



Research Article

Fuzzy *k*-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA

Peter A. Burrough^{1*}, John P. Wilson², Pauline F.M. van Gaans¹ & Andrew J. Hansen³

¹Utrecht Centre for Environment and Landscape Dynamics, Faculty of Geographical Sciences, University of Utrecht, P.B. 80115, 3508 TC Utrecht, The Netherlands; ²Department of Geography, University of Southern California, Los Angeles, CA 90089-0255, USA; ³Department of Biology, Montana State University, Bozeman, MT 59717, USA (*author for correspondence, e-mail: p.burrough@geog.uu.nl)

Received 25 January 2000; Revised 15 December 2000; Accepted 30 January 2001

Key words: digital elevation models, Fuzzy *k*-means, GIS, landscape classification, topo-climatic analysis, vegetation mapping

Abstract

Previous attempts to quantify topographic controls on vegetation have often been frustrated by issues concerning the number of variables of interest and the tendency of classification methods to create discrete classes though species have overlapping property sets (niches). Methods of fuzzy *k*-means classification have been used to address class overlap in ecological and geographical data but previously their usefulness has been limited when data sets are large or include artefacts that may occur through the derivation of topo-climatic attributes from gridded digital elevation models. This paper presents ways to overcome these limitations using GIS, spatial sampling methods, fuzzy *k*-means classification, and statistical modelling of the derived stream topology. Using data from a ca. 3600 km² forested site in the Greater Yellowstone Area, we demonstrate the creation of meaningful, spatially coherent topo-climatic classes through a fuzzy *k*-means classification of topo-climatic data derived from 100 m gridded digital elevation models (DEMs); these classes were successfully extrapolated to adjacent areas covering a total of ca. 10 000 km². Independently derived land cover data and middle infrared corrected Landsat TM derived estimates of Normalised Difference Vegetation Index (M_NDVI) at 575 independently sampled sites were used to evaluate the topo-climatic classes and test their extrapolation to the larger area. Relations between topo-climatic classes and land cover were tested by χ^2 analysis which demonstrated strong associations between topo-climatic class and 11 of the 15 cover classes. Relations between M_NDVI and topo-climatic classes proved to be stronger than relations between M_NDVI and the independent cover classes. We conclude that the fuzzy *k*-means procedure yields sensible and stable topo-climatic classes that can be used for the rapid mapping of large areas. The value of these methods for quantifying topographic controls on biodiversity and the strength of their relations with computed NDVI values warrant further investigation.

Introduction

The conservation of biodiversity remains one of our greatest challenges as we search for models of sustainable human settlement and economic development. Hansen and Rotella (1998) recently argued that in order to improve landscape management we need to quantify the links between key abiotic factors, ecological processes, vegetation and land use. Similarly,

Mackey (1996) has argued that we need to be able to investigate how projected changes in land use and environmental conditions may affect biodiversity. These aims require understanding based on sufficient appropriate data and the tools to analyse them properly. Until recently our ability to reach sufficient understanding has been limited by the multi-scaled and multivariate nature of key biophysical processes, large numbers of species, the relative paucity of biological

data and the need to extrapolate from point samples to larger areas. The advent of Geographic Information Systems (GIS), Remote Sensing (RS), and related digital data sets provides new opportunities to circumvent these obstacles (e.g., Gerrard et al. 1997).

Mackey (1996) applied elements of hierarchy theory (Koestler, 1967) to ecological phenomena as a means of coping with scale effects. Figure 1 shows five possible scales at which environmental regimes can be modelled and gives examples of the processes and factors operating at each scale for a forested site in a humid, erosional landscape. Primary environmental regimes (PERs) determine the distribution of light, heat, water and mineral nutrients for photosynthesising plants, and the processes at higher levels set constraints for those operating at lower levels. There is some feedback up this hierarchy and biologically mediated processes will usually dominate in defining the effective environment for a given organism. Most of the ecological research this century has been conducted at the global-, micro-, and nano-scales, although an increasing number of recent studies have examined the controls on vegetation patterns and biodiversity at meso- and topo-scales.

The work at these two intermediate scales has capitalized on the increasing availability of high-resolution, continuous, gridded Digital Elevation Models (DEMs) and the development of new computerized terrain analysis tools as exemplified by the PCRaster (Wesseling et al., 1996) and TAPES tool sets (Wilson and Gallant, 1998). DEMs with grid sizes of 100 m, 30 m, 10 m, 5 m, and even 1 m resolution can be obtained from which many different topographic attributes can be derived for every grid cell as a function of its surroundings (Burrough and McDonnell 1998; Burrough et al., 2000). Topographical attributes such as drainage networks and ridgelines that affect the distribution of water, and slope, aspect, and horizon shading that modify the amount of solar radiation received at the surface, have important ecological consequences in many landscapes (Moore et al. 1991, 1993). These topo-climatic attributes can easily be derived as gridded but continuous surfaces from DEMs using GIS: the challenge is to determine if this information usefully quantifies topographic controls on vegetation patterns.

Mapping vegetation requires two components, first a meaningful classification and second a means of extrapolating the classes to areas. For practical purposes such as the monitoring and management of natural forest stands over large areas it is necessary to re-

duce the multivariate abiotic and biotic attributes to a meaningful, but manageable number of classes or relationships that have a clear spatial expression. The range of methods for relating abiotic factors to vegetation response at sample sites includes aerial photo interpretation, the use of logical relationships and 'expert knowledge' (e.g., the Ellenberg numbers used in northwest Europe to indicate species-site affinities – Ellenberg et al. 1992), gradient and site analysis (cf., Austin et al. 1984, Tilman 1994), multivariate statistical methods such as principal component analysis and correspondence analysis (e.g., Jongman et al. 1995), logistic regression (Hosmer and Leemshow 1989), and more recently neural networks (Fitzgerald and Lees 1996). A major drawback of many of these methods is that of necessity they may be limited to a relatively small number of data points or small areas. Another problem in the case of 'expert knowledge' and methods such as aerial photograph interpretation is that the methods may not be truly reproducible – see Albrecht (1992) for a discussion on the problems of ecological modelling on the basis of phenomenological investigations.

Extrapolating classes to areas is usually done in one of two ways – either the classes are interpolated from field observations (e.g., Rossi et al. 1992), or the classes make use of data that vary continuously over space. In many cases, the current operational method for mapping vegetation is to use remotely sensed infra red data, and from these to derive vegetation indices to indicate vegetation response through 'greenness' – the so-called NDVI (Normalised Difference Vegetation Index – cf., Walsh et al. 1994). Field checks are used to determine relations between these indices and what is on the ground (field samples) and these are used to set up a training set of cells so that all other cells in the study area can be classified.

There are several problems with using RS data for mapping vegetation. IR reflectance data are passive attributes; their values not only depend on time of year (which could be useful), but also on cloud cover and moisture stress, and each observation in time needs to be geographically registered. In short NDVI is but one possible, short term indicator of vegetation response to ecological driving forces that are considered in ground-based ecological classifications. It would be useful to have a spatially contiguous, high resolution, cheap attribute that provides information on some of the more important long term ecological conditions affecting plant growth.

DEMs provide ground based data, which are usually strongly spatially correlated, so the results of ecological classifications based on the derivatives of elevation data may be both ecologically sensible and mappable. Recent work in North America on the relations between topographic controls and the spatial distributions of boreal forest tree species has examined the extent to which meaningful topographic and climatic differences can be adequately captured by digital terrain analysis (Mackey et al. 2000; Nemani et al. 1993; Walsh et al. 1994). Mackey et al. (2000) obtained the spatial distributions of Jack Pine (*Pinus banksiana*), trembling aspen (*Populus tremuloides*), and black spruce (*Picea mariana*) from published air-photo interpreted (API) maps, and compared them with 20 m gridded data sets of incoming shortwave solar radiation, topographic wetness index, and local topographic position. The local topographic position attribute varied from 0 to 1 based on the proportion of cells in a user-defined circle that were lower than the centre cell. Using standard combinatorial analysis in GIS, the first two attributes were paired with this local topographic position attribute and used to generate two sets of 10 000 environmental domains by dividing each variable into 100 equally spaced classes across its range. An observed probability of occurrence of each of the target tree species was then calculated for each environmental domain by dividing the frequency of occurrence of the target species by the frequency of occurrence of all grid cells. The calculated probabilities were used to identify the environmental conditions where each of the target species is more or less likely to dominate. The results were promising from the point of view that the distribution of the three tree species in the 2-D environmental domains was consistent with field observations and patterns noted in the ecological literature.

The approach of Mackey et al. (2000) cannot easily be extrapolated to other species and/or regions, however, because of four sets of factors.

(1) It requires high-resolution, independently derived, and accurate maps of existing vegetation across large areas. There are at least two problems here. First, most published regional maps of predicted vegetation utilize some combination of topographic variables derived from DEMs and spectral reflectance variables derived from satellite imagery (which introduces circular reasoning for the problem tackled in this paper). Second, the likelihood that individual species have overlapping property sets (niches) and the role of disturbance (succession) and biotic (competition) factors may im-

pose fundamental limits on the success of predictive mapping in many landscapes (Franklin 1995).

(2) Using simple combinatorial analysis means that because of the huge number of potential combinations, only two variables can be analysed at a time. In practice, this is insufficient because there are many topographic attributes that have ecological significance, and the numbers of classes created by combinatorial analysis grow exponentially. For example, using eight attributes (as we do in this study) each with 10 equally spaced sub-classes would yield 100 million logically separate combinations (10^8). Each of these classes has by definition no overlap with its neighbours, yet it is common knowledge that vegetation associations often gradually merge from one to the other, both in terms of composition and spatial extents. Clearly there is a requirement for a reproducible, data driven method of abiotic landscape classification that can create a reasonable number of classes meaningful in biotic terms that preserve important features of inter-site variability and overlap while only requiring a modest computational effort. In previous work, Mackey et al. (1988) had used agglomerative factor analysis to identify clusters (groups of grid cells that shared similar PER index values) and represent their relative inter-group similarity as a non-overlapping hierarchy, but they had not included detailed information on the spatial variation of topographical control variables, which is the aim of this paper.

(3) Non-systematic and systematic errors occurring in DEMs may confound the expected relationships between vegetation patterns and terrain-controlled site conditions (cf., Wheatley et al. 2000). These problems may be amplified of course when first- and second-order derivatives like slope or convexity are calculated. Additional problems may arise when the analysis is limited to only two or three variables because the key topographic variables used to construct the domains must be selected prior to the environmental domain analysis.

(4) Finally, there are the issues of locating plots accurately and precisely with reference to the topographic data and the possibility that the field observations represent areas smaller than the spatial resolution of the topographic data because the vegetation composition varies over short distances (e.g., Franklin et al. 2000). This is tricky because coarser grid resolutions reduce the magnitude of the first difficulty, but increase the likelihood of mis-classification through short range variation.

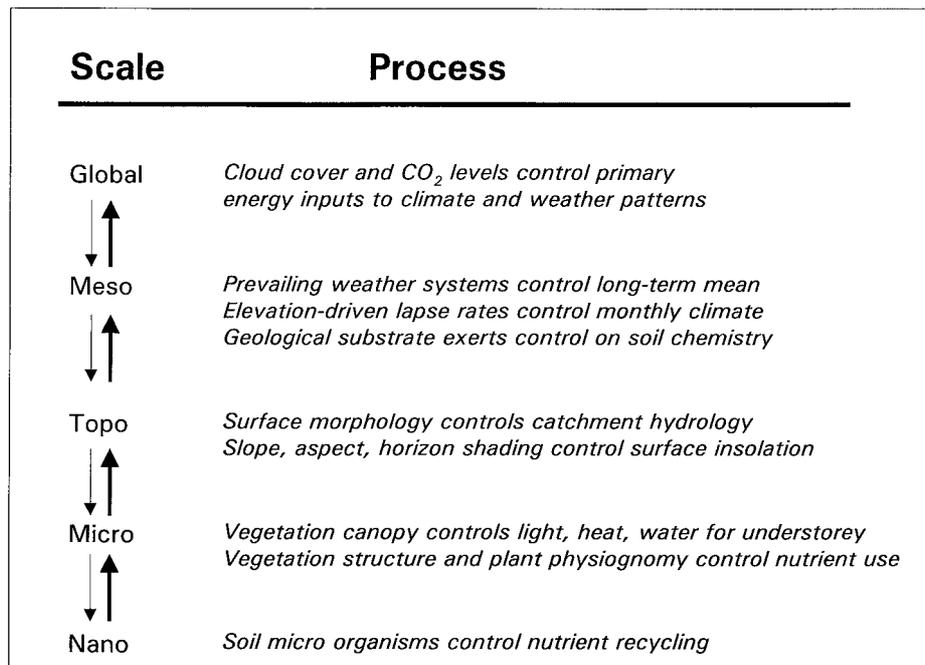


Figure 1. Scales at which various biophysical processes dominate calculation of primary environmental regimes (from Mackey 1996).

The first two problems are potentially the most serious and point to the need to use some other method to set up the classification. This method must be computationally tractable, and ideally should provide an easy-to-interpret classification that offers continuity in both attribute and geographic space that is not constrained by grid resolutions nor by the size of the database. Methods of fuzzy *k*-means classification have been used by other workers to overcome the problem of class overlap (problem 1) but until recently their usefulness has been limited to small data sets (problem 2) and by artefacts induced by deriving attributes from gridded DEMs (problems 3 and 4).

This paper describes the creation and testing of meaningful bio-physical topo-climatic classes that overcome these limitations by using geographical information systems (GIS), spatial sampling methods, statistical modelling of the derived stream topology, and fuzzy *k*-means classification. Using GIS makes it possible to handle very large amounts of spatial data, and to perform a wide range of computations uniformly to spatial or attribute data so that new attributes can easily be computed at the meso- and topo-scales identified in Figure 1.

The method of fuzzy *k*-means serves as an exploratory technique that suggests how best to divide the landscape into meaningful groups, both in terms

of the number of classes and their definition. Given a suitable set of polythetic, overlapping classes, the allocation of an individual z to each of a given set of classes is expressed not in terms of a binary 'Yes' or 'No', but by a set of continuous memberships $0 < \mu_A(z) < 1$ to each of the classes, that add up to 1 (Bezdek et al. 1984; Vriend et al. 1988; Burrough and McDonnell, 1998). The value of $\mu_A(z)$ presents a way of giving a graded answer to the question 'what is the affinity of observation z with the central concept of class A?'. Note that the fuzzy *k*-means approach initially says nothing about geographical contiguity of these optimal, overlapping classes, though if the source data are spatially correlated (as most derivatives from DEMs are) then it is likely that the resulting membership values will also be spatially correlated (Burrough et al. 1997).

Finally, the computed topo-climatic classification is compared with an independently derived land cover data set to investigate the relationships between these topographically-based classes and biological diversity.

To sum up, the questions we address in this paper are not whether the methods presented here give a better ecological classification of forest types than multivariate clustering or logistic models, but whether the fuzzy *k*-means means of classifying the derivatives of DEMs, (a) provides a useful, generic and mappable

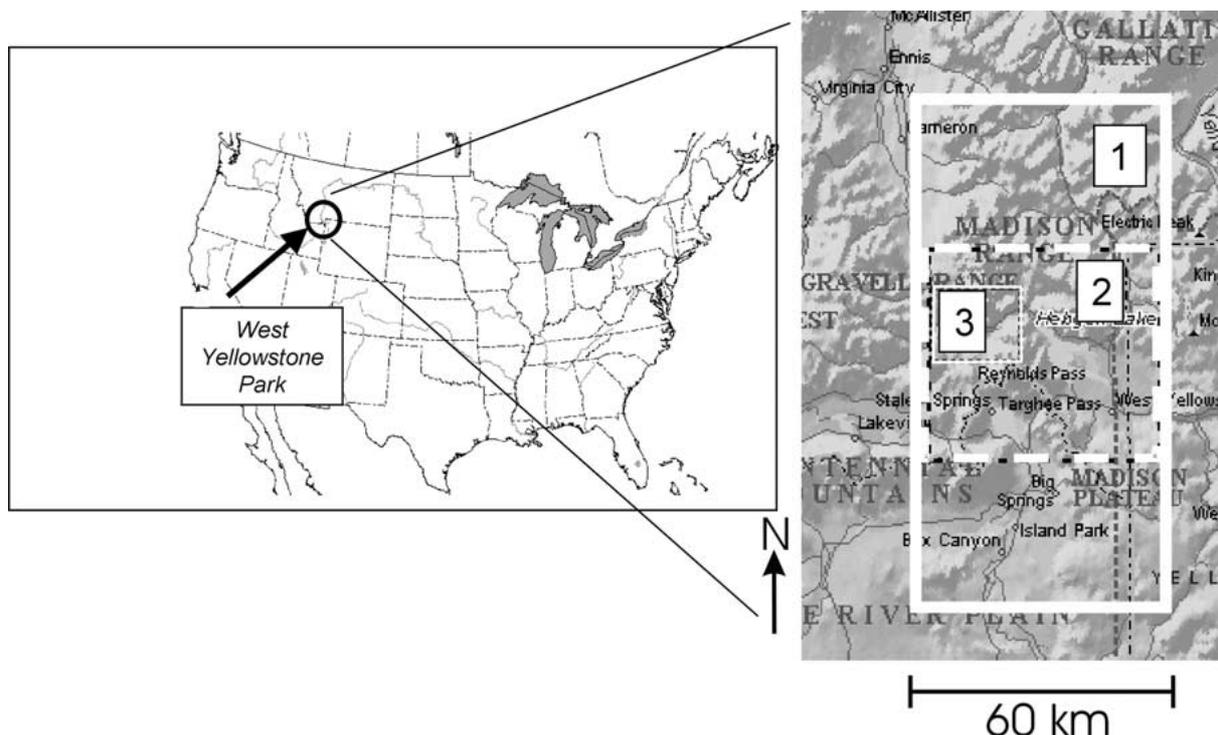


Figure 2. Location of the study area in the West Yellowstone National Park, in southern Montana. The numbers on the detailed map show: 1. The whole 10 000 sq km study area; 2. The central area of 3500 sq km used for setting up the classification; and 3. The area used to illustrate the source data and fuzzy k -means results given in Figures 3, 6 and 7.

procedure that is ecologically meaningful for vegetation response in the long term, (b) makes sense in terms of local field observations and (c) scores at least as well or better than the current NDVI methods.

Study area

We used a 10 000 km² area that includes the 6500 km² Greater Yellowstone study area surveyed by Hansen and Rotella (1998) in their pilot study of bird abundance and diversity. This rectangular-shaped study area combines parts of the Gallatin (GNF) and Targhee National Forests (TNF) and Yellowstone National Park (YNP), and includes the upper catchments of the Gallatin, Madison, and Henry's Fork Snake Rivers (Figure 2). This region is very diverse. The growing season in the high elevation areas of YNP may be as short as two months and the young volcanic soils are very infertile and subject to summer drying (Despain, 1990). These areas are dominated by a crown fire vegetation system with large wildfires occurring at approximately 250-year intervals (Romme, 1982). Over 40 percent of YNP was burned in the exten-

sive fires of 1988 (Christensen et al. 1989) and many of the native species are adapted to this disturbance regime. Some species require periodic fire to prosper. This pattern can be contrasted with TNF to the west of YNP where approximately 55% of the landscape has experienced clear-cut logging (Hansen and Rotella 1998). However, the topography is much more varied and the trees less amenable to harvest in the GNF portion of the study area to the north of TNF and YNP. The growing seasons in the lower valleys of the Gallatin, Madison, and Henry's Fork Snake Rivers approach five months in length. The combination of milder climate and deeper soils generates much higher net primary productivity in these areas compared to the surrounding uplands (Hansen and Rotella 1998).

The biological significance of different cover types was investigated by Hansen and Rotella (1998) who in 1995 sampled approximately 100 species of breeding birds at 97 field sites. They found that bird abundance and richness increased by a factor of two as one moved from lodgepole pine stands at higher to lower elevations. In addition, cottonwood, aspen, and willow stands had much greater bird abundance and richness compared to other coniferous and herbaceous

stand types. Finally, they determined that aspen is a very important population source for yellow warblers in contrast to cottonwood stands that act as a population sink because of lower nest success rates caused by the presence of larger numbers of predators. Based on these findings, Hansen and Rotella (1998) hypothesized that the controls on biodiversity in this region included climate, topography, and soils in addition to disturbance, patch dynamics and biotic interactions, but they were unable to test this hypothesis because of the difficulty of data collection and synthesis. One aim of this paper is to build on their work and show how landforms can be classified to help with: (1) the identification of the topo-climatic controls on land cover; and (2) the development of methods that might be used to generate improved maps of vegetation cover in this region.

Data sources and computational procedures

Our approach has four key steps. First, we built a gridded DEM for the study area and calculated selected topographic attributes. Second, we took a moderately large stratified random sample of the grid cells, extracted their attributes, and submitted this smaller data set to fuzzy *k*-means classification. Third, the parameters of the fuzzy classes were used to classify all unsampled cells on the basis of their topographic attributes thereby creating gridded topo-climatic class maps of large contiguous areas. Fourth, we compared a moderately large data set of point-sampled, spatially referenced vegetation cover classes and co-located M_NDVI data with the crisped fuzzy classes in order to evaluate the usefulness of the landscape classification as an independent predictor of vegetation types.

The digital terrain model and derived attributes

A total of 156 7.5-min USGS 30 m DEMs and 12 0.5° by 1° USGS 1:100 000-scale hydrography files were acquired and used with ANUDEM (Hutchinson 1989) to construct a 100 m square-grid DEM. ANUDEM takes irregular point data or contour data and creates square-grid DEMs at user-defined resolutions. The program automatically removes spurious pits within user-defined tolerances, calculates stream lines and ridge lines from points of locally maximum curvature on contour lines, and (most importantly) incorporates a drainage enforcement algorithm to maintain

fidelity with a catchment's drainage network (Hutchinson 1989). This process eliminated some but not all of the systematic and non-systematic errors that were present in the original USGS DEMs (as discussed below).

Using PCRaster, eight topo-climatic attributes were computed for each cell from the gridded DEM: elevation, slope, profile curvature, plan curvature, distance from ridgelines, total annual incident solar radiation, topographic wetness index, and sediment transport capacity index (Figure 3). All attributes can be easily computed (see Burrough and McDonnell 1998; Moore and Wilson 1992, 1994; Moore et al. 1993, for algorithms). Total annual incident solar radiation was computed instead of aspect, because the latter is measured on a circular scale, and energy input to the terrain is a related, but ecologically more useful and interpretable attribute. The Topographic Wetness Index (TWI) and Sediment Transport Index (STI) were computed using the surface topology of the DEM as determined by the well-known D8 algorithm (Burrough and McDonnell, 1998; Moore et al. 1993). While TWI is defined as a natural logarithmic attribute, the distance from ridgelines and STI were converted to natural logarithms because their distributions were strongly positively skewed. As Table 2 shows, the correlations between these data are not large and therefore they provide largely independent contributions to the classification.

Data quality

No data are perfect and errors in DEMs may produce artefacts that will later show up as outliers in the classification. In this case errors arose because of differences in base heights between adjacent USGS source DEMs. Thanks to the large difference in relief, most errors were concentrated in the large flat area that drains into Hebgen Lake (Figure 4). The first six topographic attributes listed above were computed directly from the derived 100 m DEM and given the data ranges, these are not very sensitive to small, local variations in elevation. Moreover, the affected area contained very few sites where vegetation cover had been recorded so errors arising in these attributes had little impact on the final evaluation (Figure 9).

Topographic wetness and sediment transport capacity indices, however, are very sensitive to the presence of errors in elevation data in areas of low relief, particularly when using the D8 algorithm (Burrough and McDonnell 1998; Wilson et al. 2000). This

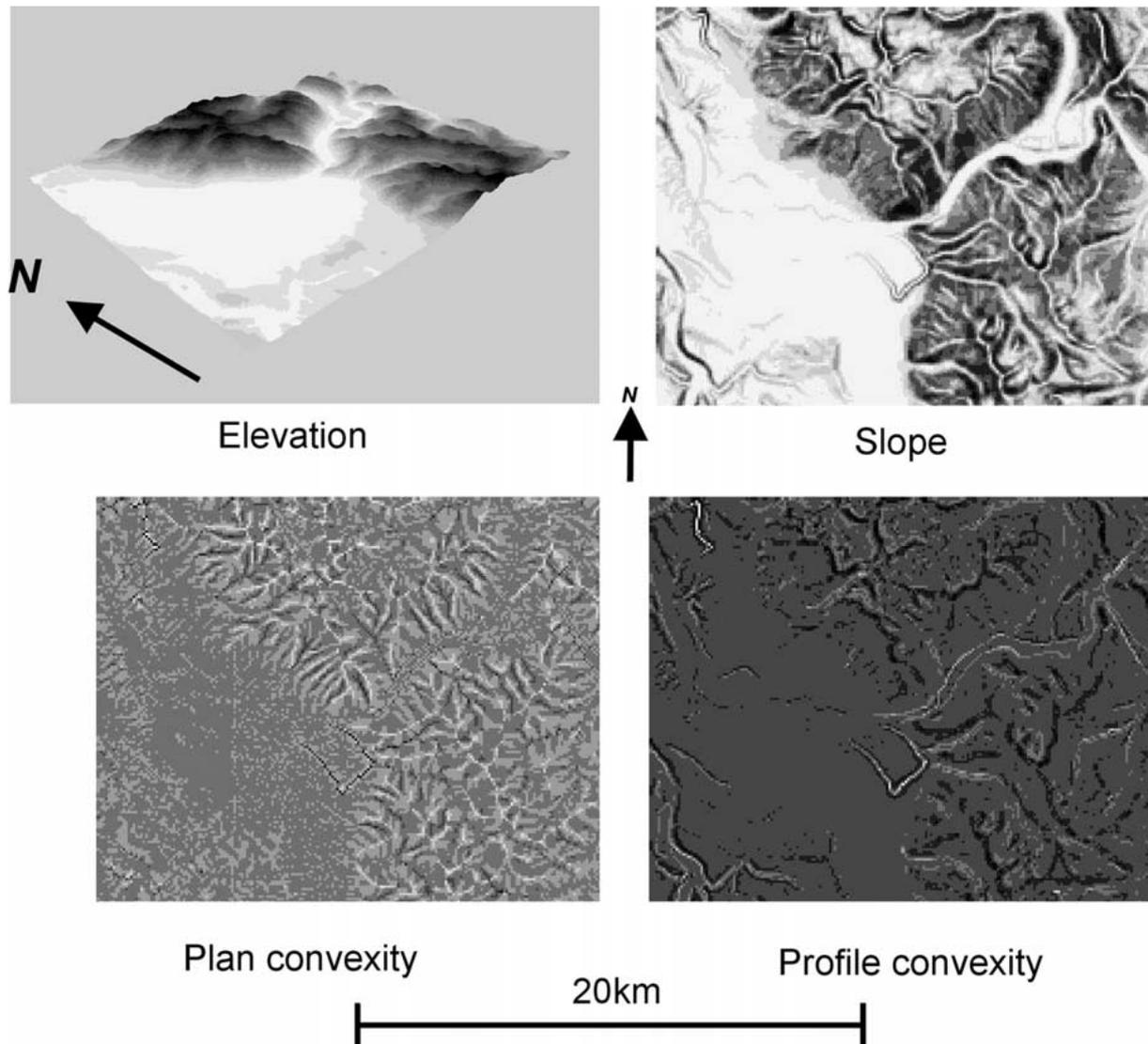


Figure 3a and 3b. Samples taken from area 3 in Figure 2 of grey scale maps of input data used for the fuzzy k -means classification. The elevation map is shown as a 3D image at the top left in Figure 3a – the rest are two-dimensional figures. Grey scales range from low (white) to high (black), except for $\ln(\text{Ridge proximity})$ [top left in Figure 3b] which has reversed tone for clarity. Size of area is 20×16.3 km.

algorithm routes flow from each cell to one of the eight surrounding cells in the direction of steepest descent. The use of eight flow directions typically results in a poor representation of flow networks where the surface is not oriented in exactly one of those eight directions. In addition, this method only represents parallel or converging flow, so that divergent areas in the upper parts of the landscape are not represented accurately (Wilson and Gallant 1998). Moreover, small differences in elevation in almost flat areas can yield very different drainage nets. Therefore we computed mean values of both indices from mean specific catch-

ment areas and slopes, where these mean areas and slopes had been averaged from 20 separate realizations of the drainage topology made by adding a normally distributed RMS error of ± 1 m to each cell of the DEM. This procedure removes the artefacts obtained with a single derivation of the surface topology of a mathematically smooth surface (Burrough et al. 2000).

Fuzzy topo-climatic classification – the training set

We split the study area north-south into three roughly equal quadrangles of approximately 3000–3500 km²

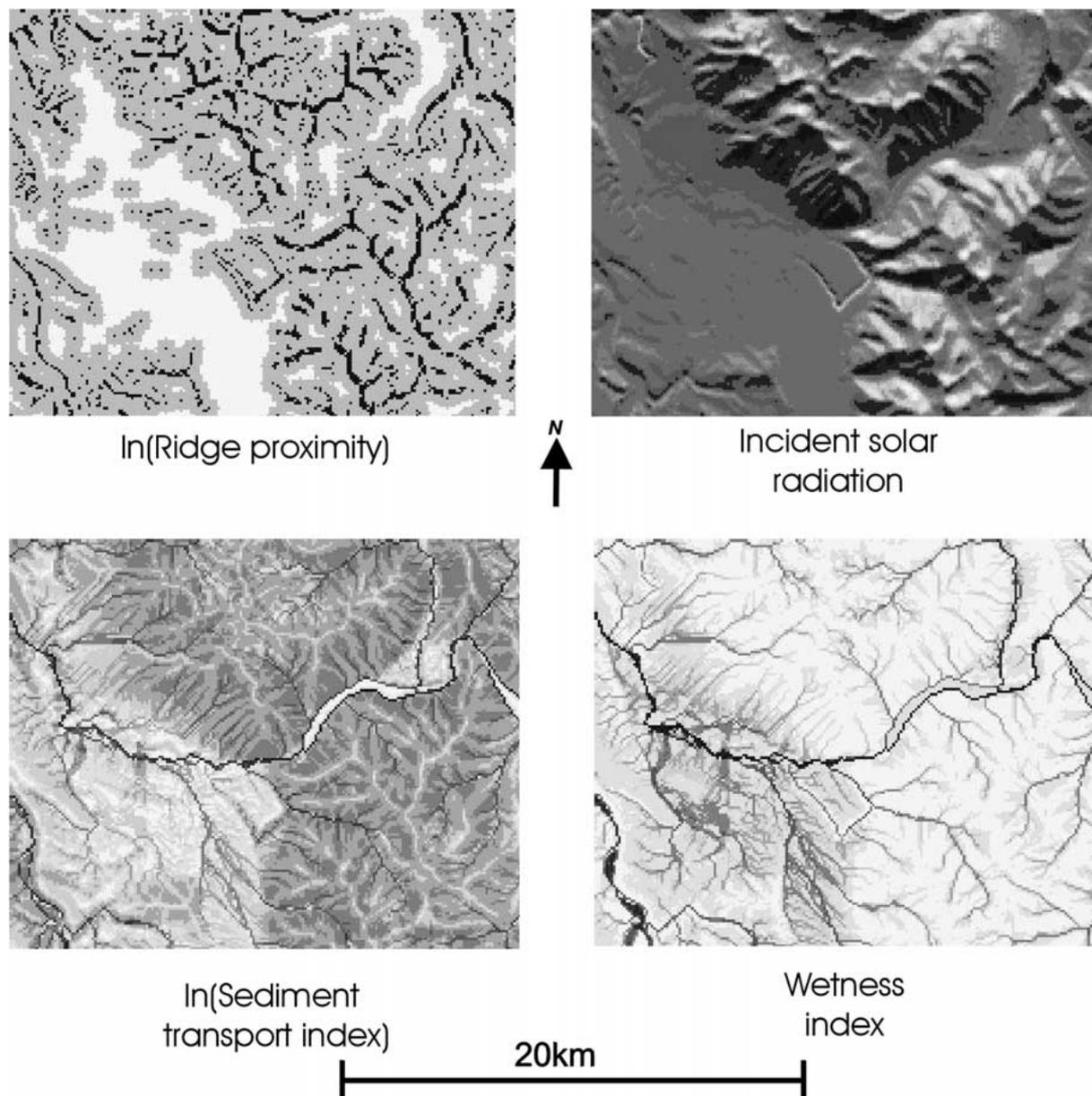


Figure 3b. Continued.

so we could test whether or not a fuzzy classification developed for one area could be extrapolated to adjacent areas (see Hypothesis 1 below). The centre quadrangle of $593 \text{ rows} \times 607 \text{ columns}$ (359951 cells or 3599.5 km^2) was used to derive the classification. A training set of 699 sample points needed to assemble the data for the fuzzy classification was obtained using stratified, random sampling of this centre quadrangle (Figure 4). Table 1 gives the basic statistical

information for these sample data and Table 2 their correlations.

The methods of fuzzy k-means

The fuzzy k-means algorithm is described in detail in Burrough and McDonnell (1998) and Burrough et al. (2000); only the main points are summarized here. Fuzzy *k*-means uses an iterative procedure that usually starts with an initial random allocation of N objects to

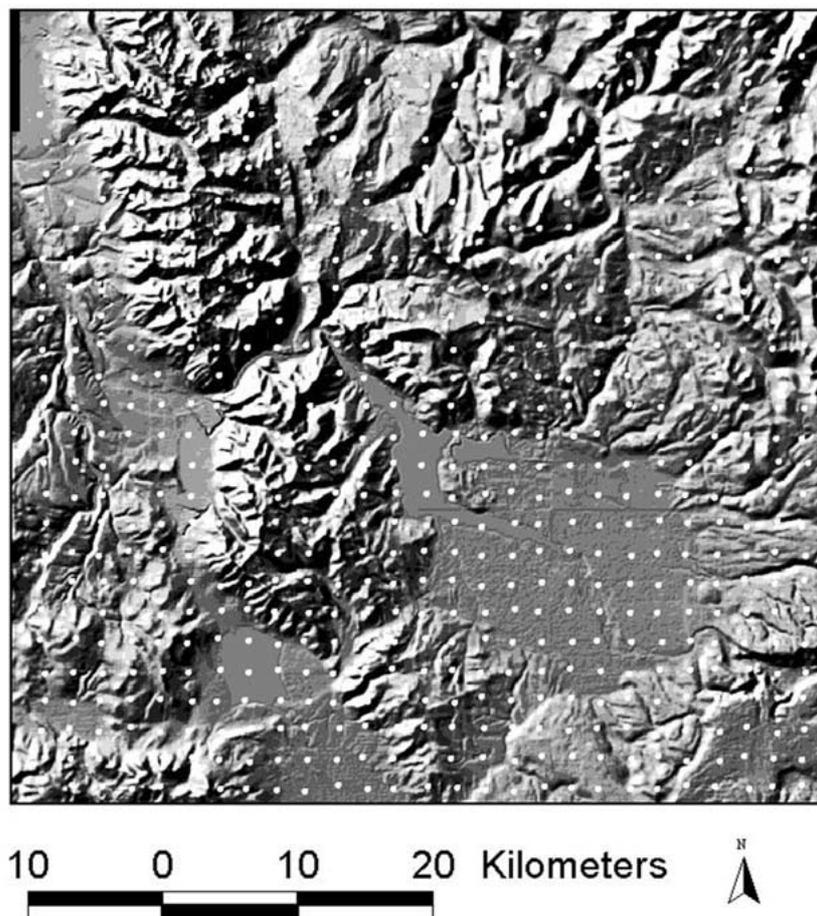


Figure 4. Shaded relief map of the central study area showing the locations of the samples used to extract data for the fuzzy k -means classification. The unevenness in the image in the flat area bottom right is caused by systematic differences in the digitized elevation data between neighbouring map sheets.

Table 1. Basic statistics for the sampled input data ($N = 699$).

Input data and units (where appropriate)	Mean	Standard deviation	Minimum	Maximum
ELEV (m)	2034	291	1781	3223
SLOPE (degrees)	10.83	8.97	0.00	41.25
PROFC (deg m^{-1})	-0.10	1.19	-5.20	6.60
PLANC (deg m^{-1})	-0.02	1.27	-5.25	8.09
RDPRX (ln (m))	5.39	1.60	0.00	8.25
SOLAR ($\text{kJoule y}^{-1} \text{m}^{-2}$)	9.11	1.28	3.51	11.89
WET20 (ln (m degrees $^{-1}$))	12.70	2.19	9.49	22.25
SED20 (ln (m degrees))	5.24	1.40	1.73	9.11

Table 2. Correlations between input data.

Attributes	ELEV	SLOPE	PROFC	PLANC	RDPRX	SOLAR	WET20	SED20
ELEV	–							
SLOPE	0.567	–						
PROFC	0.078	–0.053	–					
PLANC	–0.031	–0.024	–0.447	–				
RDPRX	–0.355	–0.233	–0.483	0.241	–			
SOLAR	–0.124	–0.195	0.000	–0.022	0.113	–		
WET20	–0.421	–0.560	–0.321	0.373	0.393	0.113	–	
SED20	0.321	0.605	–0.348	0.342	–0.028	–0.142	0.177	–

Figures in bold are significant at $\alpha = 0.01$.

K clusters. Given the cluster-allocation (expressed in terms of the memberships μ_{ic} in the range 0–1), the cluster centre C of the c th cluster for the j th attribute x is calculated as a weighted average:

$$C_{cj} = \frac{\sum_{i=1}^N (\mu_{ic})^q x_{ij}}{\sum_{i=1}^N (\mu_{ic})^q}, \quad (1)$$

where the fuzzy exponent q determines the amount of fuzziness or overlap – see next section.

In the next step, objects are reallocated among the classes according to the relative similarity between objects and clusters. In ordinary k -means, the membership μ of the i th object to the c th cluster is determined by:

$$\mu_{ic} = [(d_{ic})^2]^{-1/(q-1)} / \sum_{c'=1}^K [(d_{ic'})^2]^{-1/(q-1)}, \quad (2)$$

where d is the distance measure used for similarity. Using the Diagonal metric (in which attributes are scaled to have equal variance), we obtain with the sample variances s_j^2 :

$$(d_{ic})^2 = \sum_{j=1}^V [(x_{ij} - C_{cj})/s_j]^2. \quad (3)$$

Other metrics that are frequently used are the Euclidian (no scaling) or Mahalanobis (both variance and co-variance are used for distance scaling).

Reallocation proceeds by iteration until a stable solution is reached where similar objects are grouped together in a cluster. Once the sample variances and the optimal class centroids for the sample have been computed, unsampled objects (cells) can also be assigned membership values using equations (2) and (3).

The ratio of the dominant and first sub-dominant membership value for each object is a useful index of

the degree of class overlap. It is defined as:

$$CI = \mu_{i,\max 2} / \mu_{i,\max 1} \quad (4)$$

and is termed the Confusion Index. Mapping the confusion index may indicate parts of the landscape where spatial change in classes is clear and abrupt, or diffuse and vague (Burrough et al. 1997).

The classification may be repeated for a range of numbers of classes. The optimal number of classes can be determined using two parameters that express the overall fuzziness of the classification: these are the partition coefficient F and the classification entropy H . They are computed as:

$$F = 1/N \sum_{i=1}^N \sum_{c=1}^K (\mu_{ic})^2, \quad 1/K \leq F \leq 1, \quad (5)$$

$$H = 1/N \sum_{i=1}^N \sum_{c=1}^K -\mu_{ic} \ln(\mu_{ic}), \quad 1 - F \leq H \leq \ln(K). \quad (6)$$

Both F and H depend on the number of clusters K , hence scaled values are often more useful (Bezdek et al. 1984):

$$F_{\text{scaled}} = (F - 1/K)/(1 - 1/K) \quad (7)$$

$$H_{\text{scaled}} = (H - (1 - F))/(\ln(K) - (1 - F)) \quad (8)$$

Applying fuzzy k -means to the training set

In this study the fuzzy k -means procedure was carried out using the FUZNLM program (van Gaans and Vriend 1995 – basic algorithms taken from Bezdek et al. 1984) on the 699 sample points to yield from two to nine classes. The success of the classification was assessed in terms of the scaled H and F parameters (Figure 5). The maximum F and minimum H

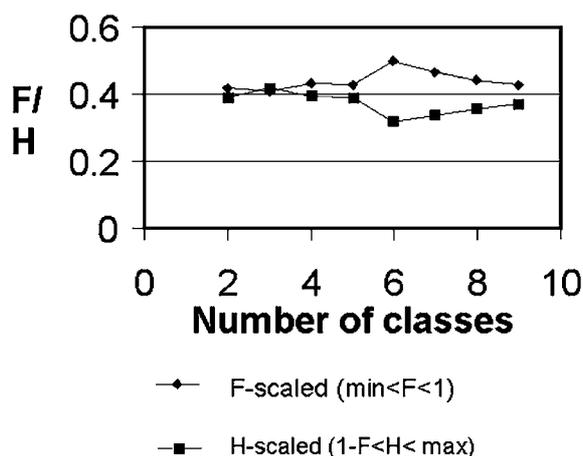


Figure 5. Determining the optimal number of fuzzy classes.

Table 3. Cluster centres for the six topo-climatic classes – units as in Table 1.

Input data	C1	C2	C3	C4	C5	C6
ELEV	2094	2175	2316	2522	2540	2599
SLOPE	1.97	5.04	9.26	21.61	13.10	22.18
PROFC	0.01	-1.03	0.00	-0.47	2.58	-0.44
PLANC	-0.15	1.00	-0.24	0.01	-1.34	0.13
RDPRX	6.23	6.04	5.54	5.40	0.10	5.48
SOLAR	9.44	9.22	9.18	6.79	8.72	10.42
WET20	13.17	16.92	12.02	11.32	10.77	11.47
SED20	3.30	6.64	5.13	6.24	4.77	6.34

values both suggest that six classes are most appropriate. Note that in this study the value of q (the fuzzy overlap in parameter equation (1)) was set at 1.5, this being a much used compromise between $q = 1$ (gives crisp classes with no overlap) and larger values giving more overlap. The same value was used throughout as we were not investigating the effects of modifying the fuzzy k -means parameters on the classification – this would require separate study.

Tables 3 and 4 present the class centroids and class ranges for each attribute for this six-class solution. The class centroids in Table 3 show clearly that fuzzy k -means has established classes with clear differences while Table 4 shows that all classes have considerable overlaps in the ranges of all input data, as was expected. These topo-climatic classes have been given the following generalized names: Class 1 – valley bottoms; class 2 – main drainage lines; class 3 – lower slopes; class 4 – steep, shaded north-facing upper slopes; class 5 – narrow ridges; and class 6 – steep, south-facing, drier upper slopes and broad ridges.

Classifying the remainder of the central quadrangle and the adjacent areas

Each unsampled cell was assigned a membership value for each class using equations (2) and (3) and the parameters of the training set. This procedure yielded six maps (i.e., one for each class – Figure 6). These maps demonstrate that the classes can be clearly interpreted in landscape terms. The maximum membership for each cell, and the confusion index were also computed (Figure 7). Finally, a map of the defuzzified, crisp topo-climatic classification was made by assigning a code to each cell corresponding to the class with which it has the largest membership (Figure 8). Table 5 lists the absolute areas and proportions of the crisped topo-climatic classes for the whole of the central area.

The procedure of assigning membership values to unsampled cells was repeated in exactly the same manner for the adjacent northern and southern quadrangles. A composite map of the whole study area was

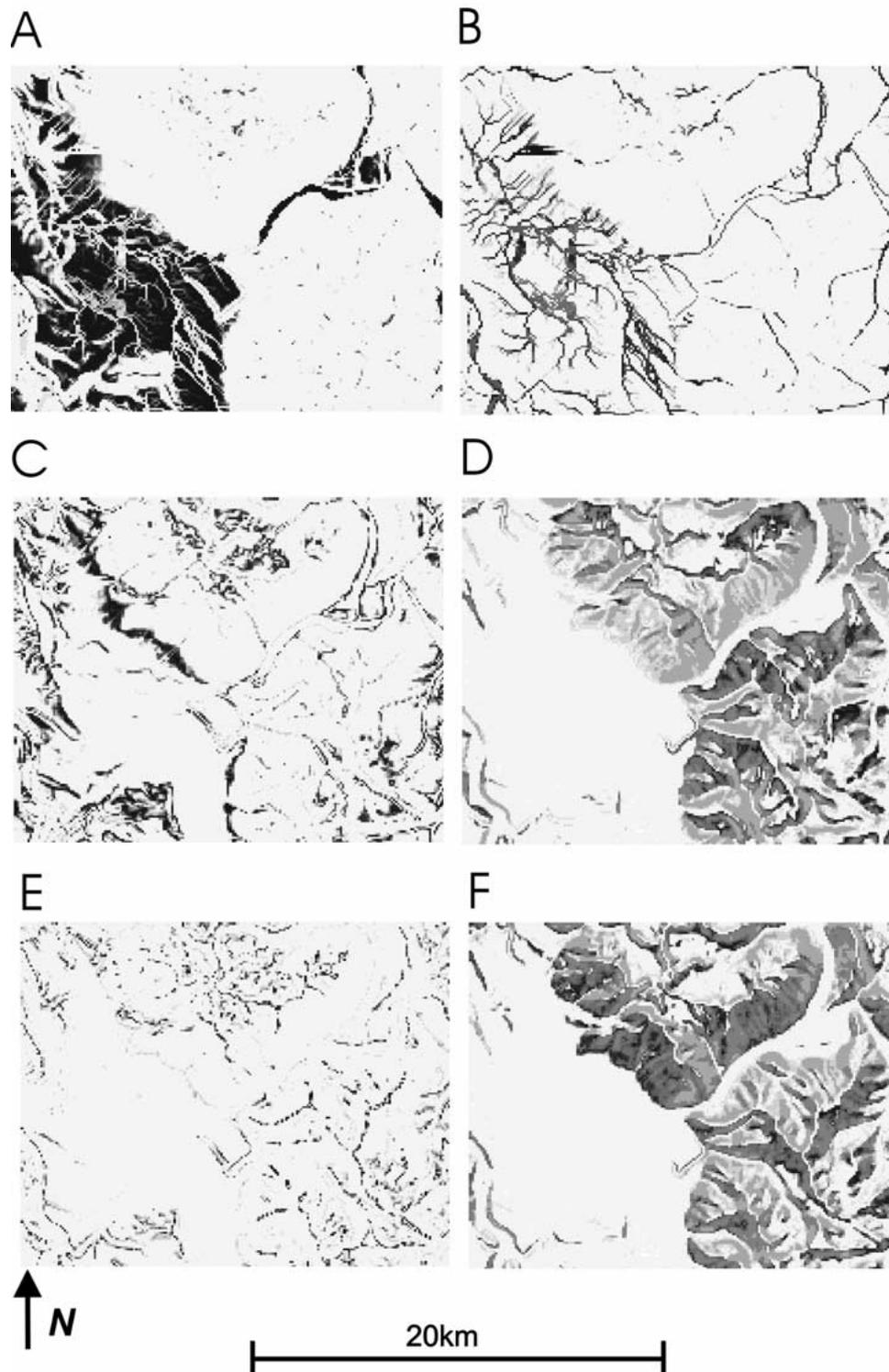


Figure 6. Grey scale maps of area 3 showing individual fuzzy k -means membership values for the six optimal classes (values range from 0 – white to 1 – black) showing how each class represents a different aspect of the landscape. A. Valley bottoms; B. Drainage channels; C. Lower slopes; D. N-facing steep slopes; E. Ridges; F. S-facing lower slopes.

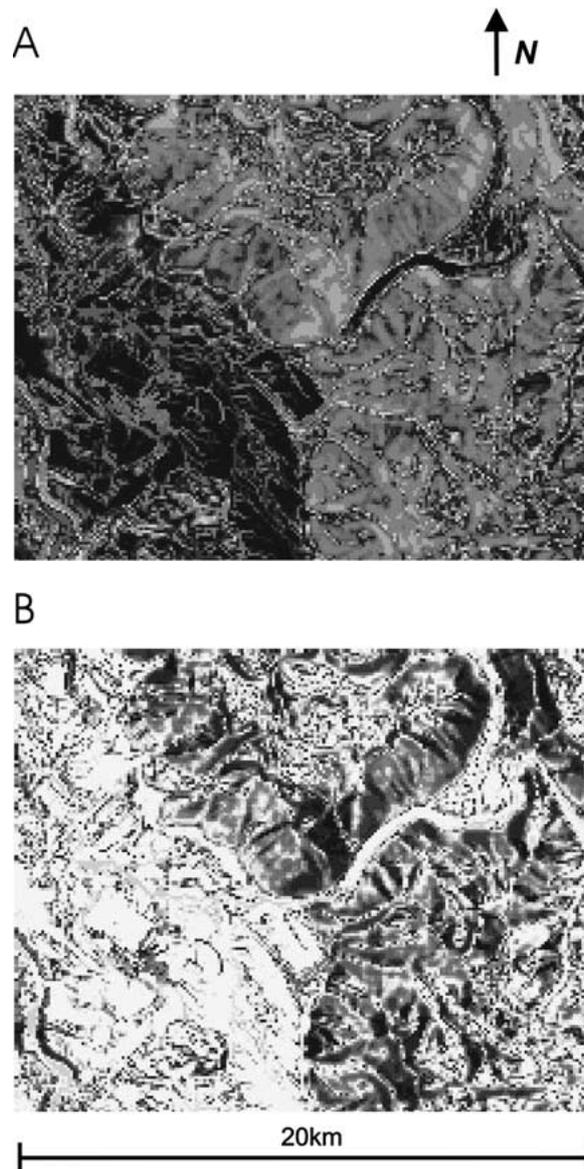


Figure 7. Grey scale maps of area 3 showing A. the variation of maximum membership values, and B. the variation in confusion index (in both cases the scale ranges from 0 – white to 1 – black). These show clearly that lowlands are more easily classified than mountains.

produced by combining the three individual maps in the GIS. The joins between the quadrangles may be seen from the east-west rows of empty cells where edge effects in the computations of slope had generated missing values along the map borders (Figure 8).

Field data on vegetation cover classes

A sample of 575 air-photo interpreted observations of vegetation cover collected independently by one of the authors (Hansen) was used to evaluate the

crisped topo-climatic classes for vegetation mapping. The sample sites for these data were chosen on the basis of size, visibility and ease of access; they were assigned to one of the fourteen land cover classes listed in Table 6. Size was important because the remote sensing classification method of Ma and Redmond (Wheatley et al. 2000) for which the data were originally compiled used a minimum mapping unit of 2 ha in area. Similarly, only those vegetation patches that were visible in both the Landsat TM scene and

Table 4. Cluster ranges for the six topo-climatic classes – units as in Table 1.

Input data	C1	C2	C3	C4	C5	C6
ELEV	1840–2731	1781–2749	1807–2962	1929–3118	1972–3101	2029–3223
SLOPE	0–8.5	0–22.4	2.2–27.2	8.9–34.3	1.8–32.4	8.1–41.3
PROFC	–1.1–1.0	–5.2–0.8	–3.0–1.4	–4.5–1.7	1.0–6.6	–4.1–1.1
PLANC	–1.5–1.3	–0.6–8.1	–5.2–2.3	–5.2–3.4	–5.0–2.1	–3.9–5.8
RDPRX	4.6–8.3	4.6–8.1	4.6–7.2	4.6–7.1	0.0–4.6	4.6–7.1
SOLAR	8.7–10.2	7.4–10.6	7.6–10.3	3.5–8.8	4.7–11.2	8.7–11.9
WET20	11.0–17.4	13.0–22.2	9.8–14.4	9.6–14.5	9.6–12.2	9.5–14.4
SED20	1.7–5.4	4.5–9.1	3.5–6.8	5.4–7.4	2.9–6.0	5.1–8.4

aerial photographs were selected as plots, and ease of access was necessary so that the cover types at some locations could be checked in the field. The test data also included single values for seven Landsat TM bands recorded on June 21, 1991, and several other attributes. We used only the cover classes and NDVI data computed from the TM bands at the 575 sites to carry out the tests described below, referring to this data set as the VEGPNT data throughout the remainder of the paper.

The VEGPNT data were converted to a gridded map showing the locations of the samples (Figure 10). This map was used to determine the crisped topo-climatic class for each VEGPNT sample location, so that cover classes, NDVI data and crisped topo-climatic classes could be compared at all 575 data points.

Testing the methodology: Results

To evaluate the usefulness of the combined fuzzy k -means classification and allocation procedure we examined the following hypotheses, which may be divided into procedural (1+3) and ecological/functional groups (2+4+5):

- *Hypothesis 1.* The fuzzy topo-climatic classification developed for the training data set from the central quadrangle can usefully be extrapolated to the central area as a whole and also to the adjacent north and south areas, thereby reducing computational loads on sampling and classification for areas larger than the study area (cf., Lagacherie et al. 1997).
- *Hypothesis 2.* The spatial distribution of the VEGPNT cover class data is not biased in favour of particular types of topo-climatic classes that may

have been indicated by the fuzzy k -means classification of the morphological data from the DEM. If this is so, then we may use the VEGPNT data with confidence to examine the relations between VEGPNT cover classes and the crisp topo-climatic classes derived using fuzzy k -means.

- *Hypothesis 3.* The VEGPNT point samples have been located in the correct grid cell. If they had been displaced by even a single 100m cell, the topo-climatic classification could be different, particularly in the vicinity of narrow drainage nets and ridges which may be only one cell wide (classes 2 and 5). Therefore we investigated whether moving the VEGPNT samples systematically 100 m east-west or north-south had any effect on the quality of the registration of the VEGPNT data with the crisped fuzzy classes.
- *Hypothesis 4.* The cover class composition of a given crisped topo-climatic class is significantly different from random, but is this true for all cover classes identified, or is it only true for certain classes? Assuming that the VEGPNT data are a representative sample (Hypothesis 2), we wish to test the degree to which the crisped topo-climatic classes give useful information about the cover classes.
- *Hypothesis 5.* The topo-climatic classes will show an affinity with differences in vegetation response as measured in terms of NDVI indices provided by remote sensors.

The fuzzy topo-climatic classes (Hypothesis 1)

The maps of the maximum class memberships and the confusion index show that much of the central area has been unambiguously allocated to a single class (Figures 7a, b). This is a reflection of the strong spatial

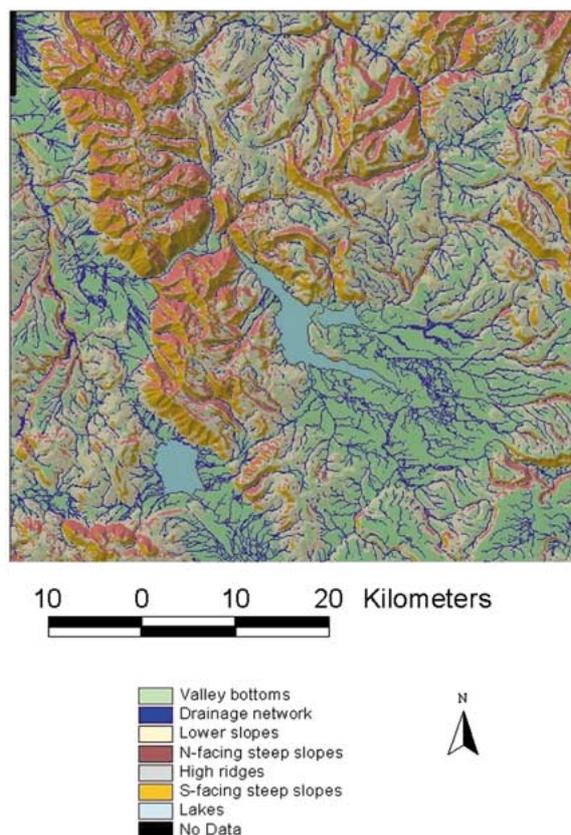


Figure 8. Map of crisped topo-climatic classes for the central quadrangle.

Table 5. Data on the crisped topo-climatic classes for the central quadrangle.

Topo-climatic class	Area (km ²)	Area(%)
1. Valley bottoms	930.31	26.18
2. Drainage channels	369.81	10.41
3. Lower slopes	1048.33	29.51
4. N-facing steep slopes	370.85	10.44
5. Ridges	260.27	7.33
6. S-facing steep slopes	519.33	14.62
Lakes	53.53	1.51

correlation structure of the input data (Burrough et al. 1997). In areas dominated by topo-climatic classes 1, 2, 3 and 5, the thin lines indicating large *CI* show sudden changes where spatial boundaries between classes occur (Figure 7a). The largest values of the confusion index occur within classes 4 and 6, the north and south facing upper slopes. This is presumably because not all upper slopes face strictly neither north nor south;

many have clearly different orientations. In general the spatial distributions of the fuzzy topo-climatic classes are so different that most areas may be assigned to the single 'hard class' with which they have most affinity (Figure 8).

Similarly, it is clear that the classification derived from the training set of 699 points can be applied directly to the northern and southern quadrangles without serious loss of information. For the sample data set and each of the map quadrangles, Table 7 lists the mean value of cell maximum membership for each of the six classes together with the F and H statistics. It shows that all three quadrangles are classified equally well (or even better) than the training set. Overall differences arise from differences in the proportions of mountainous and flat areas, as the latter are more easily classified. Figure 8 shows in detail the transition from the central to the northern quadrangle from which it is apparent that the pattern of classes is disrupted neither at the join between the areas nor within the body of the area. Running the same classification procedures on the adjacent areas produced

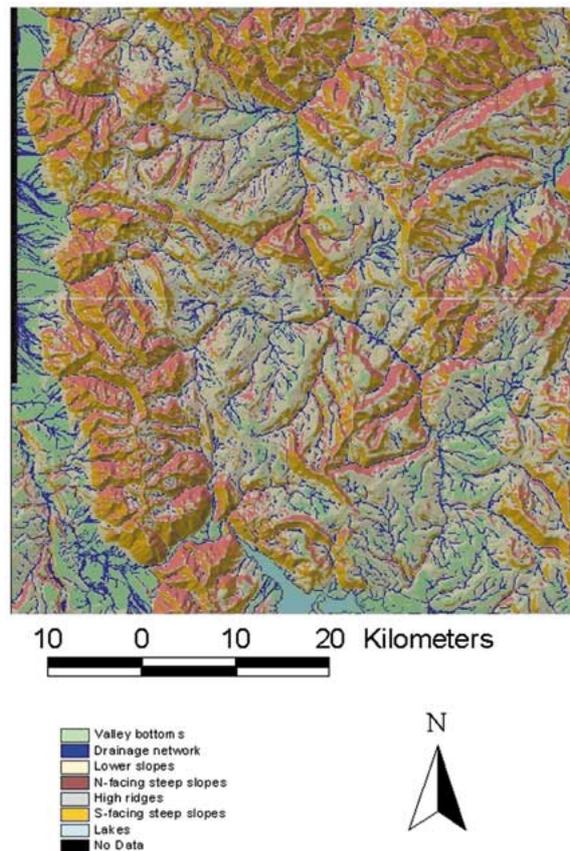


Figure 9. Map showing edge matching between the central and northern quadrangles.

similar results, so we conclude that in this case, the topo-climatic classification based on a single training set from the central quadrangle seems to be suitable to classify a much larger area. In other words, the pattern of landscape represented by the central quadrangle extends at least to the immediately adjacent areas.

Comparison of map of crisped fuzzy k-means and VEGPNT data (Hypothesis 2)

The VEGPNT data had not been collected evenly over the whole study area but were clustered in north- and south-central locations. In particular, the Madison Range in the northwest and the Gallatin Valley to the north were not sampled so they were masked out for the rest of the analysis (Figure 10). All comparisons between VEGPNT data and the map of crisped fuzzy k -means discussed in the remainder of this section were made using this map.

The degree of representativeness of the VEGPNT data was examined as follows. The map of crisped topo-climatic classes was used to predict how many

of the 575 VEGPNT sites should fall in a given topo-climatic class based solely on the proportions of the map it occupies. The numbers of expected VEGPNT samples per topo-climatic class were then compared with the observed number of VEGPNT observations in the same topo-climatic class and evaluated using the χ^2 statistic with 5 degrees of freedom (Table 8). These comparisons show clearly that the proportions of the mapped area falling in the different crisped classes as estimated by the VEGPNT data do not differ significantly from the proportions computed from the map itself. Therefore we can use the VEGPNT samples to test hypotheses about the relations between the occurrence of cover classes in VEGPNT and the crisped topo-climatic classes.

The geometric registration of the point data set (Hypothesis 3)

The geometric registration of the VEGPNT samples was examined by creating eight copies of the VEGPNT data set for which the X and Y coordinates

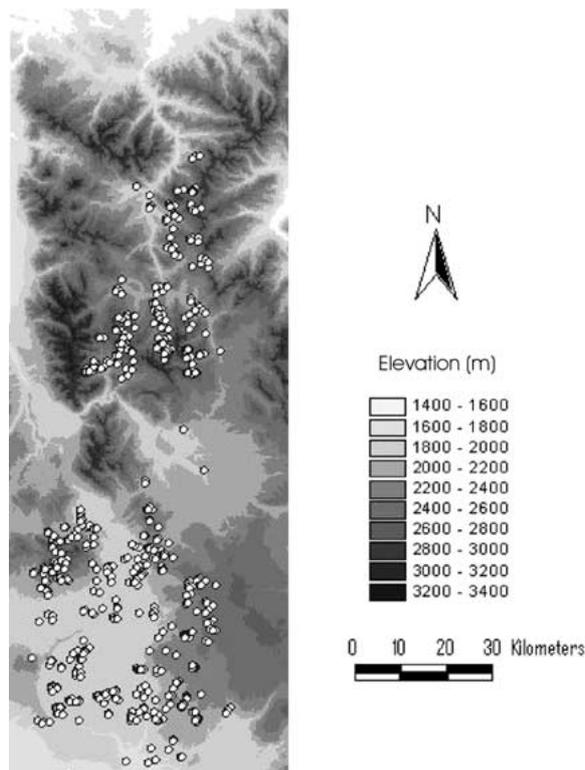


Figure 10. Grey scale map showing elevation of the whole area and the locations of the independent vegetation sample sites used to test the fuzzy k -means classification.

respectively had been offset by 100 m (i.e., 1 grid cell) from the original values. These were then overlaid on the crisped topo-climatic class map to give new values for the topo-climatic classes at the offset data points. Each new data set was compared with the original by computing the χ^2 value between the new proportions of points and the predictions from the whole map to yield a 3×3 matrix of χ^2 values with 5 degrees of freedom centered on the original data (Table 9).

These results demonstrate in all cases that the fit between the proportions of the area estimated by the original VEGPNT data and the proportions derived from the map of crisped classes deteriorates if the VEGPNT data are relocated by 100 m in any direction. The reason seems to be that moving the VEGPNT sites by only 100 m causes the topo-climatic classes representing spatially linear features such as drainage networks (riparian areas) and ridges to be under-sampled. Evidently, these areas had been specifically and correctly sampled by the VEGPNT survey. The conclusion is clear that the registration of VEGPNT is good enough to support an analysis of the

distribution of cover classes with respect to the crisped fuzzy classes.

The value of the crisp topo-climatic classes for predicting land cover (Hypothesis 4)

VEGPNT is now used to examine the hypothesis that the crisped topo-climatic classes provide meaningful predictions of land cover. Figure 11 presents histograms of the proportions of each land cover class associated with each crisped topo-climatic class. These data show that the proportions of VEGPNT sites falling in the crisped topo-climatic classes clearly differ from cover class to cover class. To test the significance of these apparent differences we computed χ^2 values for the following comparisons.

We computed the expected number of cover class samples per crisped topo-climatic class simply as a function of the proportion of the area of the topo-climatic class and the number of VEGPNT samples in the given cover class. We then computed χ^2 for the comparison of these predictions with the actual number of cover class observations given by VEG-

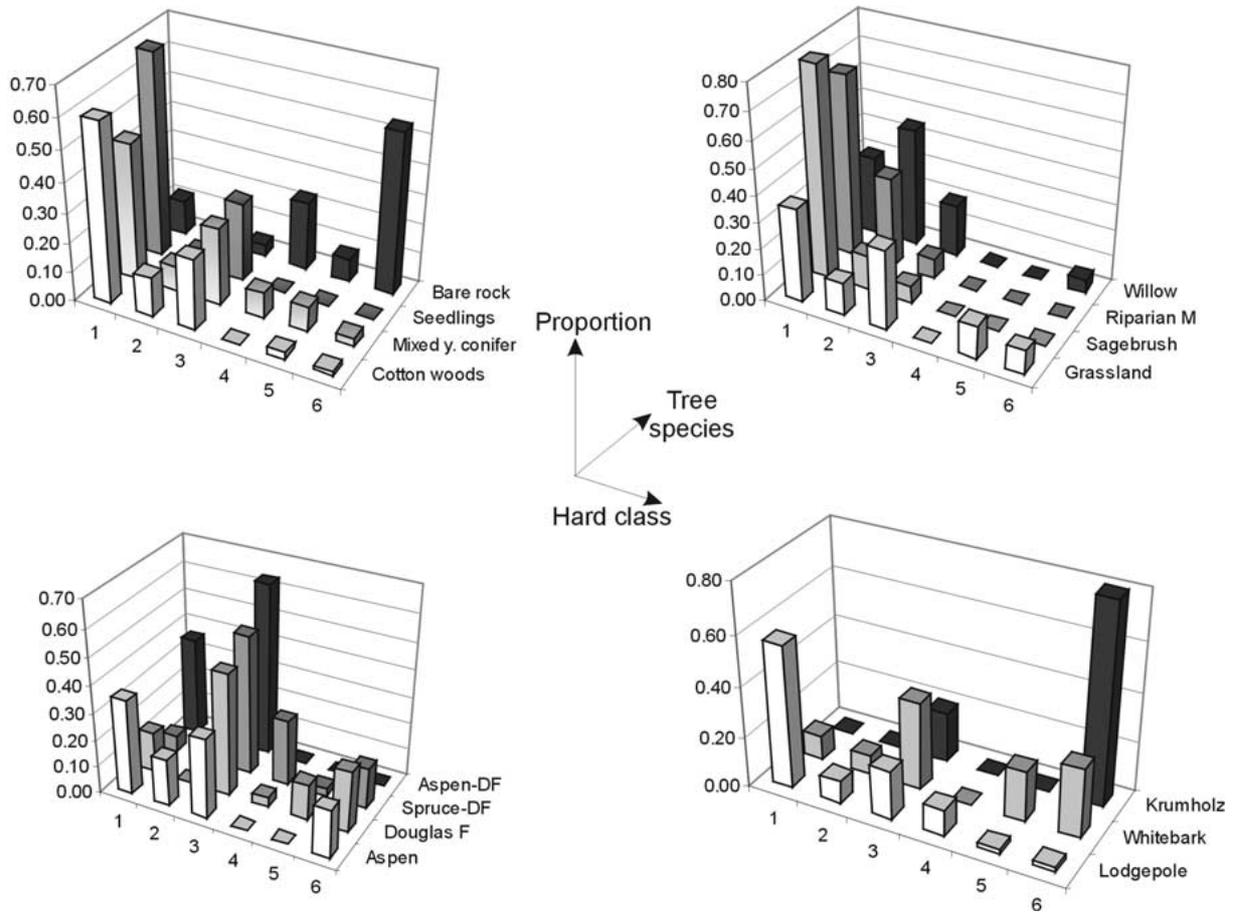


Figure 11. Set of 3-dimensional histograms showing the relations between the proportion of each vegetation type (vertical axis), the topoclimatic classes (long horizontal axis) and the vegetation types at the independently sampled vegetation sites.

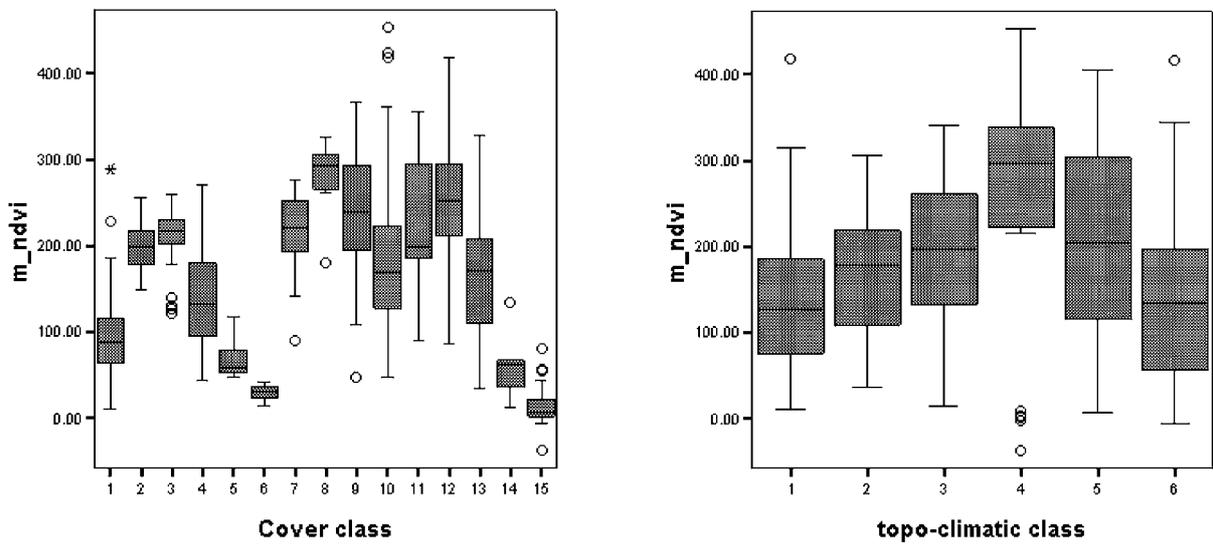


Figure 12. Box and whisker plots of M_NDVI for (a) Cover classes, (b) topo-climatic classes.

Table 6. Land cover classes used in this study.

Class name	Code #	No. obs.	Class descriptions
Cottonwoods	1	90	Cottonwood (<i>Populus</i> spp.) dominated deciduous forest on large floodplains.
Willow	2	26	Riparian communities dominated by several willow (<i>Salix</i>) species.
Riparian meadow	3	29	Moist grassland communities.
Grassland	4	57	Herbaceous communities dominated by grasses and containing forbs.
Sagebrush	5	15	Big sagebrush (<i>Artemisia tridentata</i>) and grass (<i>Festuca idahoensis</i>) dominated communities on drier sites.
Young conifers	6	19	Young conifer seedlings and saplings growing on areas burned in 1988 Yellowstone fires.
Aspen	7	19	Deciduous forests dominated by quaking aspen (<i>Populus tremuloides</i>).
Aspen-Douglas fir	8	11	Mixed deciduous and coniferous forests dominated by these two species.
Douglas fir	9	54	Coniferous forests dominated by Douglas fir (<i>Pseudotsuga mensiesii</i>).
Lodgepole pine	10	125	Coniferous forests dominated by lodgepole pine (<i>Pinus contorta</i>).
Mixed conifer	11	35	Coniferous forests containing lodgepole pine, Englemann spruce (<i>Picea engelmannii</i>), subalpine fir (<i>Abies lasiocarpa</i>), and other tree species.
Spruce-Douglas fir	12	33	Coniferous forests dominated by Engelmann spruce and Douglas fir.
Whitebark pine	13	40	Subalpine savannas dominated by whitebark pine (<i>Pinus albicaulis</i>).
Krumholz	14	5	Subalpine communities containing dwarf whitebark pine, lodgepole pine, and other conifer species.
Bare rock	15	26	Rock outcrops with little or no vegetation cover.

PNT. From the results presented in Table 10, it is clear that there are strong associations between crisped topo-climatic class and cover class for all cover classes except Aspen, Mixed Conifer, Conifer seedlings on burnt areas, and the Aspen-Douglas Fir association. These results support the histograms linking crisp topo-climatic class to cover class given in Figure 10. Therefore we conclude that the fuzzy classification of sites in terms of the abiotic attributes that can be derived from a DEM provides a useful and reproducible tool for predicting land cover in the area studied. Note also that Table 10 shows that the different topo-climatic classes differ greatly in the number and variety of cover classes present.

Fuzzy Classes derived from a topo-climatic analysis show an affinity with differences in vegetation response as measured in terms of NDVI indices provided by remote sensors

As noted in the Introduction, Landsat TM derived estimates of Normalised Difference Vegetation Index (NDVI) have been used to map forest types in mountainous areas. Walsh et al (1994) suggest that for the Glacier National Park, Montana, in spite of large local variations in biophysical conditions, separation of different forest types (e.g. lodgepole pine versus spruce and fir) is possible with the aid of NDVI. As the vegetation data set used by the current study also includes measurements of TM reflectance for 7 bands, NDVI was computed for each test location. Though not a

Table 7. Goodness of classification for the different samples and map quadrangles expressed as Mean μ_{\max} , and in brackets the % of cells, for each data set or quadrangle and topo-climatic class.

Topo-climatic class and code	699 cell training set	Sampled quadrangle	Northern quadrangle	Southern quadrangle	575 point land cover data
1. Valley bottoms	0.825 (22)	0.861 (27)	0.776 (12)	0.900 (65)	0.894 (41)
2. Drainage channels	0.726 (14)	0.768 (11)	0.741 (8)	0.745 (13)	0.707 (10)
3. Lower slopes	0.633 (30)	0.779 (29)	0.733 (29)	0.769 (17)	0.827 (27)
4. N-facing steep slopes	0.663 (12)	0.636 (10)	0.634 (18)	0.614 (1)	0.623 (6)
5. Ridges	0.830 (6)	0.607 (7)	0.610 (9)	0.549 (2)	0.512 (6)
6. S-facing steep slopes	0.640 (15)	0.638 (14)	0.647 (23)	0.655 (2)	0.686 (11)
Overall	0.706 (100)	0.752 (100)	0.689 (100)	0.843 (100)	0.797 (100)
F	0.583	0.663	0.583	0.772	0.717
F scaled	0.500	0.596	0.499	0.727	0.661
H	0.856	0.637	0.787	0.441	0.541
Hscaled	0.319	0.206	0.269	0.136	0.171

Table 8. Observed number of vegetation samples per topo-climatic class and number predicted from map proportions.

Crisp topo-climatic class and code	Number of points from VEGPNT overlay	Number of points predicted by map proportions
1. Valley bottoms	234	221
2. Drainage channels	59	63
3. Lower slopes	153	146
4. N-facing steep slopes	33	45
5. Ridges	33	33
6. S-facing steep slopes	63	67
Total	575	575
χ^2	4.793 n.s.	

major aim of this study, these data provide an opportunity to compare the observed Vegetation cover classes with the crisped topo-climatic classes in terms of differences in NDVI.

Several workers have noted that simple NDVI data may give inflated values for broad leaved understorey and other background vegetation – this seems to be a particular problem in mountain ecosystems (Nemani et al. 1993). Nemani et al. (1993) showed that changes in middle Infra Red (MIR) may be used to correct this when estimating conifer forest leaf area index. In this study we use the changes in MIR to correct the NDVI computed from TM3 (Red) and TM4 (Near Infrared - NIR). We use a formula that was empirically adapted from Nemani et al. (1993) to calculate a MIR-corrected NDVI from TM bands 3, 4, and 5 for each

test location namely:

$$M_NDVI = \frac{(TM4 - TM3) / (TM4 + TM3 + 1) * (256 / (TM5 + 1)) * 100}{(9)} \quad (9)$$

Exploratory data analysis confirmed that M_NDVI displays relations with both vegetation and topo-climatic classes that are both stronger and more capable of being interpreted than the simple NDVI, which returned little more than noise. For the topo-climatic data, simple correlation analysis showed weak but significant positive correlations between M_NDVI and sediment transport and slope, and significant negative correlations with insolation (direct received solar radiance) and elevation (Table 11).

Figure 12 shows box and whisker plots of mean M_NDVI for a) each cover class, and b) each

Table 9. χ^2 values for offset sample points.

Offset (m)	100 W	0	100 E
100 N	15.61**	9.89	14.22*
0	17.61**	4.79	12.56*
100 S	15.56**	11.21*	12.18*

*Significant at 0.05.

**Significant at 0.01.

crisped topo-climatic class. Figure 12a shows clearly that M_NDVI is definitely lowest for cover classes ‘Krumholz and ‘Bare rock’ (14 and 15), and also for the lowland cover classes Cottonwoods, Sagebrush and Young Conifers (1,5,6). M_NDVI values for Willow, Riparian Meadow and Grassland (2,3,4) overlap with the M_NDVI ranges for the forested cover classes, but given their topographic location it is easy to separate these cover classes (and also Krumholz and Bare rock) by combining the M_NDVI data with information from the fuzzy DEM analysis.

In spite of this success, Figure 12 indicates that the four cover classes Douglas fir, Lodgepole Pine, Spruce, and mixtures of these (9, 10, 12,11), show little if any differences with respect to M_NDVI. These cover classes dominate the mountains and cover the major parts of the crisped topo-climatic classes ‘Lower slopes’, ‘N-facing steep slopes’, ‘Ridges’ and ‘S-facing steep slopes’ which together cover almost two-thirds of the original study area. As these topo-climatic classes are so dominant, it would be useful to see if they differ in their mean M_NDVI. To do this, the VEGPNT data were reduced to include only those 134 sites that are located in both the four cover classes and the four topo-climatic classes just listed. Vegetation cover classes including Aspen were omitted because they did not occur on N-facing steep slopes and Ridges. The Whitebark cover class was excluded because there were no occurrences of this cover class on ‘N-facing steep slopes’.

Box plots and ANOVA on these 134 data points showed that it is difficult to distinguish all topo-climatic classes in terms of M_NDVI: a Scheffé test of contrasts (a standard conservative test of differences between groups of means in ANOVA) shows that the groups ‘lower slopes and South-facing steep slopes’ and the groups ‘North-facing steep slopes and Ridges’ have similar values: however, M_NDVI values for the N-facing steep slopes and Ridges are significantly greater than for the lower and S-facing steep slopes. Unfortunately there are too few data to

detect any significant association when the land cover and topo-climatic classes are combined.

Discussion and conclusions

This paper clearly demonstrates that applying the methods of fuzzy *k*-means to topo-climatic derivatives of DEMs results in spatial classifications that are reproducible and intuitively meaningful for attributes other than those from which they were derived. Using the methods of allocation developed here there is no limitation in the size of database or area being classified so that problems related to the size of data sets that can be classified are avoided. Note that all work presented in this paper was accomplished on a modern personal computer so computational requirements are not extreme: the most resources were consumed for the stochastic modelling of the mean topographic wetness and sediment transport indices. This study has also demonstrated that when landscapes contain the same mix of morphological features over large areas, for reconnaissance classification as used in this study, it is not necessary to spread the allowed number of point samples over the whole area. Indeed, it is probably better to ensure that a sufficient number of samples are located in such a way as to pick up important features at different levels of resolution – e.g., drainage lines versus flat riparian areas. Stratified random sampling is an elegant and easy method for sampling an already digitized database though in the field it may prove expensive in difficult terrain to reach all points on the ground.

Although we only used eight attributes in the topo-climatic classification, this number may be easily extended without computational problems. Indeed, we expect that the classification will be improved by adding information from geological and soil maps. Multi-temporal data from remotely sensed images giving NDVI for different times of the year may also provide useful additional information. The topo-climatic classes may also be valuable as ancillary data for creating classifications of existing vegetation cover using remote sensing.

Although the computations are strictly repeatable, errors in classification may arise through systematic or random errors in the data. Certain errors may manifest themselves as outliers in the classification, and can therefore be identified and corrected, but the artefacts produced by algorithms used to generate DEMs and their attributes need careful attention. We believe

Table 10. Results of χ^2 analysis comparing observed number of VEGPNT samples per topo-climatic class and number predicted from proportional map area and points in cover class for combined areas.

Topo-climate class		1	2	3	4	5	6	Sum	χ^2	Significance
Cover class	Code	No. of samples per topo-climatic class								
Cottonwoods	1	50	11	20	0	2	1	84	27.52	0.001
Willow	2	6	9	4	0	0	1	20	28.70	0.001
Rip. Meadow	3	18	9	2	0	0	0	29	27.03	0.001
Grassland	4	20	7	17	0	7	5	56	10.49	0.09
Sagebrush	5	12	2	1	0	0	0	15	12.25	0.02
Young conifers	6	13	1	5	0	0	0	19	9.64	0.1
Aspen	7	6	3	5	0	0	3	17	3.39	>0.1
Aspen-D fir	8	4	0	7	0	0	0	11	9.33	>0.1
Douglas fir	9	8	0	24	2	7	12	53	34.84	0.001
Lodgepole pine	10	70	11	24	13	2	2	123	28.24	0.001
Mixed conifers	11	16	3	9	3	3	1	35	3.69	>0.1
Spruce-D fir	12	2	0	17	8	1	5	33	32.52	0.001
Whitebark pine	13	4	3	14	0	8	11	40	38.17	0.001
Krumholz	14	0	0	1	0	0	4	5	23.33	0.001
Bare rock	15	3	0	1	6	2	14	26	62.38	0.001
% of map		38.5	11.0	25.4	7.8	5.7	11.6	100		

Table 11. Pearson Correlations of topo-climatic variables with M_NDVI (N=566).

	ELEV	SLOPE	PROFC	PLANC	RDPRX	SOLAR	WET20	SED20
R	-.115**	.144**	-.043	-.011	-.074	-.240**	-.074	.213**

**Correlation is significant at the 0.01 level (2-tailed).

more work is necessary on the question of errors in the derivatives of DEMs (cf., Burrough and McDonnell 1998; Davis and Keller 1997; Wilson et al. 2000).

This study has demonstrated the value of combining quantitative map analysis with fuzzy k-means classification for generating and testing hypotheses about the topo-climatic relations between landscape and ecology. On the basis of the results obtained in this study it seems possible to investigate other relations, such as topo-climatic zones and the avian fauna of the area.

This analysis demonstrates that even though simple, single time NDVI values showed no significant differences for land cover or topo-climatic classes, the modified M_NDVI index seemed to be able to distinguish vegetation types that differ greatly in form and morphology. Together with topo-climatic information from the DEM certain land cover classes such as riparian vegetation, krumholz and bare rock were easily and unambiguously distinguished and mapped. The major forest vegetation types that occur on the

upper topo-climatic units of lower slopes, North and South-facing slopes, and ridges, cannot be so easily separated: here the M_NDVI response appears to be dominated by site factors that affect all tree species more or less equally. This problem may have been aggravated because not all test data were collected in the field and unknown errors may have arisen during the aerial photo interpretation. The conclusion is that it is very difficult to distinguish differences in Douglas Fir, Lodgepole and Spruce on mountain sites using single time NDVI or M-NDVI data alone though there are indications that topo-climatic information may be of help (cf., Walsh et al. 1994). Clearly, more research is needed to improve understanding of the relations between NDVI response and topographic controls on vegetation response in these areas.

Acknowledgements

Robin Patten (Department of Biology, Montana State University) and Zhenkui Ma (Wildlife Spatial Analysis Lab, University of Montana) compiled the API-derived vegetation data set used in this study. Robert Snyder (Geographic Information and Analysis Center, Montana State University) helped with the development of the 100 m DEM and derivation of terrain attributes. Ed DeYoung (GIS Lab, University of Southern California) helped with the preparation of the study area map. These tasks were partially funded by grants from NASA (Grant No. NGT-51196) and the National Science Foundation (Grant No. OSR-9554501). This paper was prepared in the summer of 1998 when the second author served as the Belle van Zuylen Chair in the Faculty of Geographical Sciences at the University of Utrecht. The support from all of these sources is gratefully acknowledged.

References

- Albrecht, K.-F., 1992. Problems of modelling and forecasting on the basis of phenomenological investigations. *Ecol. Model.* 63: 45–67.
- Austin, M.P., Cunningham, R.B. and Fleming, P.M. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55: 11–27.
- Bezdek, J.C., Ehrlich, R., Full, W. 1984. FCM: The fuzzy c-means clustering algorithm. *Comp. Geosci.* 10: 191–203
- Burrough, P.A. and McDonnell, R.A. 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford.
- Burrough, P.A., van Gaans, P.F.M. and Hootsmans, R.J. 1997. Continuous classification in soil survey: Spatial correlation, confusion and boundaries. *Geoderma* 77: 115–35.
- Burrough, P.A., van Gaans P.F.M. and MacMillan, R.A. 2000. High-resolution landform classification using fuzzy *k*-means. *J. Fuzzy Sets Syst.* 113: 37–52.
- Christensen, N.L. and 12 others. 1989. Interpreting the Yellowstone fires. *BioScience* 39: 678–685.
- Davis, T.J. and Keller, C.P. 1997. Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: Slope stability predictions. *Int. J. Geogr. Inf. Sci.* 11: 409–434.
- Despain, D. 1990. Yellowstone Vegetation. Roberts Reinhart Publishers, Boulder, CO.
- Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. and Paulißen, D. 1992. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18: 1–248.
- Fitzgerald, R.W. and Lees, B.G. 1996. Temporal context in floristic classification. *Comp. Geosci.* 22: 981–994.
- Franklin, J. 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* 19: 474–499.
- Franklin, J., McCullough, P. and Gray, C. 2000. Terrain variables used for predictive mapping of vegetation communities in southern California. *In* Terrain Analysis: Principles and Applications. Edited by J.P. Wilson and J.C. Gallant. John Wiley and Sons, New York, pp. 331–354.
- Gerrard, R.A., Church, R.L., Stoms, D.M. and Davis, F.W. 1997. Selecting conservation reserves using species covering models: Adapting the ARC/INFO GIS. *Transactions GIS* 2: 45–60.
- Hansen, A.J. and Rotella, J.R. 1998. Abiotic factors and biodiversity. *In* Managing Forests for Biodiversity. Edited by M.L. Hunter. Cambridge University Press, Cambridge.
- Hosmer, D.W. and Lemeshow, S. 1989. Applied Logistic Regression. John Wiley and Sons, New York.
- Hutchinson, M.F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *J. Hydrol.* 106: 211–232.
- Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, O.F.R. (eds). 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge.
- Koestler, A. 1967. The Ghost in the Machine. Macmillan, New York.
- Lagacherie, P., Cazemier, D.R., van Gaans, P.F.M. and Burrough, P.A. 1997. Fuzzy *k*-means clustering of fields in an elementary catchment and extrapolation to a larger area. *Geoderma* 77: 197–216.
- Mackey, B.G. 1996. The role of GIS and environmental modeling in the conservation of biodiversity. *In* Proceedings of the Third International Conference Integrating GIS and Environmental Modeling, Santa Fe, NM, 21–25 January, 1996. Edited by Goodchild et al. Santa Barbara, National Center for Geographic Information and Analysis, University of California (<http://www.ncgia.ucsb.edu>)
- Mackey, B.G., Nix, H.A., McMahon, M.F. and Fleming, P.M. 1998. Assessing the representativeness of places for conservation reservation and heritage listing. *Environ. Manag.* 12: 501–514.
- Mackey, B.G., Mullen, I.C., Baldwin, K.A., Gallant, J.C., Sims, R.A. and McKenney, D.W. 2000. Towards a spatial model of boreal forest ecosystems: The role of digital terrain analysis. *In* Terrain Analysis: Principles and Applications. Edited by J.P. Wilson and J.C. Gallant. John Wiley and Sons, New York, pp. 391–422.
- Moore, I.D. and Wilson, J.P. 1992. Length-slope factors for the revised universal soil loss equation: simplified method of estimation. *J. Soil Water Cons.* 47: 423–428
- Moore, I.D. and Wilson, J.P. 1994. Reply to 'Comment on Length-slope factors for the Revised Universal Soil Loss Equation: Simplified method of estimation' by George R. Foster. *J. Soil Water Cons.* 49: 174–180.
- Moore, I.D., Grayson, R.B. and Ladson, A.R. 1991. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Proc.* 5: 3–30.
- Moore, I.D., Turner, A.K., Wilson, J.P., Jenson, S.K. and Band, L.E. 1993. GIS and land surface-subsurface modelling. *In* Environmental Modeling with GIS. Edited by M.F. Goodchild, B.O. Parks and L.T. Steyaert. Oxford University Press, New York, pp. 197–230
- Nemani, R., Pierce, L., Running, S. and Band, L. 1993. Forest ecosystem processes at the watershed scale: sensitivity to remotely-sensed leaf-area index estimates. *Int. J. Remote Sensing* 14: 2519–2534.
- Romme, W.H. 1982. Fire and landscape diversity in subalpine forests of Yellowstone National Park. *Ecol. Monog.* 52: 199–221.
- Rossi, R.E., Mulla, D.J., Journel, A.G. and Franz, E.H. 1992. Geostatistical tools for modelling and interpreting ecological spatial dependence. *Ecol. Monog.* 62: 277–314.
- Tilman, D. 1994. Competition and biodiversity in spatially structured habitats. *Ecology* 75: 2–16.

- van Gaans, P.F.M. and Vriend, S.P. 1995. FUZNLMEEX computer program. Utrecht, Department of Physical Geography, Utrecht University.
- Vriend, S.P., Gaans, P.F.M. van, Middelburg, J. and de Nijs, A. 1988. The application of fuzzy c-means cluster analysis and non-linear mapping to geochemical datasets: Examples from Portugal. *Appl. Geochem.* 3: 213–224.
- Walsh, S.J., Butler, D.R., Brown, D.G. and Bian, L. 1994. Form and pattern in the alpine environment: an integrated approach to spatial analysis and modelling in Glacier National Park, USA. In: *Mountain Environments & Geographic Information Systems*. Edited by M.F. Price and D.I. Heywood. Taylor and Francis, London.
- Webster, R. and Burrough, P.A. 1972. Computer-based soil mapping of small areas from sample data. *J. Soil Sci.* 23: 210–234.
- Wesseling, C.G., Karssenber, D., Burrough, P.A. and van Deursen, W.P.A. 1996. Integrating dynamic environmental models in GIS: The development of a dynamic modelling language. *Trans. GIS* 1: 40–48.
- Wheatley, J.M., Wilson, J.P., Redmond, R.L., Ma, Z. and DiBenedetto, J. 2000. Automated land cover mapping using Landsat TM images and topographic attributes. In *Terrain Analysis: Principles and Applications*. Edited by J.P. Wilson and J.C. Gallant. John Wiley and Sons, New York, pp. 355–390.
- Wilson, J.P. and Gallant, J.C. 1998. Terrain-based approaches to environmental resource evaluation. In *Landform Monitoring, Modelling, and Analysis*. Edited by S.N. Lane, K.S. Richards and J.M. Chandler. John Wiley and Sons, Chichester, pp. 219–240.
- Wilson, J.P., Repetto, P.L. and Snyder, R.D. 2000. Effect of data source, grid resolution, and flow routing method on computed topographic attributes. In *Terrain Analysis: Principles and Applications*. Edited by J.P. Wilson and J.C. Gallant. John Wiley and Sons, New York, pp. 133–162.