

# Using a Fuzzy Surface-based Cartographic Representation to Decrease Digitizing Efforts for Natural Phenomena

**Kim Lowell and Christopher Gold**

**ABSTRACT:** On a previously-digitized map, a series of 685 points were screen-digitized so as to represent major spatial features of each polygon composing 17 forest types having inherently uncertain boundaries; three forest types having precise boundaries were treated as such. From these points and three exact types, surfaces were generated which provided weights of the certainty of having each of the 20 forest types at a location. These surfaces were "reconstituted" into a thematic map by assigning each location to the type having the highest weight. When compared to the precisely digitized and rasterized map, areal and locational inaccuracy were 5% and 18%, respectively, for the reconstituted thematic map. For certain situations-such as the preparation of long-term forest management plans-these results show promise for adequately representing maps of natural features while reducing digitizing efforts.

**KEY WORDS:** interpolation, spatial certainty, geographic information systems

## Introduction

The digitizing of thematic maps in order to convert them to a computer-compatible format appears to be a straightforward process. One places an analog map on a digitizing tablet and, with a puck, traces those lines which comprise the polygons of interest. That this is a time-consuming, operator-intensive, and tedious task is well accepted. Nonetheless, to obtain what has been considered the necessary precision for digital spatial data bases, the arduous tracing of every pertinent line on a map and the subsequent editing of the inevitable errors has remained the most common way of inputting spatial data into a computer.

This process of data input seems reasonable, given that the lines on the map have been identified by a trained cartographer, photo-interpreter, and so on. Thus it is reasonable to assume that the thematic map lines are delimiters of real-world polygons. For thematic maps of certain spatial phenomena, however, a closer examination suggests that this data input process may not be as widely applicable as has generally been assumed.

Suppose that one is interested in producing thematic maps using census data collected for a given

urban area. Census data are often collected by census district and the boundaries of these districts usually coincide with natural or man-made features-e.g., city streets, township boundaries, or edges of rivers. In such cases, the boundaries being mapped are real and can be readily identified on the ground with a minimum of positional error. While the edge of a river may be difficult to find, if a river or similar feature is used as a legal demarcation it can be reasonably assumed that its boundaries are readily identifiable. Thus, in digitizing a map of a city one can digitize the census district boundaries confident that these are features having a high level of certainty of existing and being located in the right place.

The previous example fits well with accepted choropleth mapping (Dent 1985) and existing geographic information system (GIS) techniques: readily identifiable polygons exist and each polygon has associated attribute data. But consider an alternative situation in which one has a forest-type map produced by a trained photo-interpreter from a set of aerial photographs. The resulting thematic forest map would look very similar to the city map in that dimensionless boundary lines would delimit polygons which have a number of attributes (e.g., forest density, height). Thus, in either the urban or forest setting, one apparently has the same situation in relation to mapping. However, it must be recognized that the two are fundamentally very different in relation to real-world conditions (Mark and Csillag 1989).

In contrast to census boundaries, the boundaries between forest-type polygons may or may not be readily identifiable. Certainly the boundary

---

Kim Lowell is Professeur Titulaire and Christopher Gold is Professeur Titulaire at the Centre de Recherche en Géomatique, Pavillon Casault, Université Laval, Ste-Foy, Québec G1 K 7P4, Canada.

---

	New England 1,166-point random sample	New England 263-point systematic sample	Nebraska 1,402 -point random sample	Nebraska 307-point systematic sample
<b>Lambert  k - 1 </b>				
mi ni mum	0. 012420714839	0. 012363590439	0. 000000105656	0. 000000026782
maxi mum	0. 019518537294	0. 019145526405	0. 000865845106	0. 000853359897
medi an	0. 015126128433	0. 015210325420	0. 000139462271	0. 000134937983
mean	0. 015333822728	0. 015377209726	0. 000219354466	0. 000221059109
std. dev.	0. 001808530342	0. 001804960020	0. 000214439383	0. 000220340705
<b>Lambert ω</b>				
mi ni mum	0. 024687794223	0. 024574952441	0. 000000211312	0. 000000053565
maxi mum	0. 038658579692	0. 037926840647	0. 001730940741	0. 001705991778
medi an	0. 030024610644	0. 030190469750	0. 000278905093	0. 000269857759
mean	0. 030430503592	0. 030515977571	0. 000438614882	0. 000442020977
std. dev.	0. 003561334705	0. 003554254596	0. 000428728588	0. 000440526833
<b>Lam bert z</b>				
mi ni mum	0. 000215491734	0. 000213526268	0. 000000000000	0. 000000000000
<b>maximum</b>	0. 000528413434	0. 000508597687	0. 000001059301	0. 000001028985
medi an	0. 000318732313	0. 000322263593	0. 000000027502	0. 000000025747
mean	0. 000331889942	0. 000333700911	0. 000000132957	0. 000000137467
std. dev.	0. 000077700815	0. 000077617362	0. 000000222741	0. 000000229202
<b>Albers  k - 1 </b>				
mi ni mum	0. 000011551382	0. 000030821077	0. 005569818468	0. 005580511348
maxi mum	0. 008123329343	0. 008021655312	0. 009071795671	0. 009068814790
medi an	0. 003449399555	0. 003528361366	0. 007746334766	0. 007674313855
mean	0. 003696526486	0. 003710334579	0. 007608117079	0. 007572308990
std. dev.	0. 002267272702	0. 002265338671	0. 000974445404	0. 000975846273
<b>Albers ω</b>				
mi ni mum	0. 000023102630	0. 000061641205	0. 011170717410	0. 011192222732
maxi mum	0. 016312825843	0. 016107829630	0. 018226137674	0. 018220121592
medi an	0. 006910711151	0. 007069186709	0. 015552830172	0. 015407673461
mean	0. 007408276369	0. 007435569885	0. 015275220381	0. 015203058852
std. dev.	0. 004552141643	0. 004548018939	0. 001963465081	0. 001966271152
<b>Albers z</b>				
mi ni mum	0. 000000000189	0. 000000001343	0. 000044118364	0. 000044288397
maxi mum	0. 000094084530	0. 000091734724	0. 000117449267	<b>0.000117371744</b>
medi an	0. 000016885017	0. 000017668302	0. 000085522078	0. 000083933133
mean	0. 000026724130	0. 000026832583	0. 000083858371	0. 000083081183
std. dev.	0. 000026651695	0. 000026700323	0. 000020805588	0. 000020812505

between a lake and a mature forest is readily locatable, probably with the same precision as the river in the urban setting. But the boundary between a 40%/60% spruce/birch mix and a 60%/40% spruce/birch mix is less identifiable. In fact, it may not even exist in the real world. This type of boundary can be considered, therefore, to be extremely "fuzzy." When one digitizes the lines of a forest map, one is sometimes digitizing the features of high spatial certainty. But, more often, one may be digitizing those features which have inherently low associated spatial certainty. More accurately, perhaps, one is digitizing lines whose fuzzinesses are variable, yet which are treated as having uniform spatial certainty.

This suggests that, for digitizing and spatial data representation, the traditional GIS model may be severely lacking for natural phenomena—that is, for those phenomena whose boundaries are naturally fuzzy. (Soils and ecological zones are further examples of this.) The painstaking digitizing of the lines on a forest (or soils or. . .) map may be rather a waste of time. That is, given that the spatial precision of the real-world boundaries is relatively low, it may be preferable to utilize alternative digitizing techniques and spatial data representations that reflect the fuzziness of the factors being mapped and which do not insist on the exact preservation of highly imprecise boundaries.

This is not to suggest that the low spatial precision of real-world boundaries makes the conversion of forest maps to a digital format a nonsensical exercise. Instead, it is to say that recognition that this imprecision exists is critical to understanding the central premise of this paper. Specifically, if a "suitable approximation" of the real-world forest can be obtained by not digitizing every line on a forest-type map, then the painstaking task of converting large map-based (forestry) databases to a digital format may be satisfactorily reduced. The purpose of this paper is to evaluate an alternative way of producing maps of natural features and to evaluate the accuracy of that alternative. A secondary objective is to demonstrate one method for producing a cartographic representation which may be more suitable for digital computers than the traditional fixed-line model inherent in paper maps. The test-domain for this work is the production of forest maps.

## Study Site

A portion of Montmorency Forest—the research forest of Laval University—was selected for use. The forest is located approximately 80 km north of Quebec City, Canada. The area used for this study is approximately 1500 hectares in size, is irregularly-shaped and approximately 6.5 km (east-west) by 3.5 km (north-south). The predominant vegetation type is boreal forest which is the dominant forest association of the region (Society of American Foresters 1979). A forest type map had been produced for the area through the interpretation of 1984 aerial photographs by trained photointerpreters in 1985. On the portion of Montmorency Forest used in this study, this resulted in 242 different forest stand polygons representing 20 different forest and non-forest categories. For the purposes of this study, all land-cover types were arbitrarily assigned numbers 1 to 20 and are referred to throughout as FT1 (forest type 1), FT2, etc. Among the 20 categories were water (lakes and rivers: FT20), disturbances (complete blow-downs and clearcuts: FT15), and unproductive land (bare rock and other anomalies: FT16). It was felt that these three types might have fairly precise boundaries compared to the other types; this was verified by visual examination of the aerial photographs. Thus, in subsequent sections, these three types are treated slightly differently than the others.

## Experiment

The raster<sup>1</sup> map of the study area was displayed on a color computer monitor to screen-digitize points within polygons that would be spatially representative of each polygon for those types having uncertain boundaries—all types except FTs 15, 16, and 20 which will be discussed shortly. (The raster map will subsequently be referred to as the "digitized map" to reflect how it was created.) Originally, it was intended to place one point per polygon only. However, most polygons were irregularly-shaped, making it difficult and inappropriate to use a single point to represent a polygon. An alternative point-digitizing strategy comprised of several rules was adopted (Figure 1).

1. At least one point per polygon was digitized.

---

<sup>1</sup>A paper map of the area was digitized by a Master's level student as part of his thesis project. This entailed registering the map to the Universal Transverse Mercator (UTM) projection, tracing the boundaries of all forest stand polygons with a digitizing puck, and assigning the appropriate forest type value (1 to 20) to each completed stand. This digitized map was then converted into a raster thematic map composed of square cells that were 13 m on a side or 0.169 ha in area.

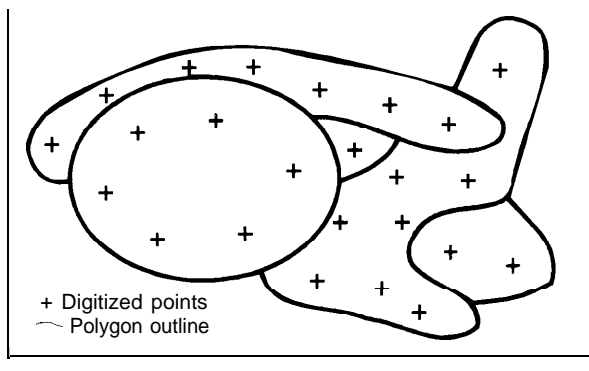


Figure 1. Location of digitized points in various polygons.

2. If a polygon was a regular geometric shape, such as circular or rectangular with a well-defined centroid, and was relatively small, a single point was placed at the center.
3. If a polygon was regularly-shaped but relatively large, points were distributed near its perimeter but far enough inside to make reasonably certain that on the ground the point would be the type digitized.
4. If the polygon was an irregularly-shaped blob, a point was placed within each major feature of the polygon. If a major feature was relatively large, the preceding rule applied.
5. If the polygon was linear in form, points were placed at regular intervals along its length.

It is evident that the placing of points within a polygon was an extremely subjective process. However, the process closely resembles what would occur if one were to tell a digitizing operator to not digitize lines, but to digitize instead one or more points within each polygon which have high certainty of actually being the type attributed to the polygon. That is, the operator would place points in the “fattest” part(s) of each polygon. Furthermore, the operator was made aware of how the points were to be treated subsequently (see below). Because one goal of this research was to allow a digitizing operator to produce an acceptable map representation by placing such points, this subjectivity is actually considered beneficial in a study which seeks to produce an operational methodology that will ultimately include the same subjectivity.

In all, 685 points were placed in 210 polygons of the 17 forest types having uncertain boundaries. This provided a per-polygon average of 3.26 points for the forest type polygons having uncertain boundaries. The 32 polygons of the three types having precise boundaries—FTs 15, 16, and 20—were treated differently. Since high certainty could be inferred for these types, their boundaries

were digitized and the resulting polygons rasterized. Their treatment is described subsequently. For the moment it is the 685 points digitized for the types having uncertain boundaries which are of interest.

These points can be considered to be places having a near-certainty of actually being a given type; this also means that they are certain to not be another type. Viewed a different way, one can locate all digitized points on the area and use as attribute data their certainty values (100 or 0) of being a given type. The result is a set of spatially distributed points showing the certainty of having, say, FT 1 at each location. One can then interpolate among these points to produce an isarithmic map of certainty for that type. This certainty weight is akin to the “fuzzy set membership function” described by Zadeh (1965). The same procedure can be followed for each forest type to produce a set of surfaces containing values between 0 and 100 (or 0 and 1.0) showing the certainty of having a given type at a given location.

The actual interpolation technique employed is important, however. One would like to use a technique which preserves the information at the original points and ensures that the values at any given location sum to 100 over all types. In this study, a variant of natural-neighbor interpolation known as Voronoi diagram area-stealing (Gold 1989) was employed. In area-stealing, one creates a Voronoi diagram (composed of Thiessen polygons) around the data points (Figures 2a and 2b). For any point at which one would like to have an interpolated value, one locates that point on the Voronoi diagram (Figure 2c) and constructs the Thiessen polygon around that query point (Figure 2d). This Thiessen polygon is then overlaid on the original Voronoi diagram and the percent of area that the point would “steal” from the original Thiessen polygons of each type is determined (Figure 2e). These percentages become the certainty scores for each type for this point. In the example presented (Figure 2), the query point would have certainties of 20, 15, and 65 for FTs A, B, and C, respectively (Figure 2e). Note that these values sum to 100 as desired. Further note that, if a query point is placed directly on a data point, its Thiessen polygon will match the Thiessen polygon of the data point exactly. This means that the query point’s Thiessen polygon would steal 100% of its area from the data point, thereby preserving the data point as desired.

While theoretically this procedure can be conducted for every location on the map, computational constraints make it more efficient to sample the map surface systematically with query points

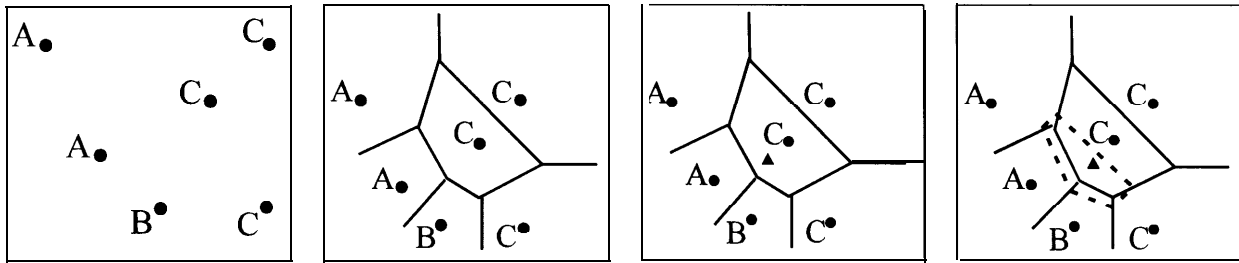


Figure 2. Example of area-stealing techniques. Above, left: Digitized points and forest types (A, B, etc.). Above, second from left: Voronoi diagram (i.e., Theissen polygons). Above, second from right: Single sample grid-point inserted (triangle). Above, right: Creation of new Voronoi diagram as if grid point were a data point (dotted lines show boundaries of grid-point Theissen polygon). Left: Sample point Theissen polygon showing percentage of area stolen from original polygons of a given type.

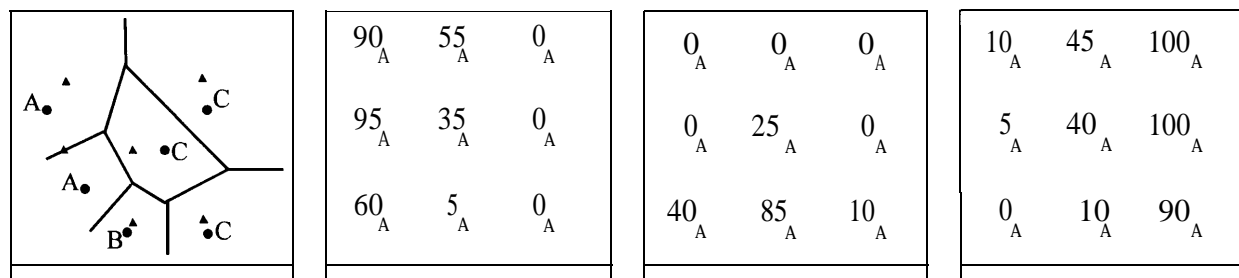
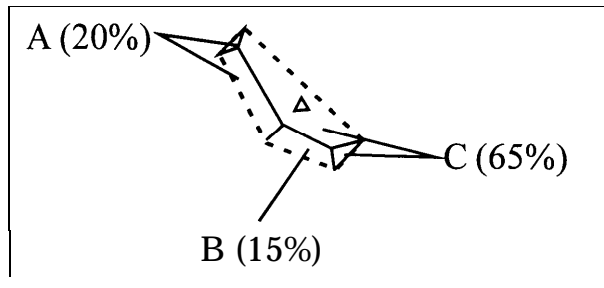


Figure 3. Grid of sample points imposed on Theissen polygon boundaries and resulting weights. (Example is hypothetical.) Above, left: Voronoi diagram and sample grid. Above, second from left: Certainties for forest type A. Above, second from right: Certainties for forest type B. Above, right: Certainties for forest type C.

in order to estimate certainty scores for each type at regular intervals (Figure 3). Once this has been done, a less computationally demanding interpolation method can be employed to estimate certainties at intermediate locations such as the 13 m raster cells used in this study. The interpolation method selected for this purpose is relatively unimportant since data are now regularly spaced at a known distance—an optimal situation for interpolation.

Thus to produce certainty surfaces by the method described, a user need only specify a grid of sample points. Because it was felt in this study that the spacing of the grid might affect results, two grid sizes were used: 50-by-50 m and 100-by-100 m. Certainty values for each grid point were estimated using the 685 points for each of these spacings. Linear interpolation was then conducted on the resulting grid to provide raster surfaces with 13 m cells showing the certainty of having a given FT in a given cell. Note that, with 17 forest types in the original data, 17 raster surfaces resulted. Furthermore, instead of having a deterministic map surface (Figure 4a),

one now has a fuzzy surface representation for each type (Figure 4b).

The three FTs with fixed boundaries remained to be incorporated into this set of fuzzy surfaces. To do this, any cell having been digitized as FT 15, 16, or 20 was assigned a value of 100 on fuzzy surfaces 15, 16, and 20, respectively; the remaining cells on these surfaces were then assigned values of 0 (zero). These three surfaces were overlaid on each of the 17 fuzzy surfaces produced as described earlier. At each location on fuzzy surfaces 15, 16, or 20 at which the value of 0 was present (absence of one of these types), values on the remaining fuzzy surfaces were not altered. In cases where a value of 100 was present on fuzzy surfaces 15, 16, or 20, the value on all of the remaining 17 surfaces was assigned a value of 0. This procedure ensured that the certainty values for all cells would sum to 100 over all types, and that the fixed boundaries of FTs 15, 16, and 20 would be preserved. This was done for the surfaces produced using both the 100-by-100 m and 50-by-50 m grid-point spacing.

Each set of fuzzy surfaces was then used to produce a deterministic map by assigning each cell to a

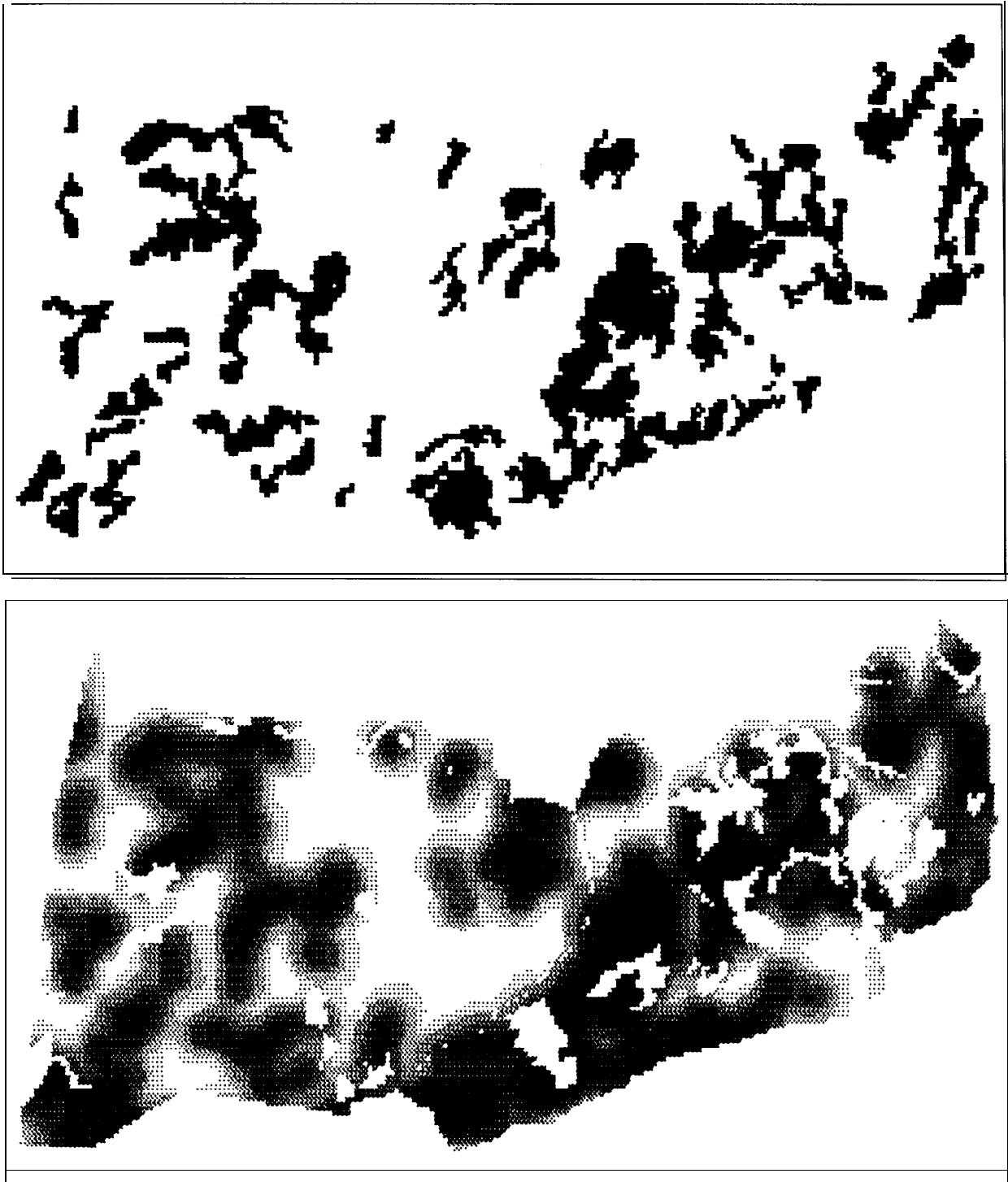


Figure 4. Forest type 2. Top: Digitized (raster) map. Above: Fuzzy surface.

forest type using a maximum-membership decision rule. For a given cell, the weights on all surfaces were examined and the cell was assigned to that forest type for which its certainty value was the highest (as is done in the maximum-likelihood classifier in remote sensing). Two thematic maps resulted from the procedures described:

- a maximum-membership map based on a 100-by- 100 m sample grid
- a maximum-membership map based on a 50-by-50 m sample grid.

Moreover, the original digitized map was present in the digital database.

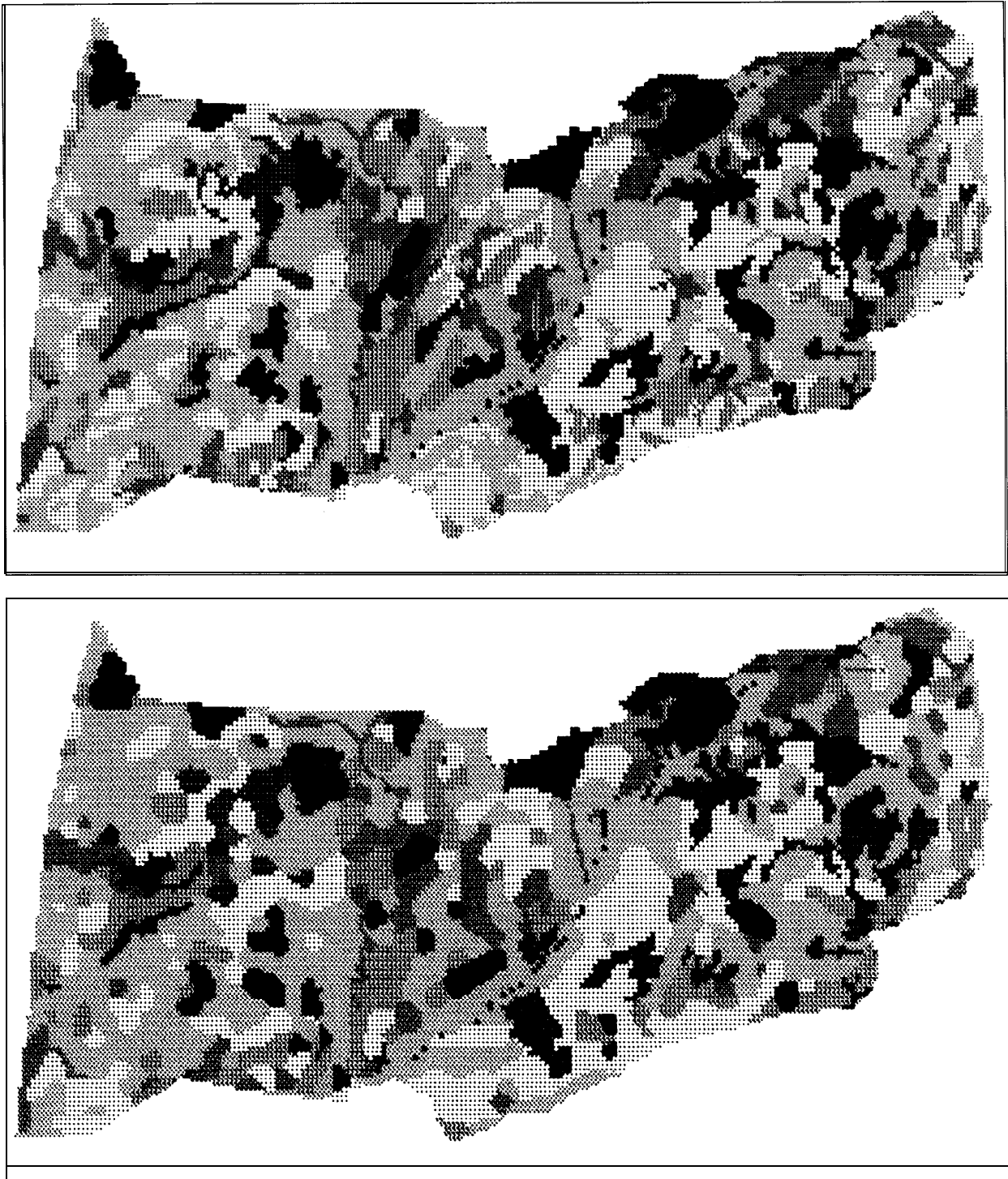


Figure 5. Maps of the study area. Top: Digitized map. Above: Maximum-membership map (50-by-50 m grid).

The maximum-membership maps were evaluated in two ways. First, for a given map the area of each of the twenty forest types was determined and compared to the amount on the initial digitized map. Because the only concern in this evaluation is that the amount-not the location-of a given forest type on both maps be the same, this will be referred to as an assessment of "area" accuracy." Second, the

maximum-membership map being considered was overlaid with the digitized map and the areal coincidence of each type determined. This is referred to subsequently as an assessment of "locational accuracy" because the concern is that forest types are in the same location on both maps. This latter is comparable to the diagonal of error matrices that are used to evaluate the classification of satellite images.

Map class (1)	Digitized map (cells) (2)	Max-membership polygons (cells) (3)	Difference (%) (4)	Locational agreement (cells) (5)	Difference (%) (6)
1	21,012	20,693	1.5	16,381	22
2	29,945	31,224	4.3	25,460	15
3	11,340	10,168	10.3	8,288	26.9
4	1,721	1,727	0.3	1,265	26.5
5	1,460	1,427	1.5	903	38.2
6	1,412	1,838	30.2	1,216	13.9
7	215	169	21.4	119	44.6
8	261	285	9.2	220	15.7
9	5,007	4,788	4.4	3,520	29.7
10	2,832	2,556	9.7	2,077	26.7
11	380	297	21.8	234	38.4
12	272	320	17.6	226	16.9
13	1,214	1,478	21.7	1,032	15
14	414	571	37.9	354	14.5
15	1,951	1,951	0	1,951	0
16	9,308	9,308	0	9,308	0
17	146	69	52.7	67	54.1
18	211	292	38.4	208	1.4
19	270	210	22.2	204	24.4
20	1,542	1,542	0	1,542	0
Totals:		Number of cells ( $\sum(\text{column (2)})$ ) = 90913			
		Areal differences ( $\sum (2)-(3) $ ) = 4558			
		Locational differences ( $\sum (2)-(5) $ ) = 16338			
Total areal inaccuracy:		5%			
Total locational inaccuracy:		18%			

Table 1. Area and location evaluation of maximum-membership polygons for 50-by-50 m grid.

## Experimental Results

Figure 5 presents the two types of maps produced—digitized and maximum-membership. It is evident (Table 1) that if one is asking the question “How much of each type is present?” then the maximum-membership map performs fairly well with a total areal inaccuracy of 5%. Though the average for any given type is only 5%, for individual types, this can vary radically (Column 4, Table 1). Of particular interest is that estimates of the less common types (e.g., FTS6, 14) are particularly poor. Of further interest is that these types were neither consistently overestimated nor consistently underestimated. Visual inspection of the digitized map suggested that, if less common types were present in “tentacled” polygons instead of those with more regular geometric shapes, their areas tended to be

underestimated. Thus, in an operational context, it may be useful to represent such types by using more points, though it is difficult when looking at a forest map to determine the general shape characteristics of polygons of a given type. Locational inaccuracy for this maximum-membership map is 18%. Therefore, if one is concerned not with the total amount of each type but with the location of each, the accuracy will be less than the areal accuracy. Again, the exact amount of locational inaccuracy varies with type, but it appears to be more consistent for each type than areal accuracy.

If the spacing of the points on the sample grid is doubled (Table 2), the areal inaccuracy increases to 9% and the locational inaccuracy increases to 22%. The same tendencies as were noted for the nature of the inaccuracy for individual types occur. Thus, the spacing of the sample grid is important for the

Map class (1)	Digitized map (cells) (2)	Max-membership polygons (cells) (3)	Difference (%) (4)	locational agreement (cells) (5)	Difference (%) (6)
1	21012	20719	1.4	15338	27
2	29945	33346	11.4	25267	15.6
I 3	I 11340	I 9492	I 16.3	7551	33.4
4	1721	1689	1.9	1107	35.7
5	1460	1219	16.5	726	50.3
6	1412	1718	21.7	1136	19.5
7	215	88	59.1	68	68.3
8	261	281	7.7	207	20.7
9	5007	4265	14.8	3029	39.5
10	2832	2168	23.4	1775	27.3
11	380	251	33.9	196	48.4
12	272	296	8.8	219	19.5
13	1214	1445	19	970	20.1
14	414	615	48.6	346	16.4
15	1951	1951	0	1951	0
16	9308	9308	0	9308	0
17	146	41	72.9	146	71.9
18	211	286	35.5	209	1
19	270	193	28.5	188	30.4
20	1542	1542	0	1542	0
Totals:		Number of cells ( $\Sigma(\text{column (2)})$ ) = 90913			
		Areal differences ( $\Sigma  (2) - (3) $ ) = 8516			
		Location differences ( $\Sigma  (2) - (5) $ ) = 19739			
Total areainaccuracy:		9%			
Total locati onaccuracy:		22%			

Table 2. Area and location evaluation of maximum-membership polygons for 1 00-by-100 m grid.

maximum-membership approximation of the map. However, increasing the spacing above the 50-by-50 m grid examined here would require considerably more computational time than what was used in this study (see, further, the discussion of practical implications, below). However, given that human operator time would not increase, if inaccuracy were further reduced it might be desirable to use a 13-by-13 m grid spacing (the size of individual cells).

Given these findings, it was necessary to examine the source of the inaccuracies in greater detail. There are two potential sources of these errors. One is that, in reality, one cannot base the certainty of observing a particular forest type at any given sample point solely on the amount of area stolen from a given polygon. In truth, this certainty should also be affected by the aspatial characteristics of the forest type-in particular the variability of the type-and

the spatial characteristics (e.g., size and shape) of a polygon. If a forest type is highly variable, the "true certainty" of observing it at any given point is not necessarily 100%. Similarly, if one has a perfectly round polygon that is "big," the point in the center may indeed have a 100% certainty of being that type. However, a point located in a "long, narrow" polygon might not actually have a 100% certainty of being that type. In this study, this type of error was not examined and is, instead, considered a necessary subject for future research.

The second potential source of inaccuracy is that differences between the digitized and maximum-membership maps were due to the "incorrect placement" of a digitized point within a polygon. One might argue that objective algorithms might produce better polygon centroids. This may or may not be true. But even if it were true, such an approach

would require the precise digitizing of map to be treated algorithmically. This-digitizing forest type lines exactly-is precisely what one is trying to avoid by using this approach. Though in this study points were located within polygons of an already digitized map, operationally one would register the map to a digitizing table and extract subjective centroids directly without the digitizing of boundaries as was done in this study. Thus the manner in which centroids were located within polygons in this study is comparable to what would be found in an operational context.

## Discussion

### Practical Implications

It is clear that, in its present form, the digitizing technique described is not capable of exactly reproducing a forest map as drawn, nor as it would be digitized by tracing lines exactly. This is true regardless of the sampling density of the grid used for area-stealing, though a finer density does improve the approximation. It must be understood that a portion of this inability is the irregular form of most forest polygons. Undoubtedly, if similar techniques were applied to an urban setting having relatively regularly-shaped and uniformly-sized polygons-e.g., city blocks-it is probable that both the areal and the locational accuracy would improve. However, given that some inaccuracy would undoubtedly remain, and that boundaries of such polygons are locatable and verifiable in "the real world," the techniques would not be applicable to such a situation.

The map used in this study contained 20 forest types. It is likely that a map containing fewer categories treated in the proposed fashion would have greater agreement with a conventional digitized product, assuming that polygons had the same degree of spatial irregularity. The overall effect of fewer categories is bigger polygons and, subsequently, fewer boundaries. In the digitizing method presented, boundaries are treated as the least certain spatial features on the map. If fewer boundaries exist, one would expect better agreement with a conventionally digitized map.

The inability to reproduce a digitized map is only a problem if one considers the physical map to be "truth," and/or if areal and locational inaccuracies are seen to exceed those inherent in the physical map. In cases in which all boundaries exist and are readily identifiable (on the ground and the map), the technique described is clearly

inappropriate. However, its applicability to natural phenomena must remain the choice of the individual user. If one believes, for example, that a soils map has an inherent 25% margin of error, then an areal and locational digitizing inaccuracy of 5% and 18% is probably acceptable. Similarly, if one believes there is a 10% margin of error in the map, and one is only concerned that the total amount of area of each forest type is represented adequately, then a 5% error is well within the limits of map accuracy. In fact, unpublished work by the first author suggests that the areal agreement among interpretations of the same aerial photographs for forestry purposes is approximately 60%; both of the fuzzy map representations presented were more precise than this.

Finally, the strongest argument for the approach presented is in time savings for data entry. In this study, a 1536-ha area having 20 forest types was digitized from a 1:20,000 map. This required approximately ten hours of human time during which a computer was in use as well. (Though a relatively powerful 486 was being used, this step could occupy a far less powerful machine if necessary.) The fuzzy surface-based approach required approximately two hours of human time plus another three hours of computer processing time on a 486 machine. Thus the approach adopted saved 80% of the human time required for digitizing an area in a conventional line-tracing procedure, but used more computer time and required a more powerful computer.

### Context

This paper discusses one possible way of deriving a fuzzy cartographic representation and one use to which it may be put. It is worthwhile here to provide additional background on fuzzy set theory (Zadeh 1965). One of the first discussions of using fuzzy set theory in geographic information systems can be found in Robinson (1988). However, the subject had been considered earlier in the classification of remotely sensed images (Robinson and Thongs 1985; Kent and Mardia 1988) in large measure because of a recognition that mixed pixels do not belong to a single type (Marsh et al. 1980). Indeed, such work continues, as evidenced by relatively recent articles on the subject (e.g., Wang 1990a and 1990b). Employing fuzzy set theory in GIS is now well established and the idea of fuzzy classifications has migrated to data other than satellite imagery (Burrough 1989; Burrough et al. 1992). The principal focus has been the attribute and spatial uncertainty inherent in

conventional maps. Studies have focused on digitizing error (Bolstad et al. 1990; Lee and Chow 1992) as well as on descriptions of how to accommodate and model uncertainty in spatial databases (Chrisman 1984; Chrisman and Lester 1991; Leung 1988). The concepts of fuzzy set theory have also been applied to the problem of spatial-temporal modeling (Lowell 1994). Other researchers have presented alternative methods for generating fuzzy surfaces given different types of map information than what was available here (Fisher 1992; Goodchild et al. 1992). Both Fisher (1992) and Goodchild et al. (1992) suggest also an alternative to the maximum-membership algorithm for class assignment. They propose, instead, the use of an algorithm which controls for overall error while maintaining a certain amount of spatial autocorrelation in the classification.

## Conclusions

Area-stealing interpolation techniques can be used to produce fuzzy surfaces of natural phenomena which represent the real world as an inherently uncertain surface. That is, boundaries are not fixed phenomena and spatial locations represented cartographically may be mixtures of two or more types, and/or have a possibility of being a map type other than that mapped on a choropleth map.

A practical application of these surfaces is a potential to diminish digitizing efforts of forest type maps. A relatively simple maximum-likelihood classification algorithm was able to produce a classified thematic map from the fuzzy surfaces with 5% and 18% area and locational disagreements, respectively, relative to a digitized map of the same area. Whether these tolerances are acceptable for a phenomenon being mapped must remain the decision of the user of a spatial database. Future work must focus on spatial operators for fuzzy surfaces, improved methods for storing these surfaces, and the production of fuzzy surfaces in a manner which incorporates spatial and aspatial weights into the certainty weights on each surface.

## ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Association of Quebec Forest Industries and the National Sciences and Engineering Research Council for funding this work. The authors specifically thank Jean Nantel of Canadian-Pacific

Forest Products for an exchange of ideas concerning similar work being conducted for an operational context. The authors also thank Pierre Beaulieu for digitizing the original map and three anonymous reviewers for many useful suggestions.

## REFERENCES

- Bolstad, P. V., P. Gessler, and T. M. Lillesand. 1990.** A variance components analysis of manually digitized map data. *Surveying and Land Information Systems* 50(3): 20 1-7.
- Burrough, P. 1989.** Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science* 40:477-92.
- Burrough, P., R. MacMillan, and W. Van Deursen. 1992.** Fuzzy classification methods for determining land suitability from soil profile observations and topography. *Journal of Soil Science* 43: 193-210.
- Chrisman, N. R., M. Lester. 1991.** A diagnostic test for error in categorical maps. *Proceedings AutoCarto* 10. Bethesda, Md.: American Congress on Surveying and Mapping, 330-48.
- Chrisman, N. R. 1984.** The role of quality information in the long-term functioning of a geographic information system. *Cartographica* 21(2 and 3) :79-87.
- Dent, B. D. 1985.** *Principles of thematic map design.* Reading, Mass.: Addison-Wesley.
- Fisher, P. F. 1992.** First experiments in viewshed uncertainty: Simulating fuzzy viewsheds. *Photogrammetric Engineering and Remote Sensing* 58(3): 345-52.
- Gold, C. M. 1989.** Surface interpolation, spatial adjacency, and GIS. In J. R. Raper ed. *Three-dimensional applications in geographical information systems.* London: Taylor and Francis, 21-35.
- Goodchild, M. F., S. Guoqing, and Y. Shiren. 1992.** Development and test of an error model for categorical data. *International Journal of Geographical Information Systems* 6(2): 87- 104.
- Kent, J. T., K. Mardia. 1988.** Spatial classification using fuzzy membership models. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMZ)* 10:659-7 1.
- Lee, Y. C., and A. Chow. 1992.** A comparison of manual and semi-automatic line digitizing. *Proceedings of the Fourth Annual Canadian Conference on GZS.* Ottawa, Canada: Canadian Institute of Surveying and Mapping, 32 (Abstract only).
- Leung, V. 1988.** *Spatial analysis and planning under imprecision* New York: Elsevier Science.
- Lowell, K. E. 1994.** Probabilistic temporal GIS modelling involving more than two map classes. *International Journal of Geographical Information Systems* 8( 1): 73-93.
- Mark, D. M. and F. Csillag. 1989.** The nature of boundaries on "area-class" maps. *Cartographica* 26( 1): 65-78.
- Marsh, S., P. Switzer, R. Lyon. 1980.** Resolving the percentage of component terrains within single resolution elements. *Photogrammetric Engineering and Remote Sensing* 46(8): 1079-86.
- Robinson, V. B. 1988.** Some implications of fuzzy set theory applied to geographic databases. *Computers, Environment, and Urban Systems* 12(2): 89-98.
- Robinson, V., D. Thongs. 1985.** Fuzzy set theory applied to the mixed pixel problem of multi-spectral landcover data bases. In B. Optiz ed. *Geographic information systems in government.* 87 1-86.
- Society of American Foresters. 1979.** *Forest cover types of North America.* Bethesda, Md.: Society of American Foresters.
- Wang, F. 1990a.** Fuzzy supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 28: 194-20 1.
- Wang, F. 1990b.** Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing* 56(8): 1163-g.
- Zadeh, L. A. 1965.** Fuzzy sets. *Information and Control* 8(3): 338-53. ■