Alternative Criteria for Defining Fuzzy Boundaries Based on Fuzzy Classification of Aerial Photographs and Satellite Images

Jingxiong Zhang and Roger P. Kirby

Abstract

Results from an empirical test, using aerial photographs and satellite images of an Edinburgh suburb, show that fuzzy boundaries of land cover can be derived by using the three criteria of maximum fuzzy membership values, confusion index, and measure of entropy, with only small differences, and that slicing based on the maximum fuzzy membership values provides the easiest and most straightforward solution. This result demonstrates the suitability of using both a crisp classification and its underlying uncertainty map for deriving fuzzy boundaries at different thresholds; together, they provide flexible and compact management of categorical map data. Distinctions between fuzzy boundaries and probabilistic boundaries (such as epsilon error bands) are highlighted, thus providing useful insights to exploring heterogeneous spatial data of the real world.

Introduction

Categorical maps are a fundamental source of spatial information in geographic information systems (GISs), depicting distributions of discrete or continuous attributes in the form of exhaustive, non-overlapping areal units separated by boundary lines. The boundaries are represented cartographically by precisely defined lines of zero width, but it is frequently misleading and inaccurate to represent adjacent areal units on categorical maps in this way (Mark and Csillag, 1989). One alternative is the use of fuzzy boundaries, the subject of this paper.

On categorical maps, the inaccuracy in boundary positions is commonly known as positional error, while the inaccuracy in categorical labeling is called attribute error (Chrisman, 1982). Positional errors in categorical maps are effectively described by epsilon error band models, while attribute errors are discussed by using the concepts of frequency and, hence, probability (Perkal, 1956; Guptill and Morrison, 1995; Goodchild and Hunter, 1997). Estimation of epsilon error band widths has mainly been based on checking test positional data against an independent set of reference data, while attribute errors are estimated by constructing error matrices, based on comparing a classified data set and, again, the reference classification data (Chrisman, 1982; Congalton, 1991).

Conventional categorical mapping assumes an object-based view of reality, that is, the real world regarded as being occupied by a set of discrete point, line, and areal objects (Goodchild, 1989). Clearly, an object-based model is suitable for spatial entities whose boundaries are well defined and for which attributes are exactly valued (NCDCDS, 1988). For example, for individual land parcels, attributes such as ownership, land price, and tax liability can be exactly evaluated, and accurate boundary lines can be drawn. Corresponding to the assumption of discreteness underlying object-based models is the crisp set theory, by which boundary lines are delineated with precision subject to human and machine limitations, and only single category memberships are allowed for objects, implying that an object is either correctly labeled or totally misidentified. Probabilistic concepts and methods are therefore useful for describing attribute errors as well as positional errors in object boundaries.

In the case of well-defined spatial entities, position and attribute are discussed separately. But many spatial phenomena, including for example some of those in lithological, pedological, vegetational, and land-cover categories, may be poorly investigated and defined, so that position and attribute are not easily separable. In these cases, when specific discrete objects are being referred to, it is usually only meaningful to discuss attribute errors, and, by extension, any discussion of errors for poorly defined phenomena is centered on attributes.

Any land-cover classification should reflect the resolution of the sensors collecting the attribute data, as was recognized by Anderson *et al.* (1976) in proposing a four-level hierarchical land-cover classification designed with different sensors chosen to match the level of classification in mind. But even with this safeguard, many spatial phenomena have poor areal definition at all scales, and are therefore fuzzy (Altman, 1994; Burrough, 1996). The complexity of geographical phenomena such as land cover and soil results, for example, in many mixed pixels on remotely sensed images of coarse spatial resolution (Campbell, 1987). Even aerial photographs with high resolution do not necessarily resolve all the detail required and, however much the scale increases, spatial heterogeneity in the real world will still exist. Therefore, a model based on discrete objects works poorly for fuzzy phenomena; a more general

J. Zhang is with the Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan Technical University of Surveying and Mapping, 39 Luo-Yu Road, Wuhan 430070, China (jxz@wtusm.edu.cn).

R.P. Kirby is with the Department of Geography, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK (rpk@geovax.ed.ac.uk).

Photogrammetric Engineering & Remote Sensing
Vol. 65, No. 12, December 1999, pp. 1379–1387.

0099-1112/99/6512–1379$3.00/0

mechanism for depicting the fuzziness inherent in many phenomena is through fuzzy set theory as opposed to crisp set theory (Kaufmann, 1975). For this reason, a more sensible representation scheme for fuzzy phenomena (attributes) is by the use of fuzzy categorical maps where each location possesses multiple, partial memberships of all the candidate classes under consideration.

Fuzzy categorical maps have been promoted by researchers such as Fisher and Pathirana (1989), Berry (1993), and Lowell (1994). The widely recognized superiority of fuzzy over crisp categorical maps is due to their representational and analytical capability. A disadvantage is the increased requirement for storage: in the case of c classes, there are c layers of raster data as opposed to one layer of vector data (Zhang, 1996). While this extra cost for expanded storage may well be justified by the need for information on errors in spatial databases, it is useful also to seek as much storage compactness as possible within a fuzzy structure, and this may be achieved by reversing the process to define more discrete single-class areas within the general fuzzy structure. More importantly, for fuzzy phenomena where attribute uncertainty is of prime concern, there seems to be merit in achieving some balance between information richness of fuzzy maps and reasoning ability of discrete objects. Defining fuzzy boundaries thus serves to transfer uncertainty from attribute to positional domains, making it possible to compare between probabilistic epsilon error bands and fuzzy boundaries, highlighting the differences between them.

The interesting question arises as to how fuzzy boundaries may be derived from fuzzy categorical maps. Fuzzy boundaries are referred to here as boundaries of non-zero widths on "defuzzified," i.e., classified, fuzzy categorical maps. Research has been carried out in modeling fuzzy boundaries of different natures (Edwards and Lowell, 1996; Wang and Hall, 1996; Kiiveri, 1997). Two examples of modeling fuzzy boundaries based on a more theoretical framework, by Burrough (1996) and Lagacherie *et al.* (1996), advocated the use of continuous fields as opposed to discrete objects in an attempt to model uncertainties.

This paper seeks to redefine possible ways by which fuzzy boundaries may be derived from fuzzy categorical maps. In the next section describing the concepts underlying fuzzy categorical maps and fuzzy boundaries, emphasis is placed on how fuzzy boundaries can be generated quantitatively from a slicing process by using three criteria: the maximum fuzzy membership values (FMV), confusion index, and measure of entropy. Then follows an empirical test in which the value of the proposed approaches is illustrated by examples in the context of suburban land-cover mapping, in which real data sets of both raster and vector format over an area of varied land-cover types were incorporated. The conclusions stress the difference between probabilistic and fuzzy boundaries, and suggest the theoretical and practical importance of fuzzy boundaries for categorical mapping.

Concepts and Methods

Fuzzy Categorical Maps

The basis for fuzzy categorical maps is the concept of a field-based model, which conceives of the real world as a set of single-valued functions defined at all locations. Both numerical and categorical variables, commonly known as attributes in object domain, are relevant, elevation and rainfall being examples of the former, and land cover and soil type being examples of the latter (Goodchild, 1989). For categorical variables, every point of a field represents a discrete outcome such as a nominal or an ordinal label in a classification system. Considering a categorical variable to contain c classes, this variable can be viewed as a multi-categorical field $p_i(x)$, where $p_i(x)$ represents the probability of point x belonging to a candidate class i ($i = 1, 2, \ldots, c$). It is required that the probabilities range from 0.0 to 1.0, and sum to 1.0 across all the classes at any point. As the term "fuzzy" is quite realistic, the values of $p_i(x)$ are often known as fuzzy membership values (FMVs) (Lowell, 1994).

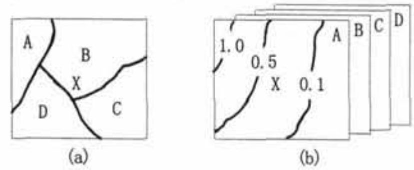

Figure 1. Two distinctive views of categorical mapping: (a) object model, and (b) multi-categorical field model.

Figure 1 is a comparative illustration of an object-based *versus* a multi-categorical field-based categorical variable with four possible classes: A, B, C, and D. Figure 1a shows a polygonal categorical map where only a single class is allowed at each location: location X belongs to class B with full membership, and zero membership in all other classes. On the other hand, under a multi-categorical field model (Figure 1b), the membership vector for location X might be {0.5/A, 0.3/B, 0.1/C, 0.1/D}. This latter view provides the basis for fuzzy categorical maps.

There are different causes of fuzziness, one common cause being related to conceptual gradation between classes. Urbanization and rainfall may, for example, each be described imprecisely as low, moderate, or high. Figures 2 and 3 indicate such class memberships by isolines and distance-decay curves, respectively; the exact function of these obviously useful graphical devices is more fully described later. Mixed pixels are another cause of fuzziness, especially in remotely sensed images of spatial resolution too coarse to resolve the ground detail sought, resulting in pixels containing mixtures of ground cover types. Mixed pixels are referred to by Franklin and Woodcock (1997) as multiscale data with respect to spatial and categorical resolution. A final source of fuzziness is where areas of one cover type are poorly represented by the training data, for example, an unusual roof material in an urban area. This source of fuzziness is relevant in Figure 2 where a fuzzy pixel p that falls near one class only would probably not be a mixed pixel or even a conceptual gradation to one of the other classes (assuming that the list of classes displayed is exhaustive).

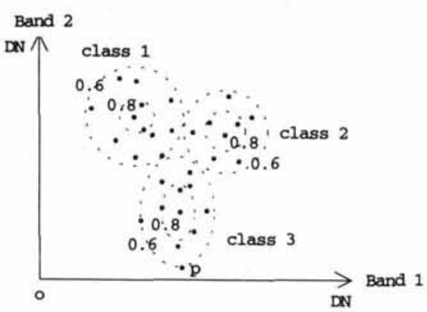

Figure 2. An example of fuzzy 3-means clustering in a two-dimensional spectral space with dashed lines representing contours of fuzzy membership values (DN: digital number).

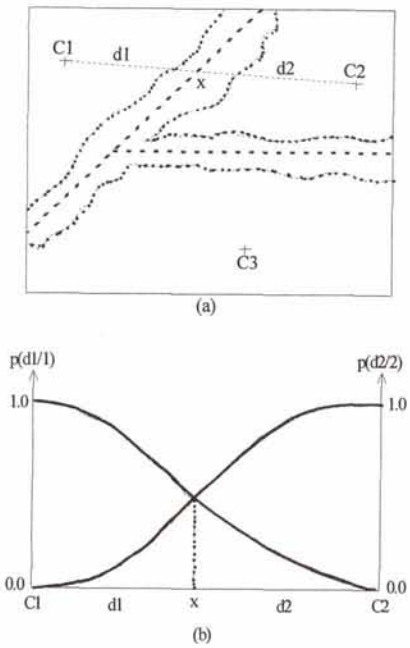

Figure 3. The process of interpolating on fuzzy maps.

The key to the derivation of fuzzy categorical maps is the process of fuzzy classification, which, in turn, relies on defining appropriate fuzzy membership functions (Klir and Yuan, 1995). For categorical mapping such as land-cover mapping, conventionally used methods include semi-automatic computerized classifications for digital images and visual interpretation and manual classification for graphical images. Various methods exist for deriving the fuzzy membership values for each type of classification. When using digital images, a typical method for fuzzy classification is the so-called fuzzy *c*-means clustering (Bezdek *et al.*, 1984), which seeks to optimize the partition of observations (pixels) among target classes by minimizing a certain distance measure adopted. Such a process assigns fuzzy membership values for each pixel belonging to all the candidate classes in an iterative way. It is illustrated in Figure 2, but a detailed discussion can be found in Bezdek *et al.* (1984).

For graphical images such as aerial photographs in graphical, not digital, form, the derivation of fuzzy classification takes place in a spatial rather than a spectral domain. Homogeneous areas are easier to identify than heterogeneous, transitional areas. Therefore, using for reference a set of classified attributes located in homogeneous areas, fuzzy mapping for graphical images at heterogeneous locations also uses interpolation methods in a spatial domain for deriving fuzzy membership values.

A graphical interpolation process is illustrated in Figure 3a. The centers C1, C2 and C3 of individual polygons, representing classes 1, 2 and 3, can be assigned with a membership value of 1.0 to their respective classes. At x (Figures 3a and 3b), the membership will reflect the component probability of each adjacent polygon. The fuzzy membership value will decrease when moving from the center towards and beyond the boundaries until it reaches 0.0 in the center of each adjacent polygon. Polygon boundaries are seen somewhere within the transitional zones indicated by dashed lines in Figure 3a. The changing pattern of class probabilities along a transect may be modelled by some function: for example, fuzzy membership functions for finding 1 or 2 along the transect of C1 to C2 are shown in Figure 3b, where p(d1/1) stands for the fuzzy membership value of class 1 at a distance of d1 away from the center C1, while p(d2/2) stands for the fuzzy membership value of class 2 at a distance of d2 away from the center C2. A theoretically more sound approach is via indicator kriging, which is described in Bierkens and Burrough (1993a and 1993b) and applied in Zhang and Kirby (1997), but not covered further here.

Fuzzy Boundaries

Suppose that fuzzy categorical maps are derived by using any of the methods for fuzzy classification described in the previous section. Then, denote a vector

$$\mathbf{P}(x) = (p_1(x), p_2(x), ..., p_c(x)) \quad (1)$$

where $p_i(x)$ ($i = 1, 2, \ldots, c$) are the fuzzy membership values of location x belonging to class i and, hence, comprise a set of fuzzy maps with a total of c classes.

To produce a crisp classification, an analogy with the classification of raster-based remotely sensed images using the maximum likelihood classifier is helpful: the maximum likelihood classifier assigns pixels to classes to which they have the maximum probability of belonging, measured by specific class membership functions. Similarly, for fuzzy categorical maps with readily available fuzzy membership values for individual grid cells, in order to generate a conventional maximum likelihood classification, vector $\mathbf{P}(x)$ is subjected to a maximization process, by which cell x is labeled as the class having the maximum values. For example, cell x is to be classified into class j on the condition as expressed in Equation 2: i.e.,

$$p_j(x) = \text{maximum } (p_1(x), p_2(x), ..., p_c(x)), \text{ for } j = 1, 2, ..., c \quad (2)$$

where class labels j form the classified data layer.

During a conventional crisp classification, information contained in a fuzzy vector $\mathbf{P}(x)$ is filtered out, leaving only the class labels having the maximum fuzzy membership values for individual locations. Boundaries in the resulting categorical map are defined where classes are separated, as shown, for example, in Figure 3. In order to acknowledge the spatial heterogeneity of class membership in the classified map, information contained in fuzzy membership values must be further explored.

First, the maximum fuzzy membership values of individual locations can be maintained to assist in defining fuzzy boundaries. Denote the maximum fuzzy membership values as p_{max}. Defining fuzzy boundaries can be done via a slicing process, by which p_{max} is examined with reference to a prescribed threshold τ (tau) (Zhang, 1996). Specifically, this process selects a fuzzy location x if the value of p_{max} is less than the value $\tau 1$. This is illustrated in Figure 4a where a two-class example is developed for the profile of C1 and C2 in Figure 3.

Second, there is the confusion index criterion for defining fuzzy boundaries, which involves two fuzzy membership values for each location, and is thus more complex than the previous criterion. An uncommon example of deriving fuzzy boundaries from a simulated set of fuzzy maps is given by Burrough (1996). He employed the concept of confusion index, evaluated as 1.0 minus the difference between the fuzzy membership values of location x belonging to the first most likely and the second most likely classes. The assumption underlying such an index is that the greater the confusion index, the smaller the difference in fuzzy membership values between the first most likely and the second most likely classes, the fuzzier is location x, and thus the more likely that location x defines a

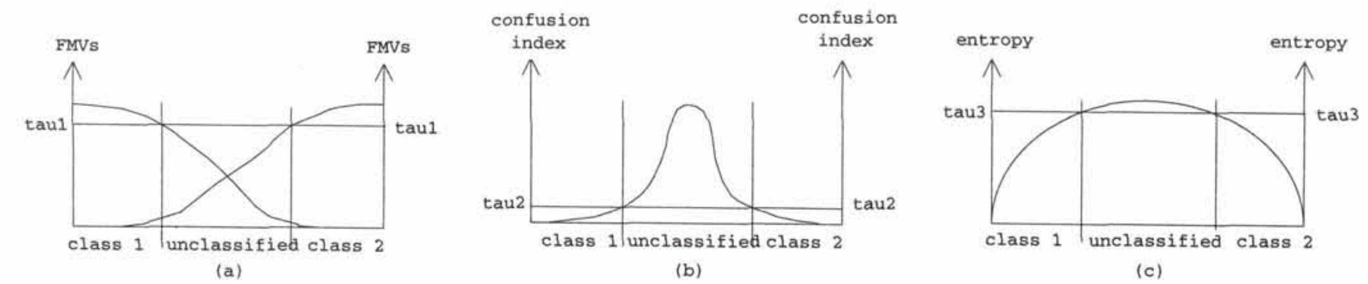

Figure 4. Three criteria applied in slicing for deriving fuzzy boundaries: (a) the maximum FMVs, (b) confusion index, and (c) entropy.

fuzzy boundary. Usually, a pre-defined threshold $\tau 2$ is applied such that location x defines a fuzzy boundary if the confusion index is greater than $\tau 2$, as shown in Figure 4b.

Finally, there is the measure of entropy for defining fuzzy boundaries, which is the most complex criterion, using the complete fuzzy membership values for each location. Foody (1995) described the use of entropy for evaluating the degree of fuzziness for fuzzy maps. Measures of entropy express the way in which the probability of class membership is partitioned between the classes. It is based on the assumption that in an accurate classification each location will have a high probability of membership in only one class. As entropy is a measure of disorder, large values indicate low accuracy in classification, while small values indicate high accuracy in classification. It is thus logical to assert that boundaries usually occur where locations have high degrees of fuzziness, that is, big values of entropy. Entropy $H(p(x))$ is measured using Equation 3: i.e.,

$$H(p(x)) = -\sum_{i=1}^{c} p_i(x) \log_2 p_i(x) \quad (3)$$

where $p_i(x)$ is the fuzzy membership value of grid cell x belonging to class i, where the index i ranges from 1 to c (the total number of classes). Again, a pre-defined threshold $\tau 3$ is applied such that location x defines a fuzzy boundary if its measure of entropy is greater than $\tau 3$, as shown in Figure 4c.

The distinctions among the three criteria described above for defining fuzzy boundaries are now related to the different sources of fuzziness mentioned in the earlier section on Fuzzy Categorical Maps. First, the maximum FMVs are straightforward to interpret and are closely associated with the conceptual gradation of class membership. Even so, the maximum FMVs may also be related to fuzziness due to mixed pixels, because FMVs can come from any possible source of fuzziness. Second, the confusion index involves FMVs belonging to the first and the second most likely classes. The interpretation of confusion index is thus not as straightforward as for maximum FMVs, suggesting a weaker relation to conceptually gradated fuzziness. Confusion index is, however, useful as a measure of fuzziness for mixed pixels. Third, there is the measure of entropy, which may appear as a measure of fuzziness on its own. However, entropy requires the complex logarithmic calculation of all elements of FMV's vectors. As the selection of the base for logarithmic calculation is largely arbitrary, the interpretation of entropy is thus not straightforward. Therefore, conceptually gradated class memberships are not apparently associated with values of entropy, but it is mathematically sound to explain, using entropy, fuzziness due to mixed pixels. The fuzziness related to certain classes not well represented by the training data is often interwoven with the other two sources of fuzziness. It is therefore not easy to establish simple relationships between it and the alternative criteria for defining fuzzy boundaries. This source of fuzziness may be looked at by increasing the categorical resolution (Franklin and Woodcock, 1997).

It is also worth stressing that fuzzy boundaries, as uncertain zones of non-zero widths that border adjacent areal units, may be considered as supersets of probabilistic epsilon error bands. Thus, these fuzzy boundaries should more truly be conceived of as transition zones of spatial variability, which is more complex than could be generalized by epsilon error band models. The distinctions between probabilistic and fuzzy boundaries are to be examined in the empirical study that follows.

An Empirical Test

The Test Site and the Data Sources

Land-cover data were used to compare the performance of different criteria in defining fuzzy boundaries from fuzzy categorical maps. The chosen test site is an area of about 2 square kilometers, located around Blackford Hill within the city of Edinburgh, Scotland. The area includes a wooded valley; residential, commercial, and academic buildings; road networks and footpaths; recreational areas; a small lake, agricultural fields and worked allotments; hills, and flat ground, as shown in Figure 5. The suburban residential districts include roads, pavements, buildings, walls, gardens, and hedges in dense spatial arrangements, creating the usual difficulties in mapping from aerial photography over selection of detail of different spatial dimensions. Further difficulties arise from the indistinct nature of boundaries between land-cover types. For example, on Blackford Hill, the dispersed individual trees or groups of trees merge with shrubs and rough grassland. The study area provides a good environment with significant fuzziness to test the alternative criteria for defining fuzzy boundaries.

Ground control points (GCPs) of three types were used: field surveyed points, densified points using photogrammetric block adjustment based on 1:5,000-scale aerial photographs, and points digitized from Ordnance Survey large-scale plans (Zhang, 1996). This set of GCPs proved sufficient for photogrammetric digitizing from 1:24,000-scale aerial photographs and remote sensing image rectification of SPOT HRV and Landsat TM data, all of similar dates. Effective pixel sizes for the three data sources were 4m, 10m, and 30m, respectively.

Deriving Fuzzy Maps of Land Cover

In order to provide a layer of reference attribute data, photogrammetric plotting was performed based on a reconstituted stereo photographic pair. A land-cover classification system was used with the following classes appropriate to the scene:

- grass (grassland and urban parks),
- built-up (built-up and barren land),
- wood (deciduous and coniferous woodland),

Figure 5. The test site – Blackford Hill, Edinburgh. Scanned aerial photograph enlarged to 1:15,000 scale.

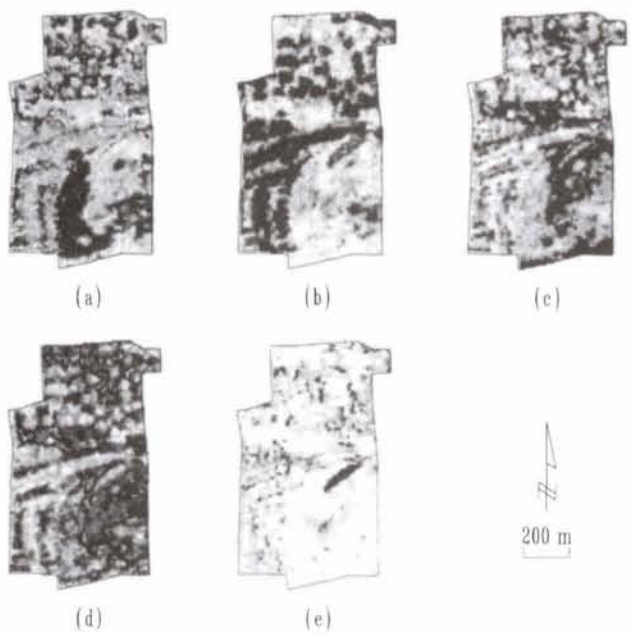

Figure 6. Fuzzy maps created from the SPOT HRV data: (a) grass, (b) built, (c) wood, (d) shrub, and (e) water. Darker grey levels indicate higher fuzzy membership values.

- shrub (shrub land, including open wooded land), and
- water (water bodies and water works).

For both SPOT HRV and Landsat TM data, a fuzzy *c*-means clustering algorithm based on Bezdek *et al.* (1984), programmed in FORTRAN 77 on a VAX/VMS, was used, with the parameter *m* set at 2.5, in order to produce fuzzy membership vectors across the five target classes pixel by pixel within the study area (Zhang, 1996). This resulted in 7133 and 779 grid cells for the fuzzy maps based on the SPOT HRV data and the Landsat TM data, respectively. As an example, fuzzy maps created from the SPOT HRV data are shown in Figure 6, where (a), (b), (c), (d), and (e) correspond to classes of grass, built-up, wood, shrub, and water, respectively. In Figure 6, the darker the grey scale, the higher the FMVs.

In order to generate fuzzy maps from 1:24,000-scale aerial photographs, indicator kriging was employed, because it has proved to be a sound approach for estimating the probabilities of all candidate land-cover types occurring at uncertain locations (Zhang and Kirby, 1997). Indicator kriging is supported by the geostatistical package GSLIB, which was used to generate fuzzy maps based on a set of representative and classified samples taken from aerial photographs, resulting in a total of 7133 grid cells of 10m size (Deutsch and Journel, 1992). The outputs from both fuzzy clustering and indicator kriging were transformed to ASCII format files using some original FORTRAN programs, and the outputs could then be loaded to ARC/INFO GRID data files in order to facilitate data management and analysis.

Deriving Fuzzy Boundaries

Following production of the three fuzzy maps, based on the 1:24,000-scale aerial photographs, the SPOT HRV data and the Landsat TM data respectively, it is possible to apply the three criteria described in the section on Fuzzy Boundaries to derive fuzzy boundaries. It is first necessary to apply the process of maximization to produce classified maps. Shown in Figure 7 are classified maps based on (a) photogrammetric data, (b) SPOT HRV data, and (c) Landsat TM data, and maps of the maximum FMVs (d), (e), and (f) corresponding to the classified maps shown in (a), (b), and (c), respectively.

Because fuzzy boundaries are to be derived following the process of maximization, the maximum FMVs, confusion

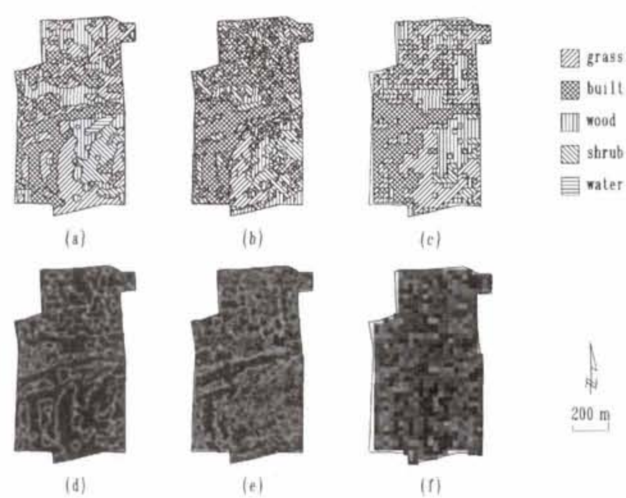

Figure 7. Classified maps based on (a) photogrammetric data, (b) SPOT HRV data, and (c) Landsat TM data; (d), (e), and (f) are maps of the maximum FMVs corresponding to the classified maps shown in (a), (b), and (c), respectively.

index, and entropy are output in the classifications, using the formulae described in the section on Fuzzy Boundaries. This resulted in three versions of each type of fuzzy map mentioned above, a total of nine maps. These maps are shown in Figure 8, where (a), (b), and (c) correspond, respectively, to the maximum FMVs, confusion index, and entropy of the classification based on the photogrammetric data; (d), (e), and (f) correspond to the classification based on the SPOT HRV data; and (g), (h), and (i) correspond to the classification based on the Landsat TM data. Note that the presence of fuzzy boundaries is indicated by the grey scale in Figure 8, where values of confusion index and entropy are illustrated inversely to provide comparable visualization between the fuzzy boundaries defined by the three criteria.

In order to check numerically if the three criteria produce similarly defined fuzzy boundaries, correlation coefficients between the maximum FMVs, confusion index, and entropy were calculated for the sets of fuzzy maps based on all three data sets (Table 1). The correlation coefficients are significant at the 5 percent level, indicating the similarity of fuzzy boundaries defined by these three criteria.

The effects of applying the three criteria can also be tested by comparing the fuzzy boundaries defined. For this comparison, it is necessary that the numbers of classified grid cells are the same when using different criteria for slicing. Each of the nine maps shown in Figure 8 was sliced by setting 29 threshold values, systematically increasing for maps of maximum fuzzy membership values, systematically decreasing for maps of confusion index and measure of entropy, so that 30 zones of equal areas (equal grid cells) were defined on each map. The zones were numbered serially from 1 to 30, the total of 30 being sufficient for statistical analysis. The process created a total of nine zoned maps, based on three criteria for each source of fuzzy map. Cell-by-cell comparison was then performed for each pair of three-zoned maps for each type of fuzzy map, creating a total of nine error matrices. As an example, Table 2 shows the 30 by 30 error matrix for fuzzy maps (d) and (e) of Figure 8, where zoned maps of maximum FMVs and confusion index correspond to the matrix's columns and rows, respectively. For clarity, diagonal elements are underlined.

Table 1. Correlation Coefficients for Pairs of Maps Showing Values of the Three Criteria for Defining Fuzzy Boundaries (FMVs = Fuzzy Membership Values)

Fuzzy Maps Based on	Maximum FMVs *versus* Confusion Index	Maximum FMVs *versus* Entropy	Confusion Index *versus* Entropy
1:24,000-Scale Aerial Photographs	−1.00	−0.93	+0.94
SPOT HRV Data	−1.00	−0.89	+0.92
Landsat TM Data	−1.00	−0.99	+0.99

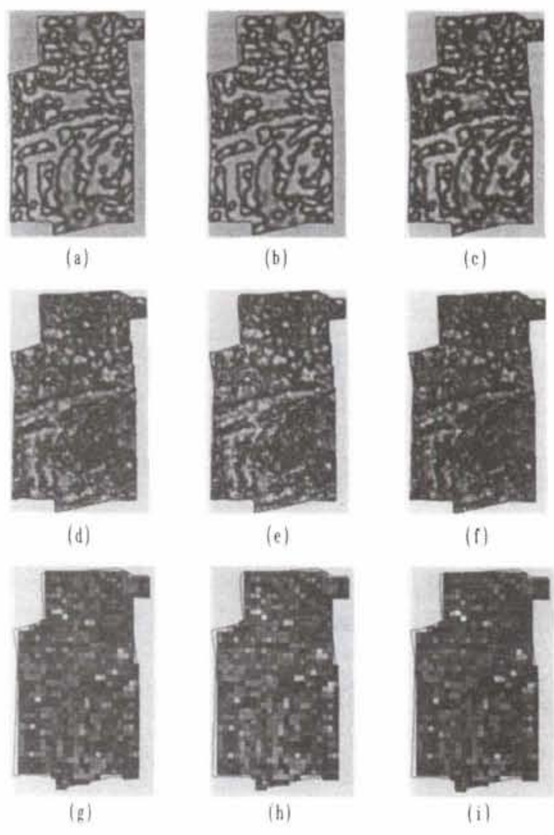

Figure 8. Maps of the maximum FMVs, confusion index, and entropy for the classifications based on (a), (b), and (c) photogrammetric data; (d), (e), and (f) SPOT HRV data; and (g), (h), and (i) Landsat TM data.

Because there were 30 zones for each map, by applying threshold values ($\tau 1$, $\tau 2$, and $\tau 3$) ranging from 1 to 30 for each of the nine zoned maps, 30 agreements were calculated for each pair of three-zoned maps for each type of fuzzy map by rearranging elements, in accordance with the value of τ applied, in their respective error matrices. The average agreements based on the nine error matrices are reported in Table 3.

It is shown in Table 3 that the comparisons between fuzzy boundaries defined by using the three different criteria in pairs are very good, all more than 90 percent, for fuzzy maps based on the three data sources. Moreover, confidence limits at the level of 95 percent are also listed in Table 3, using the statistical techniques described in Rosenfield and Melley (1980). This suggests that the three criteria result in fuzzy boundaries with similar positions. Varying the number of threshold values would have the effect of changing the confidence limits.

In addition to analysis of correlation coefficients and agreements based on error matrices, similarity between fuzzy boundaries defined by the three criteria can be examined by assessing classification accuracy against independent reference data. For this, specific threshold values of $\tau 1$, $\tau 2$, and $\tau 3$ are introduced to drive the modeling. Fifty percent of all the grid cells generated by fuzzy clustering and indicator kriging were coded into designed classes. As a result of this partial classification, the threshold values of $\tau 1$, $\tau 2$, and $\tau 3$ occur as listed in Table 4.

The slicing process creates a type of categorical map where classified locations belong to their named classes with, at least, the levels of certainty implied in the thresholds applied. Unclassified locations comprise the fuzzy boundaries that should be excluded from evaluating classification accuracy. Tests for classification accuracy are usually based on an error matrix, which is constructed by comparisons between the test data and the reference data. From the error matrix, it is possible to derive several useful classification accuracy parameters such as the overall classification accuracy and the Kappa coefficient of agreement (Congalton, 1991). In this test case, classified maps were checked against the reference map constructed employing photogrammetric digitizing. This process resulted in overall classification accuracies for fuzzy classified maps as listed in Table 5. As would be expected, the accuracy from 1:24,000-scale aerial photographs is much higher than from the satellite imagery. More importantly, Table 5 shows that the three criteria for defining fuzzy boundaries yield very similar

Table 2. Error Matrix from Cell-by-Cell Comparison of Zoned Maps from SPOT HRV Data. See Figures 8d and 8e.

15	68	19																											
56	97	74	96	43																									
5	24	21	27	45	20	7																							
2	14	49	64	83	70	79	24	2																					
		9	19	20	15	64	44	9																					
		2	17	25	19	56	40	33	22																				
		1	10	20	18	37	29	40	64	3																			
			11	10	5	23	20	31	70	32	4																		
				12	7	27	17	16	46	43	48																		
				5	8	26	12	5	48	30	72	9																	
					2	21	14	14	54	37	10	83	38	32	1														
						1	7	2	23	7	34	26	42	50	10														
								4	15	7	29	18	28	58	22														
									6	13	18	19	11	52	48	9													
									1	2	28	18	19	56	95	70	6												
											5	5	7	22	34	58	33	2											
													10	40	40	57	86	28	3										
														4	21	51	61	85	32	1									
															3	15	56	52	74	48									
																1	10	39	54	10	21								
																		4	35	94	71	21							
																			1	64	61	92	6						
																				1	31	12	11	6					
																						11	90	72					
																							38	12	85				
																								13	14	70			
																									4	20	31		
																											19	83	
																												15	53
																													22

Table 3. Average Agreements between Fuzzy Boundaries Defined by (I) Pairs of Criteria and (II) Their Confidence Limits at the Significance Level of 5 percent (FMVs = Fuzzy Membership Values) (Unit: Percentage)

	Maximum FMVs *versus* Confusion Index		Maximum FMVs *versus* Entropy		Confusion Index *versus* Entropy	
Fuzzy Maps Based on	I	II	I	II	I	II
1:24,000-Scale Aerial Photographs	99.6	98.1 ~ 98.7	92.1	91.3 ~ 94.3	92.1	90.7 ~ 95.0
SPOT HRV Data	96.5	93.8 ~ 98.5	91.8	89.8 ~ 95.6	93.9	91.1 ~ 97.3
Landsat TM Data	95.2	91.5 ~ 98.0	90.4	87.5 ~ 95.0	91.5	89.2 ~ 97.3

Table 4. Threshold Values of τs for Deriving Fuzzy Boundaries (FMVs = Fuzzy Membership Values)

	Threshold Values τ		
Fuzzy Maps Based on	Maximum FMVs ($\tau1$)	Confusion Index ($\tau2$)	Measure of Entropy ($\tau3$)
1:24,000-Scale Aerial Photographs	0.78	0.22	0.80
SPOT HRV Data	0.51	0.57	1.67
Landsat TM Data	0.40	0.67	1.67

overall classification accuracies, except for the Landsat TM data, where some variation is observed. The results using Landsat TM data slightly favor the maximum FMV method, which is the criterion preferred, given its apparent simplicity.

Results so far have shown that fuzzy boundaries can be derived by using the three criteria with only insignificant differences for fuzzy maps derived from aerial photogrammetric data and satellite digital images. Also, the results suggest that these three criteria fail to discriminate between various sources of fuzziness existing in the land-cover data derived from graphical and digital images, which may be due to conceptual gradation of cover types, mixed pixels, or lack of training data for certain cover types. A possible explanation for this is that mixed pixels constitute the dominant source of fuzziness for the study area, because mixed pixels are abundant in the SPOT HRV and Landsat TM data, and finite sampling size for photogrammetric data also leads, unfortunately, to mixtures.

As indicated in the section on Fuzzy Boundaries, fuzzy boundaries are meant to stand as a superset of all types of uncertain zones of non-zero widths in localities of heterogeneous cover types. Figure 8 reveals that fuzzy boundaries for photogrammetric land-cover data bear striking resemblance to epsilon error band models (thus termed probabilistic boundaries) while, for land-cover data based on satellite

Table 5. Accuracy Assessment for Classified Maps Excluding Fuzzy Boundaries (FMVs = Fuzzy Membership Values)

	Overall Classification Accuracies (Unit: Percentage)		
Fuzzy Maps Based on	Maximum FMVs	Confusion Index	Measure of Entropy
1:24,000-Scale Aerial Photographs	83.7	83.7	83.6
SPOT HRV Data	46.7	47.9	47.8
Landsat TM Data	36.6	34.6	31.1

images, fuzzy boundaries are hard to generalize by probabilistic epsilon error band modeling, because they are irregularly distributed. The epsilon error bands appearance of fuzzy boundaries based on the photogrammetric data is attributed to subjectivity and, hence, discreteness imposed in the identification and location of "core" and "pure" samples prior to the generation of fuzzy maps. Such subjectivity is reduced relatively in the derivation of fuzzy maps based on satellite images. Moreover, production of fuzzy maps from satellite images takes place in spectral rather than geographical space (Burrough, 1996). Thus, it is not surprising that irregular distribution of uncertain zones in geographical space may come out of spectral classes that are outstandingly ellipsoidal. The relationships between fuzzy and probabilistic boundaries represent a topic for future research.

Conclusion

Conventionally used methods for categorical mapping such as land-cover mapping include semi-automatic computerized classifications of digital satellite images and visual interpretation of graphical aerial photographs. Usually, discrete area objects, i.e., polygons, are employed in categorical maps to represent the two-dimensional distributions under study. To extend object-based data models into the domain of fuzzy categorical maps, fuzzy boundaries need to be defined properly.

Fuzzy boundaries can be derived quantitatively by using different criteria on fuzzy categorical maps. The whole process permits a theoretically sound and data-driven solution to estimating errors in attributes and boundaries of categorical maps, which is a substantial benefit. Interpretation is consistent between semi-automatic computerized classification of digital images and visual interpretation of graphical images. Although the three techniques assessed achieve similar performances, the direct extension of the maximum likelihood classification provides the easiest and most straightforward solution. Overall, the results obtained suggest that maintaining both a crisp classification and its underlying uncertainty map for deriving fuzzy boundaries at different thresholds offers a logical, flexible, and compact solution for managing categorical maps and information on the quality of categorical boundaries.

This research paper considers some of the deeper issues underlying the heterogeneous spatial data depicting the real world. The methods followed in the empirical study are not of universal application, but should be applicable to mapping land cover with remotely sensed data of compatible spatial resolutions for areas of similar categories of land cover. The methods should also be of value for further examining the effects of fuzziness for mixtures of scales and resolutions. The study has highlighted the distinctions between fuzzy and probabilistic boundaries. Probabilistic boundaries are relatively familiar and have been widely discussed by using models such as epsilon error bands. Fuzzy boundaries, on the other hand, have previously been conceived of as an extension of probabilistic boundaries and, for this reason, have been little investigated.

Acknowledgments

The constructive comments of two anonymous referees are gratefully acknowledged in the preparation of this paper.

References

Altman, D., 1994. Fuzzy set theoretic approaches for handling imprecision in spatial analysis, *International Journal of Geographical Information Systems*, 8(3):271–289.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, U.S. Geological Survey Professional Paper 964. Washington, D.C., 28 p.

Berry, J.K., 1993. *Beyond Mapping: Concepts, Algorithms, and Issues in GIS*, GIS World Inc., Fort Collins, Colorado, 246 p.

Bezdek, J.C., R. Ehrlich, and W. Full, 1984. FCM: The fuzzy c-means clustering algorithm, *Computers and Geosciences*, 10(2 & 3):191–203.

Bierkens, M.F.P., and P.A. Burrough, 1993a. The indicator approach to categorical soil data. I: Theory, *Journal of Soil Science*, 44:361–368.

———, 1993b. The indicator approach to categorical soil data. II: Applications to mapping and land use suitability analysis, *Journal of Soil Science*, 44:369–381.

Burrough, P.A., 1996. Natural objects with indeterminate boundaries, *Geographic Objects with Indeterminate Boundaries* (P.A. Burrough, and A.U. Frank editors), Taylor & Francis, London, pp. 3–28.

Campbell, J.B., 1987. *Introduction to Remote Sensing*, The Guilford Press, New York, 551 p.

Chrisman, N.R., 1982. A theory of cartographic error and its measurement in digital databases, *Proceedings AUTOCARTO'5*, pp. 159–168.

Congalton, R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data, *Remote Sensing of Environment*, 37(1):35–46.

Deutsch, C., and A.G. Journel, 1992. *GSLIB: Geostatistical Software Library*, Oxford University Press, New York and Oxford, 340 p.

Edwards, G., and K.E. Lowell, 1996. Modeling uncertainty in photointerpreted boundaries, *Photogrammetric Engineering & Remote Sensing*, 62(4):373–391.

Fisher, P.F., and S. Pathirana, 1989. Urban boundary detection from Landsat imagery: A GIS and knowledge-based approach to MSS data, *Technical Papers 1989 ASPRS/ACSM Annual Convention*, 4:93–101.

Foody, G.M., 1995. Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(5):2–12.

Franklin, J., and C.E. Woodcock, 1997. Multiscale vegetation data for the mountains of Southern California: Spatial and categorical resolution, *Scale in Remote Sensing and GIS* (D.A. Quattrochi and M.F. Goodchild, editors), CRC/Lewis Publishers, Inc., Boca Raton, Florida, pp. 141–168.

Goodchild, M.F., 1989. Modelling error in objects and fields, *Accuracy of Spatial Databases* (M.F. Goodchild and S. Gopal, editors), Taylor & Francis, London, pp. 107–113.

Goodchild, M.F., and G.J. Hunter, 1997. A simple positional accuracy measure for linear features, *International Journal of Geographical Information Science*, 11(3):299–306.

Guptill, S.C., and J.L. Morrison, 1995. *Elements of Spatial Data Quality*, Elsevier Scientific/International Cartographic Association, Oxford, 202 p.

Kaufmann, A., 1975. *Introduction to the Theory of Fuzzy Subsets*, Academic Press, New York, 416 p.

Kiiveri, H.T., 1997. Assessing, representing and transmitting positional uncertainty in maps, *International Journal of Geographical Information Systems*, 11(1):33–52.

Klir, G.J., and B. Yuan, 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall PTR, New York, 574 p.

Lagacherie, P., P. Andrieux, and R. Bouzigues, 1996. Fuzziness and uncertainty of soil boundaries: From reality to coding in GIS, *Geographic Objects with Indeterminate Boundaries* (P.A. Burrough, and A.U. Frank, editors), Taylor & Francis, London, pp. 275–286.

Lowell, K.E., 1994. An uncertainty-based spatial representation for natural resources phenomena, *Proceedings of the 6th International Symposium on Spatial Data Handling*, Edinburgh, pp. 933–944.

Mark, D.M., and F. Csillag, 1989. The nature of boundaries on area-class maps, *Cartographica*, 21(1):65–78.

NCDCDS (National Committee for Digital Cartographic Data Standards), 1988. The proposed standard for digital cartographic data, *The American Cartographer*, 15(1):9–140.

Perkal, J., 1956. On epsilon length, *Bulletin de l'Academie Polonaise des Sciences*, 4:399–403.

Rosenfield, G.H., and M.L. Melley, 1980. Applications of statistics to thematic mapping, *Photogrammetric Engineering & Remote Sensing*, 46(10):1287–1294.

Wang, F., and G.B. Hall, 1996. Fuzzy representation of geographical boundaries in GIS, *International Journal of Geographical Information Systems*, 10(8):573–590.

Zhang, J.X., 1996. *A Surface-Based Approach to Handling Uncertainties in an Urban-Orientated Spatial Database*, unpublished Ph.D. thesis, The University of Edinburgh, 228 p.

Zhang, J.X., and R.P. Kirby, 1997. An evaluation of three fuzzy approaches to mapping land cover from aerial photographs, *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(5):193–201.

(Received 12 January 1998; revised and accepted 22 October 1998; revised 10 December 1998)