attribute weights are tuned. In correspondence, we advocate the widespread adoption of an exploratory attitude in assigning attribute weights for environmental analysis or classification. Further experiments covering different landscape and/or application domains are needed to demonstrate the generality of these effects and recommendations that followed.

### Acknowledgements

Thanks are owed to Professor Peter Burrough and his colleagues Professor Simon Vriend and Dr. Pauline van Gaans at Utrecht University, who provided the FNX fuzzy $k$-means classification program.

## References

Band LE, Moore ID. 1995. Scale: landscape attributes and geographical information systems. *Hydrological Processes* **9**: 401–422.

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

Bezdek JC. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum: New York.

Bezdek JC, Ehrlich R, Full W. 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 PA, Frank AU (eds). 1996. *Geographic Objects with Indeterminate Boundaries*. Taylor and Francis: London.

Burrough PA, McDonnell RA. 1998. *Principles of Geographical Information Systems*. Oxford University Press: Oxford.

Burrough PA, van Gaans PFM, MacMillan RA. 2000. High-resolution landform classification using fuzzy $k$-means. *Fuzzy Sets and Systems* **113**: 37–52.

Burrough PA, Wilson JP, van Gaans PFM, Hensen AJ. 2001. Fuzzy $k$-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. *Landscape Ecology* **16**: 523–546.

Carrara A. 1983 Multivariate models for landslide hazard evaluation. *Mathematical Geology* **15**: 403–426.

Cheng T. 2002. Fuzzy objects: their changes and uncertainties. *Photogrammetric Engineering and Remote Sensing* **68**: 41–49.

Couclelis H. 1992. People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. In *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science 639*, Frank AU, Campari I, Formentini U (eds). Springer Verlag: Berlin; 65–77.

Deng YX, Wilson JP. The Role of Attribute Selection in GIS Representations of the Biophysical Environment. *Annals of Association of American Geographers*: In press.

Dikau R. 1989. The application of a digital relief model to landform analysis in geomorphology. In *Three Dimensional Applications in Geographic Information Systems*, Raper J (ed.). Taylor and Francis: London; 51–77.

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

Copyright © 2006 John Wiley & Sons, Ltd.

Evans IS. 1980. An integrated system of terrain analysis and slope mapping. *Zeitschrift für Geomorphologie, N.F., Supplementband* **36**: 274–295.

Fenneman NM. 1928. Physiographic divisions of the United States of America. *Annals of the Association of American Geographers* **4**: 261–353.

Fisher PE, Pathirana S. 1990. The evaluation of fuzzy membership of land cover classes in the suburban zone. *Remote Sensing of Environment* **34**: 121–132.

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

Franklin J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* **19**: 474–499.

Gallant JC, Wilson JP. 2000. Primary topographic attributes. In *Terrain Analysis: Principles and Applications*, Wilson JP, Gallant JC (eds). John Wiley and Sons: New York; 51–86.

Giles PT. 1998. Geomorphological signatures: classification of aggregated slope unit objects from digital elevation and remote sensing data. *Earth Surface Processes and Landforms* **23**: 581–594.

Gorsevski PV, Gessler PE, Jankowski P. 2003. Integrating a fuzzy *k*-means classification and a Bayesian approach for spatial prediction of landslide hazard. *Journal of Geographical Systems* **5**: 223–251.

Gregory KJ. 2000. *The Changing Nature of Physical Geography*. The Oxford University Press: New York.

Hammond EH. 1964. Analysis of properties in landform geography: an application to broad-scale land form mapping. *Annals of the Association of American Geographers* **54**: 11–19.

Herrington L, Pellegrini G. 2000. An advanced shape-of-country classifier: extraction of surface features from DEMs. In *Proceedings of the 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4)*: Banff, Alberta, Canada, unpaginated CD-ROM.

Horn BKP. 1981. Hill shading and the reflectance map. *Proceedings of IEEE* **69**: 14–47.

Hudson BD. 1990. Concepts of soil mapping and interpretation. *Soil Survey Horizons* **31**: 63–73.

Hudson BD. 1992. Soil genesis, morphology and classification: the soil survey as paradigm-based science. *Soil Science Society of America Journal* **56**: 836–841.

Irvin BJ, Ventura SJ, Slater BK. 1997. Fuzzy and isodata classification of landform elements from digital elevation data in Pleasant Valley, Wisconsin. *Geoderma* **77**: 137–154.

MacMillan RA, Pettapiece WW, Nolan SC, Goddard TW. 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 DM, Smith B. 2003. Do Mountains Exist? Ontology of Landforms and Topography. *Environment and Planning B: Planning and Design* **30**: 411–427.

Mark DM, Smith B. 2004. A science of topography: bridging the qualitative-quantitative divide. In *Geographic Information Science and Mountain Geomorphology*, Bishop MP, Shroder J (eds). Springer-Praxis: Chichester: 75–100.

McBratney AB, de Gruijter JJ. 1992. A continuum approach to soil classification by modified fuzzy *k*-means with extragrades. *Journal of Soil Science* **43**: 159–175.

McBratney AB, de Gruijter JJ, Brus DJ. 1992. Spatial prediction and mapping of continuous soil classes. *Geoderma* **54**: 39–64.

McBratney AB, Moore AW. 1985. Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology* **35**: 165–185.

Moore ID, Gesseler PE, Nielsen GA, Peterson GA. 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* **57**: 443–452.

Moore ID, Grayson RB, Ladson AR. 1991. Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes* **5**: 3–30.

Moore ID, Wilson JP. 1992. Length-slope factors for the Revised Universal Soil Loss Equation: simplified method of estimation. *Journal of Soil and Water Conservation* **49**: 174–180.

O'Callaghan JF, Mark DM. 1984. The extraction of drainage networks from digital elevation data. *Computer Graphics and Image Processing* **28**: 323–344.

Odeh IOA, McBratney AB, Chittleborough DJ. 1992. Soil pattern recognition with fuzzy-*c*-means: application to classification and soil–landform interrelationships. *Soil Science Society of America Journal* **56**: 505–516.

Park SJ, McSweeney K, Lowery B. 2001. Identification of the spatial distribution of soils using process-based terrain characterization. *Geoderma* **103**: 249–272.

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

Pfeffer K, Pebesma EJ, Burrough PA. 2003. Mapping alpine vegetation using vegetation observation and topographic attributes. *Landscape Ecology* **18**: 759–776.

Pike RJ. 1988. The geometric signature: quantifying landslide-terrain types from digital elevation models. *Mathematical Geology* **20**: 491–511.

Plewe B. 1997. A representation-oriented taxonomy of gradation. In *Spatial Information Theory: a Theoretical Basis for GIS, Lecture Notes in Computer Sciences*, Hirtle SC, Frank AU (eds). Springer: Berlin; 121–136.

Quinn P, Beven K, Lamb R. 1995. The $\ln(\alpha/\tan\beta)$ index: how to calculate it and how to use it within the TOPMODEL framework. *Hydrological Processes* **9**: 162–182.

Copyright © 2006 John Wiley & Sons, Ltd.

Robinson VB, Thongs D. 1986. Fuzzy set theory applied to the mixed pixel problem of multispectral landcover databases. In *Geographic Information Systems in Government*, Opitz B (ed.). Deerpak: Hampton: 871–885.

Robinson VB. 2003. A perspective on the fundamentals of fuzzy sets and their use in Geography Information Systems. *Transactions in GIS* **7**: 3–31.

Soil Survey Staff. 1951. *Soil Survey Manual*. US Department of Agriculture: Washington, DC.

Soil Survey Staff. 1999. *Soil Taxonomy: a Basic System of Soil Classification for Making and Interpreting Soil Surveys*. United States Department of Agriculture Natural Resources Conservation Service: Washington, DC.

Ventura SJ, Irvin BJ. 2000. Automated landform classification methods for soil–landscape studies. In *Terrain Analysis: Principles and Applications*, Wilson JP, Gallant JC (eds). Wiley: New York; 267–294.

Wilson JP, Gallant JC. 2000a. Digital terrain analysis. In *Terrain Analysis: Principles and Applications*, Wilson JP, Gallant JC (eds). Wiley: New York; 1–28.

Wilson JP, Gallant JC. 2000b. Secondary terrain attributes. *Terrain Analysis: Principles and Applications*, Wilson JP, Gallant JC (eds). Wiley: New York; 87–132.

Wood J. 1996. Scale-based characterization of digital elevation models. In *Innovations in GIS 3*, Parker D (ed.). Taylor and Francis: London; 163–175.

Zadah LA. 1965. Fuzzy sets. *Information and Theory* **8**: 338–353.

Zevenbergen LW, Thorne CR. 1987. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* **12**: 47–56.

Zhu AX, Band L, Vertessy R, Dutton B. 1997. Derivation of soil properties using a Soil Land Inference Model (SoLIM). *Soil Science Society of America Journal* **61**: 523–533.


Copyright © 2006 John Wiley & Sons, Ltd.