

Research Article

Classification and boundary vagueness in mapping presettlement forest types

DANIEL G. BROWN

Department of Geography, Michigan State University, East Lansing,
MI 48824-1115, USA
email: brownnda@pilot.msu.edu

(Received 17 November 1996; accepted 13 July 1997)

Abstract. Presettlement forest types were mapped as fuzzy sets from point data representing trees contained in General Land Office survey notes (circa 1850) for Chippewa County, Michigan. The resulting representation agreed with a polygon map of the same forest types at 66% of the locations (represented as grid cells) in the county. Boundary vagueness was defined in relation to the slope of a linear function fitted to the negative relation between entropy of forest types and distance to polygon boundaries. The similarity between forest type compositions (i.e. classification ambiguity) was shown to account for 55% of the variation in boundary vagueness.

1. Introduction

Traditional methods of mapping vegetation (Küchler 1955, Küchler and Zonneveld 1988) use a representation of stands of vegetation as discrete spatial units that belong to one of a finite number of pre-defined types. (I will refer to such representations as 'the discrete model'.) Producers and users of vegetation maps often recognize the limitations of these representations but have had few alternatives (e.g. Küchler 1988). The discrete model of vegetation patterns is particularly inadequate where representations of spatial gradients or spatial patterns of boundary uncertainty are desired.

Recent research in geographical information systems (GIS) and digital mapping has advanced methods for representing inexactness or uncertainty in the positions of spatial objects, like vegetation unit boundaries (Buttenfield 1993, McGranaghan 1993, Fisher 1994, Goodchild *et al.* 1994 a, Goodchild *et al.* 1994 b, Lowell 1994), and explicated some causes of boundary vagueness (Edwards and Lowell 1996). The focus of GIS and mapping on discrete objects is consistent with the traditional methodological approaches to mapping, but it limits representations to entities that are thematically discrete or defined by crisp boundaries. Specifically, although much work in biogeography focused on understanding natural vegetation patterns is based on the concept of a continuum of species combinations, most vegetation maps represent vegetation in spatially and thematically discrete units. A more continuous model of vegetation types is offered that can be constructed from point data. This model is used to test the role of classification ambiguity in affecting boundary vagueness.

In order to formalize the mapping process, vegetation mapping is viewed as the

simultaneous classification, aggregation, and interpolation of sampled vegetation or related data. Sample data might be in the form of species enumerations in plots, individual plant locations, or remotely-sensed pixels. Classification, the process of combining separate phenomena into classes or types is a form of *attribute generalization*. Classification of vegetation units can precede mapping or it can occur simultaneously, that is, as the mapper draws a line she/he observes and decides which vegetation types are to be created. Aggregation involves the combination of point samples or remotely-sensed pixels into polygons representing vegetation types. Interpolation is required when point data are used because vegetation is usually thought to be continuous. Aggregation and interpolation are referred to collectively as *spatial generalization*. With traditional manual mapping, this process occurs as the mapper draws lines around polygons. A polygon is an aggregated spatial unit and the location of its boundary is usually determined through an interpolation process between data points or areas of different spectral values.

The goals of this paper are to: (1) describe an automated approach for aggregating and interpolating vegetation types from point data that results in a more continuous, as opposed to discrete, representation; (2) apply the method to presettlement tree point data in Chippewa County, Michigan, using a forest type classification that is identical to that previously used to produce a 'discrete model' forest type map of the study area; (3) empirically assess the limitations of the method and consistency with a traditional, manual, and discrete map of the study area; and (4) analyse the relations between classification ambiguity and boundary vagueness. By addressing the last objective, one of several possible causes of boundary vagueness is evaluated in presettlement forest types. I hypothesize that adjacent forest type polygons with similar compositions are more difficult to delineate using manual or automated methods than adjacent polygons with very different forest types definitions, and that the boundaries between such polygons will be more vague as a result.

After a review of some of the differences between discrete and continuous models of vegetation patterns, the additional background and definitions related to fuzzy sets is provided. Descriptions of the study area and the data follow. The approach to mapping forest types from point data in two stages is outlined. First, the method used to define the relative degree of membership of each tree species in each forest type is described. Because the goal of this paper is to compare discrete and continuous representations of forest types for the same area, the method used is one that allows for such comparison by ensuring that the classifications used in each case are identical. Secondly, the interpolation process is described. Descriptions of analysis methods are followed by a description of the results and a discussion.

1.1. *Classification and continua*

An important debate has been carried out in the vegetation science literature regarding classification that has relevance to the current project. In North America the debate began to take shape in the writings of Clements (1936) and Gleason (1939), and was centred on the fundamental nature of vegetation and the degree to which plant associations were real entities or simply chance occurrences (McIntosh 1967). Clements (1936) argued for the treatment of the plant association as an 'organism' that arises, grows, and matures; whereas Gleason (1917) argued that any observed associations between species were merely the results of species and individual plant responses to environmental variations that have no direct bearing on interpretations of ecological patterns (Glavac *et al.* 1992).

Gleason's (1917) writings gave birth to the 'continuum concept', which describes vegetation variation in terms of gradual change in species composition with respect to environmental gradients, like slope steepness, soil moisture and precipitation (McIntosh 1967, Austin 1985). However, the experience of many of the investigators in mapping European vegetation (e.g. Tüxen 1954, Braun-Blanquet 1964) led them to conclude that vegetation units often are discrete entities on the landscape; this idea has been called the 'community-unit concept' (Shipley and Keddy 1987). In response to proponents of the continuum some writers argued that, regardless of whether or not communities are real entities, some form of classification was a practical necessity for mapping purposes. Küchler (1988, p. 109) stated, for example, that 'it is not often practical to map continua and the mapper usually ignores them'.

There is some truth to both the continuum and community-unit concepts (Danserau 1968). Natural vegetation is said to vary continuously with respect to environmental and resource gradients, but abrupt spatial changes in important gradients can lead to discrete or nearly discrete unit definitions on the landscape. Where environmental gradients change more slowly across space, more gradual changes in vegetation composition and structure might be expected. In some instances, disturbances (e.g. fires and wind-throw) and vegetation dynamics (e.g. succession and dispersal) help to smooth out spatial variations, in other cases these processes accentuate abrupt boundaries. Although the continuum concept argues for mapping gradual spatial variation in vegetation types, the practical needs for mapping have led to a prevailing mapping model wherein vegetation is viewed as discrete units. What are needed are alternative methods that might allow for the mapping of gradual vegetation changes (i.e. ecotones) where they are present.

1.2. Application of fuzzy sets

The central idea of fuzzy set theory is that human understanding is imperfect (i.e. inexact) and that phenomena in nature rarely fit perfectly the categories into which they are placed (Zadeh 1965). Unlike traditional set theory, where set memberships are crisp and binary (i.e. an entity is completely a member of a set or not at all), fuzzy set theory permits partial membership. The degree of membership is represented by a fuzzy membership value that usually ranges between zero and one, the extremes being the only available membership values in traditional set theory.

Fuzzy memberships differ from probabilities primarily in interpretation. Probability theory assumes that only one class or set is present and expresses the degree to which its presence is likely as a probability. The class with the highest probability is interpreted as the actual class (i.e. winner take all). Fuzzy set theory accepts that multiple classes or sets can be present at one place or at one time and expresses the degree to which each class or set is present as a membership value. A vector of fuzzy memberships is maintained and classes or sets with non-zero memberships are interpreted as present to some degree.

Two types of inexactness have particular relevance in the context of GIS: 'attribute ambiguity' and 'spatial vagueness'. *Attribute ambiguity* refers to the fact that we cannot assign membership of a given location in a singular type (e.g. a forest type) with 100% certainty. *Spatial vagueness* refers to the fact that boundaries drawn between sets as mapped in space (e.g. forest type boundaries) are not located with 100% certainty.

Ambiguity in thematic attributes, vagueness in spatial objects, or both often affect geographic data quality. A land cover map, for example, will have ambiguous

attributes when the types on the map are not well defined, distinguished from one another, or easily distinguished using the primary data (e.g. aerial photography). Boundaries between types will be vague to some extent if the resolution of the primary data is coarse, if the spatial transitions between polygons of the types are gradual, and/or if the definitions of the thematic types produced as a result of the mapping process are ambiguous in terms of the data on which the map is based. If there is ambiguity in the relations between information inputs and classification outputs then we might also expect to have difficulty locating boundaries between the output types, resulting in vague boundaries.

A need to develop geographical objects that represent natural entities with uncertain boundaries has been recognized for years (Campbell 1978, Blakemore 1984). With the exceptions of Fisher (1994), Lowell (1994), and Edwards and Lowell (1996) the focus of much of the research on fuzzy set applications in geography has been on characterizing fuzzy attribute data. In other words, the spatial units are usually represented with crisp boundaries, but the attributes describing those units are treated as ambiguous or uncertain.

1.3. *Mapped forest types as fuzzy sets*

Burrough (1989) made two arguments against the use of crisp mapping units for soil types that are relevant to an analysis of forest types: spatial variation within mapping units and the experimental error of measuring attributes of entities. Spatial variation within a forest type can be caused by environmental heterogeneity, such as might cause the presence of inclusions (very small areas of forest types included within a larger area of a different type), or by the presence of gradual transitions between forest types, such that composition of the forest type may change as distance from the boundary line increases (Glavac *et al.* 1992).

The nature of transitions between forest types is related to the causes of spatial autocorrelation in vegetation patterns. Malanson (1985) outlined two major causes of spatial autocorrelation in plant patterns: (a) spatial autocorrelation in the underlying environmental complex, and (b) seed dispersal as an inherently spatial process. The first cause suggests that where environmental gradients change abruptly in space, abrupt changes in species composition are to be expected. Similarly, where environmental gradients change gradually, gradual changes in species compositions might be expected. However, Timoney *et al.* (1993) hypothesized that changes in species patterns across gradual gradients might be more abrupt than the changes in the gradients themselves because of the vegetation switch, a positive feedback by which vegetation stabilizes its environment. This hypothesis is based on their work at the subarctic forest-tundra ecotone. The second cause of spatial autocorrelation is independent of gradients and suggests that, even in the presence of an environmental discontinuity, vegetation can have a more continuous pattern across boundaries. Situations in which this is important include frequently disturbed areas, where turnover of individuals and immigration are high, and where founder effects are important, inhibiting or facilitating certain successional pathways (Malanson 1985).

Uncertainties in boundary position can arise because of imperfect data (Burrough's second argument). This point, though true generally, is particularly pertinent to the mapping of presettlement vegetation types from tree data recorded in the public land survey records (as attempted here). A description of this data set follows in § 3.

Forest types from tree data collected at points by, first, defining the relationships between trees and forest types (i.e. attribute generalization) and, second, spatially interpolating the forest types (i.e. spatial generalization) are mapped here. As they result from the classification process, forest types are viewed here as fuzzy sets. Each tree, independent of its location, belongs to a forest type to some degree depending on its species. This relation between a tree's species and the forest type to which the tree gets assigned is ambiguous and uncertain.

As spatial objects, polygons representing natural forest types also are considered to be fuzzy. Each location belongs to a vegetation type depending on the species of the trees nearby and the spatial pattern of those trees. Manual interpretation of a pattern of points involves the placement of a line where the mapper detects a shift in the species composition from one place to another. If the variation in natural vegetation across that boundary is gradual, or the species compositions on either side of the line are similar, the mapper might expect to have more difficulty placing that line. Also, the species of the trees will suggest to the mapper that a given forest type might be present, but they will not, except in the instance of a species that is unique to an ecotone, indicate to the mapper the presence of a boundary. Boundaries are drawn *between* data points, not *at* or *through* data points, and, therefore, in the absence of data (i.e. in the spaces between trees). This effect is dependent on the sampling strategy and becomes especially important where the spaces between sample points can be large and where samples are not selected on the basis of their value as indicators of transition. Even in the presence of a real discrete boundary between forest types, the boundary line may be inexact because of uncertainty in the data or the interpretation.

2. Study area

This study was conducted using data collected (Price 1994) for the Western portion of Chippewa County, which borders Lake Superior in the Eastern portion of Michigan's Upper Peninsula (figure 1). The study area has a cool, humid continental climate. The growing season in the study area ranges from as low as 100 to 140 days (Albert 1995). Average annual precipitation is between 75 and 90 cm, with an average of between about 250 and 370 cm of snowfall.

The County was de-glaciated with the retreat of the Wisconsin glacialiation (about 10000 y.b.p.). Major landforms in the County include ground moraines, an outwash plain, lake plains and eolian deposits (Whitney 1992). The largest moraine contains coarse textured materials and runs northwest to southeast through the west-central portion of the study area. Just east of the moraine is the Raco Plains, a coarse-textured outwash plain that gradually blends with the fine textured lake plain materials (with areas of sandy lake plain and beach ridges) to the east, formed by glacial Lake Algonquin. Sandy eolian deposits can be found in the northern portion of the county with soils forming on organic deposits. Other organic deposits are found scattered throughout the county.

Landforms and, locally, soil texture and drainage controlled presettlement forest communities in the county (Albert 1995). Somewhat poorly to poorly drained clay soils on the lake plain supported stands of northern white cedar, tamarack, black spruce, hemlock, and trembling aspen, sometimes in stunted growth forms. Sugar maple, beech, basswood, and yellow birch (i.e. Northern Hardwoods) were common on the better drained soils, including those on the moraine. White pine and red pine were dominant on sandy dry sites, often on the beach ridges and sandy eolian



Figure 1. Location of the study area, Chippewa County, Michigan.

deposits. Jack pine dominated the driest sandy sites, especially the Raco Plains, and was associated with frequent fires (Price 1994). Windthrow disturbance was also common, especially on the lake plain and in the swamps (Albert 1995).

3. Public land survey records

The public land surveys conducted by the federal General Land Office (GLO) have provided an important source of historical information on the nature of North American forests near or prior to the time of European settlement. Data on individual tree characteristics (location, species, diameter) from the GLO notes have been used in numerous studies on the relations between Native Americans and primeval forests (Neuenschwander 1957, Dorney 1981), forest community characteristics at the time of European settlement (Blewett and Potzger 1950, Shanks 1953, Siccama 1971, Anderson and Anderson 1975, Delcourt 1976), natural disturbance types and frequencies in presettlement forests (Canham and Loucks 1984, Whitney 1986, Cowell 1995), and edaphic controls of presettlement vegetation (Crankshaw *et al.* 1965, Catana 1967, Whitney 1986, Leitner *et al.* 1991, Barrett *et al.* 1995).

Surveyor instructions issued in 1833 and 1850 required surveyors to identify four trees near each section corner (witness trees) and mark the trees with locational information (White 1984). The names of trees, usually specific enough to identify the species, and estimates of tree size were recorded. Along each section line, trees falling on the line (line trees) as well as one tree at the mid-point between section corners (bearing trees) were also recorded. Also, wherever the surveyor crossed a conspicuous landscape boundary (e.g. into and out of wetlands, recently burned areas, and clearings) the location was noted. Summary information describing the landscape was recorded for each section line, along with a summary of the most

abundant tree species encountered. Although this paper has focused exclusively on the location-specific tree data, automation of the other landscape information contained in the GLO data can be added to the information base for mapping presettlement landscapes (Price 1994, Barrett *et al.* 1995).

Maps of presettlement vegetation have been produced from the data contained in the GLO notes for parts of Michigan and elsewhere in the region at scales of 1:250 000 and smaller (Veatch 1959, Marschner 1975, Finley 1976, Brewer *et al.* 1984, Price 1994, Barrett *et al.* 1995). The survey data are well suited to regional-scale mapping because of the regularity of the sample, which consists of two perpendicular sets of parallel transect samples with transect spacings of approximately 1600 m (one mile), and the large area covered by the survey.

4. Methods

4.1. Data collection and crisp-boundary mapping

Survey notes from the GLO for Michigan were available at the State Archives in Lansing. Chippewa County had been surveyed between 1840 and 1850. The quality of these surveys was likely to be very high. The township lines were surveyed by William Austin Burt, one of the best known of Michigan's surveyors (Burt 1985). Whereas re-surveys were ordered in other areas because the original surveys were found to be fraudulent or inaccurate, no re-surveys were required in the study area.

There are several potential problems with the survey data for ecological analysis (Cottam 1949, Buordo 1956, Cowell 1995). The only assumption made for this project was that the relative density of each species was adequately represented throughout the study area. Some species may have been preferred or avoided by the surveyors by virtue of their average size, bark type, and branching configurations. For example, smaller trees tended to be recorded less frequently in the survey data than expected in a natural forest community (Buordo 1956). Nonetheless, the data represent the best source of pre-European settlement vegetation data available.

Data on tree location (relative to section corners), species (or genus if species was not available), and diameter at breast height (dbh), among other information, had been collected and entered into a computer database (Price 1994). Michigan State Plane coordinates of each tree were then calculated by comparing the tree location information with a listing of section corner coordinates obtained from the State of Michigan, Department of Natural Resources and assuming true north-south and east-west section lines. In all, 13 965 trees were entered for the study area, which had an area of approximately 2730 Km². Therefore, the average sampling density was approximately five trees per Km² (or about 13 trees per survey section).

Price (1994) had interpreted and mapped the boundaries of six forest species associations from the tree data (table 1). The types were manually interpreted from displays of the tree data and resulted in discrete boundaries between neighbouring forest types. Forest type definitions evolved on the basis of species associations identified on the data displays. Although topographic information, also recorded by the surveyors, was used to supplement them, the tree data were the primary data used for forest type delineation. In some cases, indicator species were used and the occurrence of a single individual was sufficient evidence for drawing a new boundary (Price 1994). Boundaries between forest types were digitized directly on the computer screen using a mouse.

Table 1. Forest types as identified by Price (1994).

| Forest type name | Abbreviation | Dominants |
|-----------------------|--------------|---|
| Northern hardwoods | hwds | hemlock, sugar maple, yellow birch, white pine, beech |
| Mixed conifer upland | c upl | white pine, hemlock, yellow birch, spruce |
| Mixed pine | pin | white pine, red pine |
| Jack pine | j pine | jack pine |
| Mixed conifer swamp | c bog | white cedar, spruce, white pine, tamarack |
| Mixed conifer lowland | c lowl | white pine, spruce, tamarack, white cedar, hemlock |

4.2. Determination of species memberships

In the automated mapping approach presented here, forest type membership values at a location are obtained using kriging (i.e. a weighted spatial averaging of membership values). Fuzzy membership values for each tree species were calculated, therefore, to describe the degree to which each tree was a component of each vegetation type on a scale from zero to one.

Multiple methods might be used to arrive at the membership values, depending on the desired classification scheme and the amount of *a priori* information available about the classification scheme (Brown, in press). In each case a list of species in each type and their relative densities is required. To get this information, one might rely on vegetation types described in the literature, for example in Burns and Honkala (1990) and Eyre (1980). A second method might use a clustering algorithm based on fuzzy sets (Bezdek 1984) to identify natural groupings of species on the landscape. Neither of these methods relies on an existing vegetation map and can be used to apply the automated mapping method anywhere that tree species point data are available. However, in order to compare the automated forest types with an analogous set of discrete types, the forest types identified and mapped by Price (1994) in Chippewa County served as the reference units for this analysis. The goal was to compare two different spatial representations of the same forest types and the types as defined on the forest type map were accepted. Price (1994) had, therefore, completed the classification (i.e. attribute generalization) phase of the project.

The first step in calculating the membership values was to tabulate the number of individuals of each species in each forest type. Next, the percentage composition of each type was calculated by dividing the number of trees of each species by the total number of trees in that type (i.e. column totals) and multiplying by 100 (table 2). Percentage Composition columns for each type in table 2, therefore, sum to 100. Finally, the degree to which each type was present given each species (i.e. membership values) was calculated by dividing the percentage value for each type by the sum of all percentage values for that species (table 2). For each species, then, the sum of membership values is one (i.e. rows of Membership Values in table 2 sum to one).

The first set of values in table 2 (Percentage Composition) indicated the relative densities of each species in each forest type. The second set of values (Membership Values) can be interpreted as the degree to which each tree species is indicative of (or is a member of) each of the forest types. Note that species that constituted only small percentages of the trees in a given type (for example, *Tilia americana* and *Ostrya virginiana* in the Northern Hardwoods type) may have had relatively high membership values by virtue of their selectivity for that type.

Table 2. Relative density of species within forest types and resulting membership values.*

| Species | N | Percentage composition | | | | | | Membership values | | | | | |
|---|------|------------------------|-------|------|--------|-------|--------|-------------------|-------|------|--------|-------|--------|
| | | hwds | c upl | pin | j pine | c bog | c lowl | hwds | c upl | pin | j pine | c bog | c lowl |
| <i>Abies balsamea</i> (L.) mill. | 1049 | 6.6 | 10.6 | 1.4 | 0.0 | 6.5 | 19.3 | 0.15 | 0.24 | 0.03 | 0.00 | 0.15 | 0.43 |
| <i>Acer rubrum</i> L. | 460 | 5.4 | 5.9 | 0.5 | 0.0 | 1.1 | 5.1 | 0.30 | 0.33 | 0.03 | 0.00 | 0.06 | 0.28 |
| <i>A. Saccharum</i> Marsh. | 1158 | 27.2 | 5.9 | 0.1 | 0.0 | 0.1 | 0.9 | 0.80 | 0.17 | 0.00 | 0.00 | 0.00 | 0.03 |
| <i>Betula Alleghaniensis</i> Britton | 1074 | 14.1 | 12.8 | 1.2 | 0.0 | 3.2 | 6.5 | 0.37 | 0.34 | 0.03 | 0.00 | 0.08 | 0.17 |
| <i>B. Papyrifera</i> Marsh. | 252 | 0.7 | 2.8 | 5.4 | 0.6 | 1.3 | 1.1 | 0.06 | 0.24 | 0.45 | 0.05 | 0.11 | 0.09 |
| <i>Fagus grandifolia</i> Ehrh. | 639 | 16.9 | 0.6 | 0.1 | 0.0 | 0.0 | 0.0 | 0.96 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| <i>Fraxinus nigra</i> Marsh. | 157 | 0.7 | 1.6 | 0.0 | 0.0 | 1.4 | 2.5 | 0.12 | 0.25 | 0.00 | 0.00 | 0.23 | 0.40 |
| <i>Larix laricina</i> (Du Roi) K. Koch | 1243 | 0.3 | 2.6 | 0.6 | 0.4 | 22.5 | 9.3 | 0.01 | 0.07 | 0.02 | 0.01 | 0.63 | 0.26 |
| <i>Ostrya virginiana</i> (Mill.) K. Koch | 21 | 0.4 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.70 | 0.26 | 0.00 | 0.00 | 0.03 | 0.00 |
| <i>Picea spp.</i> | 2126 | 3.5 | 14.8 | 7.0 | 0.8 | 28.0 | 18.9 | 0.05 | 0.20 | 0.10 | 0.01 | 0.38 | 0.26 |
| <i>Pinus banksiana</i> Lamb. | 995 | 0.0 | 1.9 | 14.4 | 94.1 | 5.0 | 0.0 | 0.00 | 0.02 | 0.12 | 0.81 | 0.04 | 0.00 |
| <i>P. resinosa</i> Ait. | 643 | 0.3 | 2.4 | 31.9 | 1.7 | 1.7 | 0.9 | 0.01 | 0.06 | 0.82 | 0.04 | 0.04 | 0.02 |
| <i>Pinus strobus</i> L. | 1206 | 3.1 | 10.2 | 31.5 | 1.7 | 5.0 | 11.2 | 0.05 | 0.16 | 0.50 | 0.03 | 0.08 | 0.18 |
| <i>Populus spp.</i> | 217 | 0.6 | 1.8 | 2.8 | 0.4 | 0.8 | 6.1 | 0.05 | 0.14 | 0.22 | 0.03 | 0.06 | 0.49 |
| <i>Thuja occidentalis</i> L. | 1424 | 4.3 | 8.1 | 0.9 | 0.4 | 20.1 | 10.0 | 0.10 | 0.18 | 0.02 | 0.01 | 0.46 | 0.23 |
| <i>Tilia Americana</i> L. | 19 | 0.4 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.62 | 0.12 | 0.00 | 0.00 | 0.00 | 0.26 |
| <i>Tsuga canadensis</i> (L.) Carr. | 1185 | 14.8 | 17.0 | 1.4 | 0.0 | 2.9 | 5.9 | 0.35 | 0.40 | 0.03 | 0.00 | 0.07 | 0.14 |
| <i>Ulmus americana</i> L. | 33 | 0.2 | 0.4 | 0.0 | 0.0 | 0.1 | 1.0 | 0.13 | 0.24 | 0.00 | 0.00 | 0.04 | 0.60 |

* Although they were included in the analysis, species with fewer than 10 occurrences in the study area are not listed in the table to save page space. For this reason the percentage composition columns do not add up to 100.

Each point in the data set, representing a tree, was assigned membership values for each forest type according to its species. These membership values were stored in the attribute table referenced to the point features.

4.3. Interpolating membership values

The membership values of each forest type were interpolated and aggregated from the tree point data to create grid-cell (raster) representations of the continuous variations (i.e. surfaces) using geostatistical methods (i.e. kriging). Several good reviews of geostatistical theory and applications are available for readers wishing more information (Isaaks and Srivastava 1989, Oliver *et al.* 1989 a,b; Rossi *et al.* 1992). The method involved the following steps: (i) calculating and plotting the sample semi-variogram (i.e. the relation between distance and the variation between membership values); (ii) fitting that relation with a positive definite function; and (iii) calculating the best linear unbiased estimate of membership values for each grid cell in each forest type (i.e. kriging). Geostatistical calculations using the FORTRAN language geostatistical subroutines by Deutsch and Journel (1992) were implemented. All subsequent grid calculations and manipulations were performed using the ArcGrid v 7.0.3 software (ESRI, Inc., Redlands, CA).

All forest type semi-variograms exhibited spatial autocorrelation, as indicated by a rise in the semi-variance value with increasing distance (figure 2). The semi-variograms were visually fitted with models using the Variowin v 2.2 program (Pannatier 1996) and are described by the nugget variance, variance contribution, and range (table 3). The *nugget* variance is the level of variance at zero distance. A *sill*, or stable level of semi-variance with distance, is usually reached at some distance (called the *range*). The sill minus the nugget is the *variance contribution*. Beyond the range, distance has little effect on the measured variance and within distances smaller than the range a variable is said to be spatially autocorrelated. The models take the exponential form:

$$\gamma(h) = c_0 + c \left[1 - \exp\left(-\frac{h}{a}\right) \right] \quad (1)$$

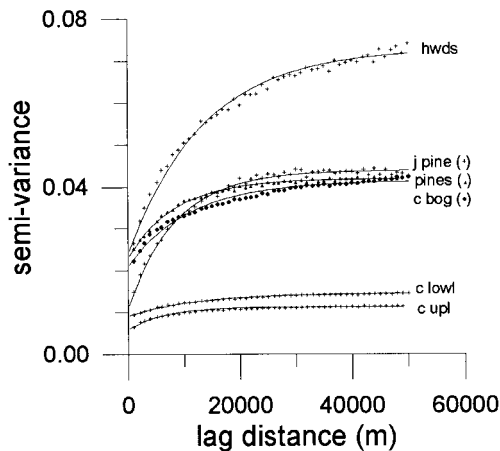


Figure 2. Experimental semi-variograms, displayed as point symbols, and fitted model variograms, displayed as solid lines, for each of the six forest types defined in table 1.

Table 3. Parameters used in exponential models (equation (1)) of forest type semi-variograms.

| Forest type | Effective range (a) | Nugget (c_0) | Variance contribution (c) |
|-------------|-------------------------|------------------|-------------------------------|
| hwds | 40 800 | 0.024 | 0.049 |
| c upl | 19 400 | 0.006 | 0.0054 |
| pin | 25 500 | 0.023 | 0.019 |
| j pine | 25 500 | 0.011 | 0.033 |
| c bog | 33 150 | 0.021 | 0.0205 |
| c lowl | 36 000 | 0.0092 | 0.0054 |

where $\gamma(h)$ is the semi-variance at lag distance h , c_0 is the nugget, c is the sill minus the nugget (i.e. *variance contribution*), and a is the range. In the exponential model, semi-variance approaches the sill asymptotically. For this reason the range parameter is referred to as the *effective range* and is equal to that distance at which 95% of the variance contribution is reached. Isotropic and stationary variation was assumed for all interpolations in this paper (Isaaks and Srivastava 1989).

Each of the forest type membership surfaces produced by this process (figure 3) consisted of grid cells (or blocks), 303 m by 303 m in size (approximately one-fifth the size of a section line on a side), for which membership values were estimated using ordinary block kriging. Membership values for the grid cells were estimated by averaging the weights that would be used to estimate multiple point values within each block. Weights were solved for sixteen points within each block using the variogram models in table 3 and a minimum of 8 and a maximum of 36 points that were no more than 40 800 m from the point being estimated. These parameters were set with the basic goals of ensuring (i) an adequate sample of actual data points on which to make the estimate (in this case, at least 8), and (ii) that no data points that were outside the range of influence (more than 40 800 m) were used in the estimation. The distance threshold was set to the longest range of the six forest type variogram models (table 3). Including points at greater distances would do little to change the interpolation because they would receive very low weights.

Most definitions of vegetation types and/or associations have included some reference to environmental homogeneity (Shimwell 1971). Although the point data on tree species were not supplemented by any ancillary data in this application, co-kriging can be used to introduce environmental co-variates into the interpolation of membership values (Brown, in press, Isaaks and Srivastava 1989).

Intuitively, the sum of membership values at all locations should be one. At each tree point, membership values sum to one, but this condition does not necessarily hold following interpolation. The interpolated membership grids were normalized so that, for each location, they summed to one using equation (2):

$$V'_j = \frac{V_j}{\sum_{i=1}^k V_i} \quad (2)$$

where V'_j is the normalized membership value for forest type j at a location, V is the non-normalized value, and k is the number of classes.

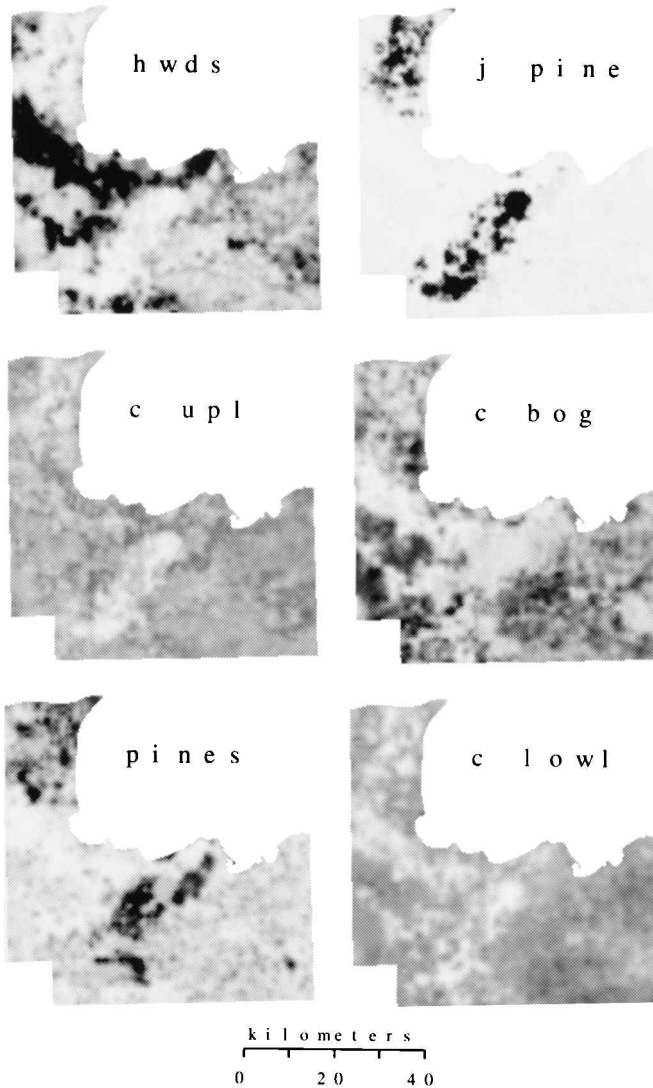


Figure 3. Fuzzy membership maps for six forest types. Darker values indicate higher membership (black saturates at a membership value of 0.55).

4.4. Comparison methods

Having automated the forest type mapping process, the automated representation was compared with a rasterized version of the manual representation to assess the consistency with traditionally accepted methods. If the automated representation agreed well with the manual, then it could serve as a model for experimentation. The more continuous representation, resulting from the above procedures and consisting of six grid-based surfaces (figure 3), was compared with the discrete-model vegetation map developed by Price (1994) to test the amount of agreement. All grid-based analyses were conducted using the same grid cell size (303 m) and with the aid of the ArcGRID software (ESRI, Inc., Redlands, CA).

The agreement between the representations by overlaying a raster version of the

discrete map with a classified version of the continuous representation was tested. In a procedure analogous to the maximum likelihood classification procedures used in remote sensing work, and treating the membership values as probabilities, a classified map was created by assigning each cell to the type for which its membership value was highest. The two classified maps were compared using the proportion correctly classified (PCC) and kappa statistics. PCC is the proportion of sites assigned to the same type on both maps (Congalton 1991). The kappa statistic is a corrected version of PCC, accounting for chance agreement between the maps (Bishop *et al.* 1975). Both PCC and kappa have theoretical ranges of zero to one, with one indicating complete agreement. The maps were compared first using all grid cells, then using only the cells with the highest maximum membership values in the fuzzy representation to test the effect of strength of membership on the level of agreement.

4.5. Assessing sources of disagreement

Gopal and Woodcock (1994) outlined procedures for using fuzzy sets in remote sensing accuracy assessment. Although this analysis is not an accuracy assessment, the calculation of levels of agreement between the two maps uses some of the same procedures as accuracy assessment. Gopal and Woodcock's *difference* measure was used here to indicate the magnitude of disagreement between the hand-drawn map and the automated representation. The difference, $\Delta(x)$, 'measures the difference between the (membership value) of the (crisp-) map category ... (at location) x and the highest score given to x among all other categories' (Gopal and Woodcock 1994, p. 185). Therefore, $\Delta(x)$ is positive if the membership value for the crisp-map category is the maximum among all categories. Negative values result when the map category has a lower membership value at location x than any of the other categories.

Two likely sources of disagreement in the representations were examined: similarity in forest type definitions and small or fragmented forest stands. Because both mapping approaches rely on having well defined forest types, it was expected that disagreement would be higher for forest types that had greatest similarities in definition to the other types. The role of type definition in affecting disagreement was assessed by summarizing $\Delta(x)$ values first by forest type category. The mean differences ($\bar{\Delta}$) for each type were compared with the sum of correlation values (Pearson's r) between species membership values of that type with all other types to assess the effect of similar type definition on distinguishability of the type on the map. The summed correlation values for each type served as an indicator of the overall uniqueness of each forest type in terms of species composition.

Spatial patterns in the input data were expected to influence the drawing of boundary lines on the map by affecting decisions (manual or automated) about boundary placement and map type assignment. A strongly clustered pattern of trees with strong membership in a given type will be easier to interpret, for example, than a more dispersed cluster of trees with strong or weaker membership in the same type interspersed with members of other types. In order to assess the influence of spatial patterns on the level of agreement between fuzzy and crisp representations, the semi-variograms were first interpreted for information on the strength and scale of spatial dependence in each of the forest types. The variance contribution of the variogram provides information on the degree to which similar values of forest type membership were clustered in space, and the range suggested the scale over which the values of each type were spatially dependent. It was expected that forest types with strongly clustered patterns (i.e. those with higher levels of variance contribution)

would be easier to identify and, therefore, to have higher levels of agreement between the manual and automated methods.

Similarly, the mapping process should be affected by stand geometry. It was expected that smaller stands would be harder to identify accurately using the automated method because the identification of these stands is based on less data that are usually in more heterogeneous neighbourhoods. Also, boundaries around smaller stands and/or stands with small perimeters might be expected to be more difficult to identify, especially with an automated method that averages out fine-scale variability, like kriging. The role of stand area and perimeter was assessed by averaging difference values for each of the 210 forest stands on the crisp-boundary map. Mean stand difference values ($\bar{\Delta}$) were regressed against the area and perimeter of separate polygons defined on the crisp-boundary type map. *T*-tests were then used, assuming unequal variances, to test for significant area and perimeter differences between stands with positive versus negative mean difference values.

4.6. Assessing ambiguity-vagueness relations

Theoretically, boundaries between forest types drawn on a map are at the location of maximum ambiguity between types (i.e. where a greater certainty of membership in one type transitions to a greater certainty of membership in the neighbouring type). For the reasons outlined above (i.e. gradual transitions between types and uncertainty in boundary location), it was considered that ambiguity would decline with distance from boundaries. A formalized measure of ambiguity between types at each location x , similar to Shannon's Diversity Index, called the classification *entropy* [$h(x)$] was used:

$$h(x) = -\frac{1}{\ln(k)} \sum_{i=1}^k V_i(x) \ln [V_i(x)] \quad (3)$$

where k is the number of types and $V_i(x)$ is the membership value for type i at location x . Entropy varies from zero to one, with a value of one indicating that all types have the same value ($1/n$) and a value of zero meaning that one type has a membership value of one and all others are zero. Values of $h(x)$ were mapped and compared with straight-line distance to the nearest crisp boundary defined by Price (1994), to assess the relationship between ambiguity and boundary vagueness.

Given this definition of classification ambiguity (i.e. entropy), a formal definition of the boundary vagueness was adopted for this analysis: a slow rate of decline in type ambiguity with distance from a boundary. A sharp ridgeline at a boundary on a three-dimensional entropy surface would indicate that the types on either side of the boundary are fairly unambiguous right up to locations near the boundary, so the location of the boundary is more certain. This should occur where fairly homogenous stands of one forest type abut homogenous stands of a very different forest type. If ambiguity decreases slowly from a boundary location (e.g. a plateau of ambiguity) then the location of the boundary might be considered more vague because the boundary conceivably could be located at some distance from its drawn location. This would occur where forest types mix gradually, are interdigitated, or where environmental heterogeneity causes less distinct occurrences of the forest types on either side of the boundary. Linear functions were fitted to model the relations between distance to boundaries and entropy.

Finally, how the similarity of neighbouring forest type definitions affected the vagueness of the boundary between them (using the above definition) was examined.

Where neighbouring forest types had more similar definitions, it was expected to find that (a) entropy measured at the boundary was greater, and (b) polygon boundaries were more vague than where neighbouring types were less similar. Boundary types were classified such that each pair of forest types on either side of a boundary constituted a unique boundary type. Separate distance-entropy relations for each boundary type were plotted, determined to approximate linear relations, and fitted with linear functions to measure boundary vagueness (i.e. using the slope of the function) for several boundary types. If the Sigmoid Wave Hypothesis (Timoney *et al.* 1993) is correct and holds for the transitions being mapped, we should expect a nearly linear relations between the degree of mixing of the two forest types and distance from the boundary at short distances that would level off at greater distances. Average entropy was calculated for grid cells within a succession of distance categories. A grid cell was included in the entropy calculation for a boundary type only if the cell's *nearest* boundary was of that type. To minimize the influence of singular or extraordinary polygons on the calculations, the functions were only calculated for each of eleven boundary types for which there were at least seven occurrences (an arbitrary number to limit small numbers of polygons). To limit the influence of small grid cell sample sizes which can bias the shapes of the functions, the functions were calculated to a maximum distance of 1000 m or to the point where the numbers of cells in a distance category fell below 20, whichever was reached first. The values of these parameters were established by identifying natural breaks in the distance functions. The relationships between forest type similarity, measured by correlating membership values, and the slope and y -intercept of entropy-distance function were assessed through simple linear regression.

5. Results

5.1. Comparison

No map or representation of known accuracy was available against which to assess the accuracy of either the automated or the manual presettlement forest mapping method. Areas of disagreement between the fuzzy and discrete representations may be attributable to errors in either map. Similarly, areas of agreement or disagreement may simply result from wrong information on both representations. However, interpretations were made assuming that where the maps were in agreement there was a greater likelihood of correct classification in both cases.

Table 4 lists the overall agreement (PCC and kappa) between discrete type maps resulting from each method. The level of agreement based on all grid cells was comparatively low (about 66% agreement). By focusing on only those cells having the highest maximum membership values in the automated representation, the

Table 4. Agreement between forest type maps for varying levels of membership.

| Membership distribution cut-off* (%) | PCC | kappa |
|---|-------|-------|
| 100 | 0.660 | 0.559 |
| 75 | 0.737 | 0.648 |
| 50 | 0.807 | 0.734 |

*Percentage of cells with highest fuzzy membership values retained for analysis.

agreement increased substantially (to nearly 81 % for the one-half of cells with the highest membership values). As the strength of membership in the forest type of maximum membership increased, therefore, there was a greater likelihood that the mapped forest type was in agreement with the type indicated on the manual vegetation map. This result suggested that the information about the degree of type membership contained in the fuzzy representation was a good indicator of the uncertainty of the classification. A cell with a higher type membership value was more likely to be correct because there was less ambiguity in the set of trees surrounding the location, resulting in similar type definitions on both maps.

5.2. Factors affecting agreement

Given that the automated and manual representations were derived from the same data set, differences between the two representations resulted from differences in the methods. Because no actual ground truth information was available, the results can only suggest potential sources of uncertainty in the maps. Table 5 lists mean difference ($\bar{\Delta}$) values for each forest type, as identified on the discrete map. Note that a higher positive mean difference value indicated stronger agreement between the representations. The negative values indicated that, on average, the discrete map type had a membership value that was lower than at least one other type. Three forest types (conifer upland, conifer lowland, and conifer bog) had difference values less than or near zero, indicating that cells assigned to those types on the hand-drawn map were least likely to be in agreement and, therefore, most likely to be incorrectly mapped.

To explain why the three mixed conifer types were more difficult to map, note the correlations between the species membership values listed in table 1 (table 6). Most correlations between types were negative, indicating that very different species mixtures defined different forest types. Where correlations were positive, the forest

Table 5. Average difference between maps and entropy within continuous representation by type.

| Forest type | Mean difference (Δ) | Mean entropy (h) |
|-------------|---------------------------------|-------------------------|
| hwds | 0.26 | 0.70 |
| c upl | -0.06 | 0.87 |
| pin | 0.18 | 0.82 |
| j pine | 0.40 | 0.61 |
| c bog | 0.03 | 0.83 |
| c lowl | -0.00 | 0.86 |

Table 6. Correlations between species membership values for each pair of forest types.

| | hwds | c upl | pin | j pine | c bog | c lowl | Sum |
|--------|-------|-------|-------|--------|-------|--------|-------|
| hwds | | | | | | | -1.31 |
| c upl | 0.13 | | | | | | -0.55 |
| pin | -0.35 | -0.20 | | | | | -1.13 |
| j pine | -0.30 | -0.53 | -0.01 | | | | -1.23 |
| c bog | -0.36 | -0.03 | -0.22 | -0.25 | | | -0.69 |
| c lowl | -0.43 | 0.07 | -0.35 | -0.14 | 0.16 | | -0.69 |

type definitions were more confused. A higher negative mean value indicated a more unique type. The three forest types with the lowest mean difference values (i.e. those least likely to be in agreement) were also those with the lowest negative sum of correlations (i.e. those with the most confused composition definitions). This indicated that type definition was a constraint on the automated (and possibly the manual) mapping process. Forest type mapping was more successful for types that were well defined and distinguished from other types by composition.

Another indicator of the distinguishability of the forest types was the variance contribution (c) of the semi-variogram (figure 2 and table 3). The variance contribution is an indicator of the degree to which like values for a given variable tend to cluster. Where this value was lower, clustering tended to be weaker. The correlation (Pearson's r) between the variance contribution of a forest type's semi-variogram (table 3) and its mean difference (table 5) was 0.75, indicating that a strong spatial clustering in a pattern of forest type membership tended to improve the ability to automate the forest type identification process. Of course, a strong clustering tendency may also be related to a lack of similarity with other type definitions.

The mean entropy values (h) (table 5) indicated the degree to which each forest type dominated the others where that type had the maximum value. The mean entropy values of the types were strongly related to the level of agreement with the discrete map, measured by mean difference (Pearson r -value was -0.95). Therefore, the inverse of average entropy of a given type was a good surrogate measure for the likelihood of accurate representation.

The degree to which the fuzzy representation was in agreement with the crisp map for each stand was significantly (at $p < 0.01$), albeit weakly, related to both the area and the perimeter of the stand. The Pearson's r correlation for the log-linear regression relations of stand mean difference to area was 0.42 and to perimeter was 0.40 (figure 4). T -tests indicated that the group of stands ($n = 68$) with positive mean difference values had a significantly (at $p < 0.01$) higher average area (29 907 ha) and perimeter (42 608 m) than the group ($n = 142$) with negative mean differences (6370 ha and 12 260 m, respectively). The hypothesized relations between stand geometry and agreement between crisp and fuzzy representations therefore was confirmed. Smaller stands and/or stands with smaller perimeters were more difficult to reproduce using the automated generalization methods.

The manual interpretation included ancillary information about topographic

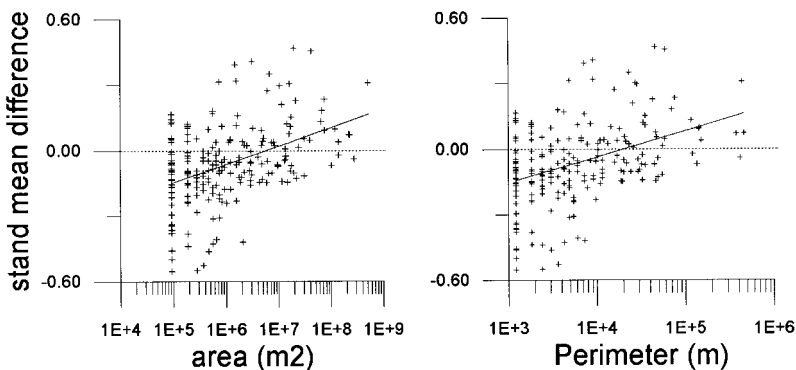


Figure 4. Relation between discrete-map polygon characteristics and mean difference with fuzzy representation (Δ).

setting in the map and was probably more sensitive to smaller distance changes in forest composition. These differences in method may have caused some of the observed differences in the mapped results, but this issue was not explored further. Co-kriging might be useful in the future to introduce independent environmental data into automated presettlement forest type mapping.

5.3. Classification ambiguity and boundary vagueness

Figure 5 is a map of entropy $[h(x)]$ with the manually interpreted boundaries overlaid on it. To assess classification ambiguity at the boundaries of the polygons, entropy was averaged and plotted for categories of distance to the nearest manually interpreted boundary. The entropy curves were plotted in grid cell increments of distance. The heavy solid line in figure 6 is the mean entropy, the lighter solid lines are the mean entropy plus and minus one standard deviation at each distance interval. The dotted line in figure 6 represents the number of grid cells used to calculate the values on the solid lines. Beyond 2100 m the number of grid cells

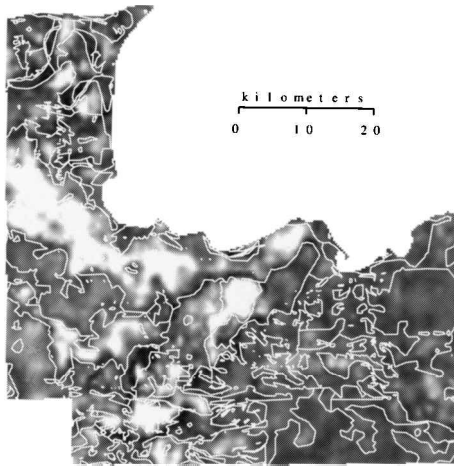


Figure 5. Shaded map of classification entropy with crisp-map boundaries overlaid. Darker shades represent higher entropy values.

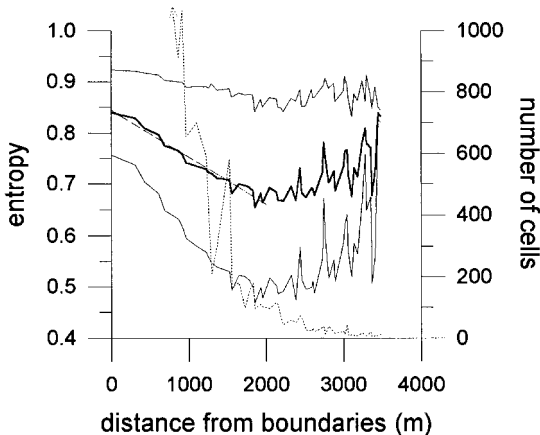


Figure 6. Average relation between distance to discrete map boundary lines and entropy.

available for calculating the average dropped below 95 and those pixels fall in no more than 10 different polygons. The increasing entropy with distance, at distances beyond 2100 m, was likely related to the fact that the number of grid cells used to calculate the average entropy was small and that a few large and compact polygons dominated at those distances. The most dominant of those polygons belonged to the conifer upland and conifer lowland types, each of which tended to have higher entropy because of their definitional similarities (table 6).

A best-fit linear function was calculated to model the observed relation between mean entropy and distance from boundaries to a distance of 2100 m (dashed line in figure 6). Interpretation beyond this distance is not warranted because the values are dominated by a limited number of forest types. The coefficients for the equation of the line are listed at the bottom of table 7. The Pearson correlation coefficient for the relation was 0.98, and the slope was significantly different from zero ($p < 0.01$).

As hypothesized, mean entropy tended to decrease with distance from the boundary, at least for cells less than 2100 m from a boundary. This general relation was true for both the plus and minus one standard deviation lines as well and it provides empirical support for a hypothesis that has been difficult to represent and model (i.e. that classification ambiguity is related to location relative to boundaries). The fact that average entropy increases at distances greater than 2100 m may indicate some significant component of the process, but conclusions based on that observation would be spurious because there are too few polygons of only three forest types contributing to the entropy averages at those distances.

Whereas the graph in figure 6 represents the average relation between entropy and boundary location, the relation was expected to vary for different boundaries depending on the similarity of the definitions of forest types on either side. The linear functions describing the relations between entropy and distance to the nearest boundary for each boundary type (i.e. separating different pairs of forest types) are given in table 7. Ten of the eleven functions had a Pearson r -value greater than 0.90; all values were greater than 0.85. Although the relations between the similarity of forest type definitions (i.e. classification ambiguity) and the y -intercept of the linear distance-entropy functions was not significant, forest type similarity (i.e. correlation coefficients in table 6) was significantly related ($p < 0.01$) to variations in the slope values of the distance-entropy functions by boundary type (figure 7). Where a

Table 7. Linear regression coefficients for distance-entropy relations by boundary type.

| Boundary type | y -intercept | Slope |
|----------------|----------------|-----------------------|
| hwds-cupl | 0.85 | -6.0×10^{-5} |
| hwds-pines | 0.91 | -1.4×10^{-4} |
| hwds-cbog | 0.84 | -1.7×10^{-4} |
| cupl-pines | 0.90 | -1.6×10^{-4} |
| cupl-jpine | 0.92 | -1.5×10^{-4} |
| cupl-cbog | 0.87 | -4.7×10^{-5} |
| cupl-clowl | 0.86 | -1.6×10^{-5} |
| pines-jpine | 0.74 | -1.1×10^{-4} |
| pines-cbog | 0.86 | -8.4×10^{-5} |
| jpine-cbog | 0.72 | -5.9×10^{-5} |
| cbog-clowl | 0.85 | -2.7×10^{-5} |
| all boundaries | 0.85 | -7.9×10^{-5} |

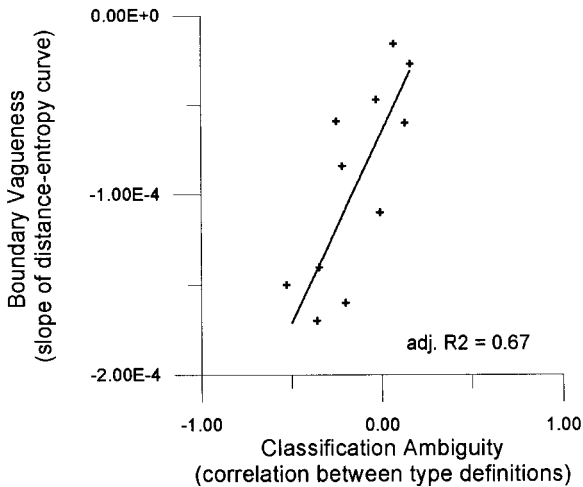


Figure 7. Scatterplot of the relation between boundary vagueness and classification ambiguity with best-fit linear function.

boundary separated a pair of types with more different definitions, at distances less than 1000m, mean entropy tended to decrease with distance from the boundary more sharply than where the forest type definitions were more similar. Therefore, increases in attribute ambiguity tended to result in increases in the vagueness of the boundary locations.

6. Discussion

The two most salient arguments for the use of a continuous model for mapping vegetation patterns are: (1) some vegetation distributions exhibit naturally continuous variation, with gradual transitions between vegetation types; and (2) data used for mapping are never perfect; they represent either a sampling of vegetation (e.g. the tree sample used in this paper) or represent variables that serve as surrogates for vegetation information (e.g. spectral information from remote sensing). Each of these arguments supports the contention that the locations of boundaries drawn on a forest type map are not 100% certain. Although, the analysis presented does not permit direct conclusions about the relative importance of these two factors, it does provide insight into the mapping process as applied to GLO survey notes and an alternative to the more traditional, discrete mapping methods.

Type maps of all kinds (e.g. soils types and forest types) are made with the implicit assumption that every location can be mapped to one type exclusively. Forest type maps might be drawn by looking at the spatial configurations of species or at reflectance patterns in an aerial photograph. Aside from the (slightly arrogant and usually erroneous) assumption that a person or automated method creating a boundary will always be able to distinguish perfectly between members (i.e. locations) belonging to sets (i.e. forest types), this approach also assumes perfect information. The assumption of perfect information implies perfect positional and attribute information and that all data are available (e.g. that all trees are known) or that other information (e.g. topographic maps) can substitute perfectly for detail in the primary data. If the data are incomplete (i.e. a sample) or uncertain to some degree

(and when are they not?) then the locations of the boundaries will be vague to some degree.

Although mappers have always been aware of gradual transitions and ecotones, the options available for representing them were limited. To address the issue of gradual transitions, Kùchler (1955) suggested the use of 'transition strips' between two vegetation types that grade into one another. Alternatively, Kùchler (1988, p. 108) suggested that one might wish to represent transition zones as 'broken lines, zigzag lines, or by an interpenetration of two contiguous vegetation types'. Blakemore (1984) suggested giving width to lines proportional to the positional uncertainty in the line (called epsilon bands). Although a more continuous representation like that in figure 3 would be preferred because it contains more information, epsilon bands or other cartographic symbols can be used to convey uncertainty information in a more traditional map form. In order to determine the appropriate distance, or width of the epsilon band, however, some understanding of the uncertainty in the boundary location is needed. Edwards and Lowell (1996) calculated epsilon band widths ('fuzzy boundary width' was their term) based on an analysis of the variability in line locations identified by multiple photointerpreters. Although more empirical work would be needed to find the appropriate cut-offs, the ambiguity-distance relations presented here (e.g. figure 6 and table 7) might also be useful for identifying epsilon band widths (i.e. by selecting a level of entropy and use, as the epsilon distance, the distance within which the average entropy is higher).

Edwards and Lowell (1996) concluded that the vagueness of boundaries was related to image textural similarities between and variability within vegetation types in simulated aerial photographs. Similarly, this analysis suggests that, when forest types are to be delineated using the sample of tree species contained in the GLO survey records, boundary vagueness is related to the similarity between the definitions of adjacent forest types. Therefore, ambiguity in the forest type definitions tended to result in more uncertainty in grid cell classification and in boundary location (i.e. vagueness). In applying this to epsilon band mapping, different boundary types might be assigned different epsilon distances according to the similarity of their adjacent forest type definitions or using their ambiguity-distance relation (table 7). Also, it is reasonable to assume that the ambiguity may fall off more rapidly on one side of a boundary than the other, depending on which types are on either side. Further work should seek explanations for such patterns that might aid in the delineation of anisotropic epsilon bands. Again, one cause of the uncertain boundary location is likely that vegetation types with similar definitions tend to have more gradual transitions or that identifying the boundary locations is more difficult. Probably there is an element of truth in both explanations.

Other processes that cause vagueness in boundaries need to be explored. For example, gradual environmental gradients might result in a gradual change between two very different vegetation types, as occurs along the arctic and alpine treelines. The relations between the slope of an environmental gradient and boundary vagueness and appropriate epsilon band widths needs to be explored.

The automation of manual tasks has several advantages. This analysis assessed one such automation (i.e. creating generalized polygons from point data). Primary among the advantages of automation is the ability to alter the classification scheme to produce alternative maps from the same data very quickly. This makes application of map data and a decision-making environment more flexible and adaptive. Furthermore, users of the generalized maps always have access to digital versions of

the original data on which they are based, which should help users to make their own decisions about data quality in the generalized representation. Also, the criteria for type delineation are much more consistent across the map. Therefore, assessment of the fitness of the map for a given purpose is not affected by variable mapping methods across the map.

The effectiveness of the map generalization process (manual or automated) for producing forest types from tree data recorded in the General Land Office survey notes was affected by ambiguity in forest type definitions and strength and scale of spatial clustering in the forest types. The analysis shows that, where a forest type was not well differentiated from the others and/or where the degree of spatial clustering was weak, the fit of locations into a forest type tended to be more ambiguous and boundaries with similar forest types tended to be more vague. Similarly, individual stands that were small were more difficult to identify consistently than larger and/or more compact stands.

In order to generalize the approach to vegetation mapping presented here, explicit recognition and incorporation of boundaries that are, indeed, discrete is needed. Although literature on the continuum concept suggests gradual vegetation change along environmental gradients, there is a clear recognition that some environmental gradients, within certain ranges of spatial scales, change quite rapidly across space leading to abrupt spatial vegetation changes. At the scale of the data used in this paper, abrupt changes in substrate type or disturbance history, for example, may have caused situations where crisp boundaries are needed. Where such rapid changes are detectable, the methods presented here must be improved to provide the ability to anchor such discrete boundaries prior to the spatial generalization process. Hutchinson's (1989) interpolation procedures, which were designed for terrain modeling, might be used for this purpose. His drainage enforcement algorithm is specifically designed to conform a continuous surface to discrete breaks where they are known.

Acknowledgments

The data used in this paper were collected from the Michigan Archives and digitized by Mr David Price and Dr Kurt Pregitzer. The author gratefully acknowledges their willingness to share the data for use in this study. This project was funded in part by the Center for Remote Sensing and an All University Research Grant from the College of Social Science, both at Michigan State University. The technical assistance of Jue Hu is also appreciated.

References

- ALBERT, D. A., 1995, Regional Landscape Ecosystems of Michigan, Minnesota, and Wisconsin: A Working Map and Classification. General Technical Report NC-178. (St. Paul, MN: US Forest Service, North Central Forest Experiment Station).
- ANDERSON, R. C., and ANDERSON, M. R., 1975, The presettlement vegetation of Williamson County, Illinois. *Castanea*, **40**, 345–363.
- AUSTIN, M. P., 1985, Continuum concept, ordination methods, and niche theory. *Annual Review of Ecology and Systematics*, **16**, 39–61.
- BARRETT, L. R., LIEBENS, J., BROWN, D. G., SCHAEZL, R. J., ZUWERINK, P., CATE, T. W., and NOLAN, D. S., 1995, Relationships between soils and presettlement vegetation in Baraga County, Michigan. *American Midland Naturalist*, **134**, 94–115.
- BEZDEK, J. C., 1984, FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, **10**, 191–203.

- BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W., 1975, *Discrete Multivariate Analysis: Theory and Practice* (Cambridge, MA: The MIT Press).
- BLEWETT, M. B., and POTZGER, J. E., 1950, The forest primeval of Marion and Johnson counties, Indiana, in 1819. *Butler University Botanical Studies*, **10**, 40–52.
- BLAKEMORE, M., 1984, Generalisation and error in spatial data bases. *Cartographica*, **21**, 131–139.
- BRAUN-BLANQUET, J., 1964, *Pflanzensoziologie; Grundz uge der Vegetationskunde*, 3rd edition (New York: Springer-Verlag).
- BREWER, L. G., HODLER, T. W., and RAUP, H. A., 1984, *Presettlement Vegetation of Southwestern Michigan*, map (Kalamazoo: Western Michigan University).
- BROWN, D. G., Mapping historical forest types in Baraga County, Michigan as fuzzy sets. *Plant Ecology*, in press.
- BUORDO, E. A., 1956, A review of the General Land Office survey and of its use in quantitative studies of former forests. *Ecology*, **37**, 754–768.
- BURNS, R. M., and HONKALA, B. H., Technical coordinates, 1990, *Silvics of North America*, Agriculture Handbook 654. (Washington, DC: US Government Printing Office).
- BURROUGH, P. A., 1989, Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science*, **40**, 477–492.
- BURT, J. S., 1985, *They Left Their Mark: William Austin Burt and his Sons, Surveyors of the Public Domain* (Rancho Cordova, CA: Landmark Enterprises).
- BUTTENFIELD, B. P., 1993, Representing data quality. *Cartographica*, **30**, 1–7.
- CAMPBELL, J. B., 1978, Locating boundaries between mapping units. *Mathematical Geology*, **10**, 289–299.
- CANHAM, C. D., and LOUCKS, O. L., 1984, Catastrophic windthrow in the presettlement forests of Wisconsin. *Ecology*, **65**, 803–809.
- CATANA, A. J., JR., 1967, Forests of the Harvey N. Ott Preserve. *American Midland Naturalist*, **78**, 496–507.
- CLEMENTS, F. E., 1936, Nature and structure of the climax. *Journal of Ecology*, **24**, 252–284.
- CONGALTON, R. G., 1991, A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, **37**, 35–46.
- COTTAM, G., 1949, The phytosociology of an oak woods in southwestern Wisconsin. Ph.D. Thesis (Madison: University of Wisconsin).
- COWELL, C. M., 1995, Presettlement piedmont forests: Patterns of composition and disturbance in Central Georgia. *Annals, Association of American Geographers*, **85**, 65–83.
- CRANKSHAW, W. B., QADIR, S. A., and LINDSEY, A. A., 1965, Edaphic controls of tree species in presettlement Indiana. *Ecology*, **46**, 688–698.
- DANSERAU, P., 1968, The continuum concept of vegetation: responses. *Botanical Review*, **34**, 253–332.
- DELCOURT, H. R., 1976, Presettlement vegetation of the north of Red River Land district, Louisiana. *Castanea*, **41**, 122–139.
- DEUTSCH, C. V., and JOURNAL, A. G., 1992, *GSLIB: Geostatistical Software Library and User's Guide* (New York: Oxford University Press).
- DORNEY, J. R., 1981, The impact of Native Americans on presettlement vegetation in southeastern Wisconsin. *Transactions, Wisconsin Academy of Science, Arts and Letters*, **69**, 26–36.
- EDWARDS, G., and LOWELL, K. E., 1996, Modeling uncertainty in photointerpreted boundaries. *Photogrammetric Engineering and Remote Sensing*, **62**, 337–391.
- EYRE, F. H., editor, 1980, *Forest Cover Types of the United States and Canada* (Washington, DC: Society of American Foresters).
- FINLEY, R. W., 1976, *Original Vegetation Cover of Wisconsin from US General Land Office Notes* (map) (St. Paul, MN: USDA Forest Service, North Central Forest Experiment Station).
- FISHER, P. F., 1994, Probable and fuzzy models of the viewshed operation. In *Innovations in GIS*, edited by M. Worboys (London: Taylor & Francis), pp. 161–175.
- GLAVAC, V., GRILLENBERGER, C., HAKES, W., and ZIEZOLD, H., 1992, On the nature of vegetation boundaries, undisturbed flood plain forest communities as an example—a contribution to the continuum/discontinuum controversy. *Vegetatio*, **101**, 123–144.
- GLEASON, H. A., 1917, The structure and development of the plant association. *Bulletin of the Torrey Botanical Club*, **44**, 463–481.

- GLEASON, H. A., 1939, The individualistic concept of the plant association. *American Midland Naturalist*, **21**, 92–110.
- GOPAL, S., and WOODCOCK, C., 1994, Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, **60**, 181–188.
- GOODCHILD, M. F., BUTTENFIELD, B., and WOOD, J., 1994a, Chapter 15: Introduction to visualizing data validity. In *Visualization in Geographical Information Systems*, edited by H. M. Hearnshaw and D. J. Unwin (New York: John Wiley & Sons), pp. 141–149.
- GOODCHILD, M. F., CHIH-CHANG, L., and LEUNG, Y., 1994 b, Chapter 7: Visualizing fuzzy maps. In *Visualization in Geographical Information Systems*, edited by H. M. Hearnshaw and D. J. Unwin (New York: John Wiley & Sons), pp. 158–167.
- HUTCHINSON, M. F., 1989, A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*, **106**, 211–232.
- ISAAKS, E. H., and SRIVASTAVA, R. M., 1989, *An Introduction to Applied Geostatistics* (Oxford: Oxford University Press).
- KÜCHLER, A. W., 1955, Comprehensive method of mapping vegetation. *Annals, Association of American Geographers*, **45**, 404–415.
- KÜCHLER, A. W., 1988, Chapter 9: Boundaries, transitions and continua. In *Vegetation Mapping*, edited by A. W. Küchler and I. S. Zonneveld (Dordrecht: Kluwer Academic Publishers).
- KÜCHLER, A. W., and ZONNEVELD, I. S. (editors), 1988, *Vegetation Mapping* (Dordrecht: Kluwer Academic Publishers).
- LEITNER, L. A., DUNN, C. P., GUNTENSPERGEN, G. R., STEARNS, F., and SHARPE, D. M., 1991, Effects of site, landscape features, and fire regime on vegetation patterns in presettlement southern Wisconsin. *Landscape Ecology*, **5**, 203–217.
- LOWELL, K., 1994, Initial studies of fuzzy surface-based cartographic representations of forests. In *Proceedings of an International Symposium on the Spatial Accuracy of Natural Resource Data Bases, Williamsburg, Va.*, edited by R. G. Congalton (Bethesda, MD: American Society for Photogrammetry and Remote Sensing), pp. 168–177.
- MALANSON, G. P., 1985, Spatial autocorrelation and distributions of plant species on environmental gradients. *Oikos*, **45**, 278–280.
- MARSHNER, F. J., 1975, *The original vegetation of Minnesota* (map) (St Paul, MN: US Department of Agriculture, Forest Service, North Central Forest Experiment Station).
- MCGRANAGHAN, M., 1993, A cartographic view of spatial data quality. *Cartographica*, **30**, 8–19.
- MCINTOSH, R. P., 1967, The continuum concept of vegetation. *Botanical Review*, **33**, 130–187.
- NEUENSCHWANDER, H. E., 1957, The vegetation of Dodge County, Wisconsin. *Transactions, Wisconsin Academy of Science, Arts and Letters*, **46**, 233–254.
- OLIVER, M., WEBSTER, R., and GERRARD, J., 1989 a, Geostatistics in physical geography. Part 1: Theory. *Transactions, Institute of British Geographers*, **14**, 259–269.
- OLIVER, M., WEBSTER, R., and GERRARD, J., 1989 b, Geostatistics in physical geography. Part 2: Applications. *Transactions, Institute of British Geographers*, **14**, 270–286.
- PANNATIER, Y., 1996, *VARIOWIN: Software for Spatial Data Analysis in 2D* (New York: Springer-Verlag).
- PRICE, D. L., 1994, An Ecological Study of the Composition, Structure and Disturbance Regimes of the Pre-European Settlement Forests of Western Chippewa County, Michigan. M.S. Thesis (East Lansing: Department of Forestry, Michigan State University).
- ROSSI, R. E., MULLA, D. J., JOURNAL, A. G., and FRANZ, E. H., 1992, Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, **62**, 277–314.
- SHANKS, R. E., 1953, Forest composition and species association in the beech-maple forest region of Western Ohio. *Ecology*, **34**, 455–466.
- SHIMWELL, D. W., 1971, *The Description and Classification of Vegetation* (London: Sidgwick & Jackson).
- SHIPLEY, B., and KEDDY, P. A., 1987, The individualistic and community-unit concepts as falsifiable hypotheses. *Vegetation*, **69**, 47–55.
- SICCAMA, T. G., 1971, Presettlement and present forest vegetation in northern Vermont with special reference to Chittenden County. *American Midland Naturalist*, **85**, 153–172.
- TIMONEY, K. P., LAROI, G. H., and DALE, M. R. T., 1993, Subarctic forest-tundra vegetation gradients: the sigmoid wave hypothesis. *Journal of Vegetation Science*, **4**, 387–394.

- TÜXEN, R., 1954, Über die räumliche, durch Relief und Gestein bedingte Ordnung der natürlichen Waldgesellschaften am nördlichen Rande des Harzes. *Vegetatio*, **5–6**, 454–478.
- VEATCH, J. O., 1959, *Presettlement Forest in Michigan* (map) (East Lansing, MI: Department of Resource Development, Michigan State University).
- WHITE, C. A., 1984, *A History of the Rectangular Survey System* (Washington, D.C.: Bureau of Land Management, Government Printing Office).
- WHITNEY, G. D., 1992, *Soil Survey of Chippewa County, Michigan* (Washington, DC: USDA Soil Conservation Service, United States Government Printing Office).
- WHITNEY, G. G., 1986, Relation of Michigan's presettlement pine forests to substrate and disturbance history. *Ecology*, **67**, 1548–1559.
- ZADEH, L., 1965, Fuzzy sets. *Information and Control*, **8**, 338–353.