

# Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues

Goodchild, Michael F.

*National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA; Tel. +1 805 893 8049; Fax +1 805 893 7095; E-mail GOOD@NCGIA.UCSB.EDU*

**Abstract.** GIS and remote sensing have emerged as distinct spatial data handling technologies with their own methods of data representation and analysis. Combining them as tools to support vegetation analysis and modeling thus presents a number of challenges. The paper begins by describing the major data sources, applications, and software characteristics of each technology, and then compares them within a consistent terminological framework that emphasizes the digital representation of continuously varying spatial data. Because the spatial continuum can be discretized in many different ways, and because each can only approximate the truth, both GIS and remote sensing are subject to error and uncertainty. Integration, and subsequent analysis and modeling, require that explicit attention be directed to uncertainty. The paper reviews the models of error that have been developed in recent years for spatial data and examines their use in the interface between GIS and remote sensing. The paper looks at the functional requirements of modeling, and includes discussion of error propagation.

**Keywords:** Error model; Error propagation; GIS; Spatial data; Uncertainty.

**Abbreviation:** GIS = Geographical Information Systems; NDVI = Normalized Difference Vegetation Index.

## Introduction

A variety of technologies for collecting, handling, and analyzing spatial data have emerged in the past three decades. If attention is limited to those that deal strictly with geographical data, rather than the more generally defined spatial data, then the list would have to include GPS, the global positioning system based on analysis of signals from a constellation of orbiting satellites; electronic surveying technology as typified by the modern 'total station'; remote sensing; and geographic information systems (GIS). Although all involve the use of sophisticated electronic technology, the last two raise interlocking issues that go well beyond those of the technology per se, particularly for scientific applications in areas such as vegetation science, and are the focus of this paper.

Although remote sensing can be broadly defined as measurement at a distance, for most purposes it implies the imaging of the surface of the Earth from an airborne or space-borne sensor, using some appropriate part of the electromagnetic spectrum. The technical issues of sensor design, orbital corrections, cloud, and noise reduction are not of concern here; for the purposes of this paper, it is assumed that imaging systems produce samplings of two-dimensional univariate or multivariate fields of measured spectral response in one or more known parts of the electromagnetic spectrum. Remote sensing software is designed to process this raw data into forms that are useful for a range of purposes, from simple mapping and monitoring to management and modeling.

A geographic information system is a much broader and more nebulous concept - a system for the input, storage, manipulation, and output of geographically referenced data. Although there is wide acceptance of this general definition, the actual specification of a GIS can vary widely depending on the assumptions that its designers are willing to make about its uses. For example, a GIS designed to process only data derived by remote sensing would be very different from one designed to process only data derived from maps. This paper will not attempt to provide a general introduction to GIS, since the topic is already addressed comprehensively in many texts (for general overviews see Burrough 1986; Maguire et al. 1991; Star & Estes 1991). Goodchild et al. (1993) include several chapters on the use of GIS for ecological modeling, and Haines-Young et al. (1993) discuss applications of GIS in landscape ecology.

The purpose of this paper is to develop a general conceptual framework for geographical data, and vegetation mapping in particular, and to embed both GIS and remote sensing within it, thus providing a theoretical and methodological basis for integration. Because both technologies create approximate representations of geographical distributions, issues of error and uncertainty are critical to successful integration. Each field has developed its own, apparently incompatible, approach to error description, and the paper develops a

general approach for overcoming this significant impediment within the general conceptual framework. Throughout the paper, the examples and illustrations are drawn from the area of vegetation mapping.

## Geographic data models

### *The nature of geographical data*

At the most fundamental level, geographical data can be defined as a collection of facts about places. The atom of spatial data is the tuple  $\langle x, y, z \rangle$ , where  $x$  and  $y$  define a place, and  $z$  some fact about the place, such as vegetation cover class. A convenient way to define the difference between spatial and geographical data is to insist that for geographical data, the coordinates  $(x, y)$  be defined by some system of measurement on the Earth, such as latitude/longitude, or the UTM (Universal Transverse Mercator) coordinate system.

Current measurement technologies place fairly coarse limits on one's ability to determine geographical position accurately. A hand-held GPS receiver, for example, can determine absolute position on the Earth to perhaps 1000 m<sup>2</sup>, or approximately the area of one arc-second of latitude and longitude, outside the polar regions. The problems of approximating the complex shape of the Earth with simple mathematical functions also place limits on absolute positioning. On the other hand, it is possible to establish relative positions with millimeter accuracy, even over distances of hundreds of kilometers, using modern surveying techniques.

For many types of geographical facts, it is appropriate to conceive of a measurement  $z$  being made at a point, or infinitely small area, even though the accuracy of that point's location may be limited. For example, ground elevation can be measured at a point, as can air temperature, rainfall, or atmospheric pressure. Although a rain gauge will average rainfall over a discrete radius, that radius will be much less than the positional error in the location of the rain gauge, and irrelevant in any likely analysis.

For other types of geographical facts, the definition of the value at a point requires an observation over some significant radius around the point. Vegetation cover classes, for example, can only be defined over discrete areas that substantially exceed the area covered by any one individual organism. This spatial resolution of the fact should not be confused with the positional accuracy of the point, or with measures of the size of the study area. In a typical example, the positional accuracy of the observation might be 1000 m<sup>2</sup> using GPS, while the spatial resolution of vegetation cover class might be 1 ha (10 000 m<sup>2</sup>), implying that it is necessary to ob-

serve an area of 1 ha to determine the cover class recorded at its central point. Unfortunately, neither measure is explicitly defined in much vegetation mapping practice.

In any discussion of positional accuracy, it is common to mix units - some parameters may be expressed as areas, some as radii, and some as diameters. To avoid confusion, area measures will be used consistently throughout the paper. Thus, the size of a pixel, minimum area of a polygon, and positional accuracy of a point will all be expressed in area measure. In the case of positional accuracy of a point, the measure will express the area of a circle centered on the point within which the true location is assumed to lie with some determined probability.

In this conceptual framework, it is possible to determine the value of any fact  $z$  at any location  $(x, y)$ . For some types of geographical data, such as ground elevation, the value at any location is potentially unique, and thus an infinite number of tuples would be required to capture geographical variation over a finite area completely. In practice, measurement systems determine values only at a discrete number of locations, or adopt other strategies to limit the volume of data that must be handled.

### *Fields and entities*

Two distinct mental models underlie the various strategies that have been devised for the representation of geographical variation (Goodchild 1992). In the field view, geographical variation can be conceived as a set of distinct variables  $\{z_1, z_2, \dots, z_n\}$ , each with a precise definition that allows it to be measured at any point  $(x, y)$ . Mathematically, geography is a multivariate or vector field. Some variables may be measured on continuous (interval or ratio) scales, while others, such as vegetation class, may be discrete (nominal or ordinal). Because all variables show some degree of spatial dependence or autocorrelation, continuous variables may be differentiable, although ridge lines, cliffs, and faults are obvious sources of exceptions. Similarly, the values of discrete variables may be constant over substantial contiguous areas.

While the field view may be appropriate for rigorous definition and precise measurement, it is generally inadequate for interpretation and many forms of modeling, where it is replaced by the entity view (the terms field and entity have many precise and imprecise synonyms in the geographical data handling field - entity is often replaced by feature or object). In this mental model, geography is conceived as an empty space littered with various kinds of objects, with associated characteristics. For example, interpretation of a raw field of ground

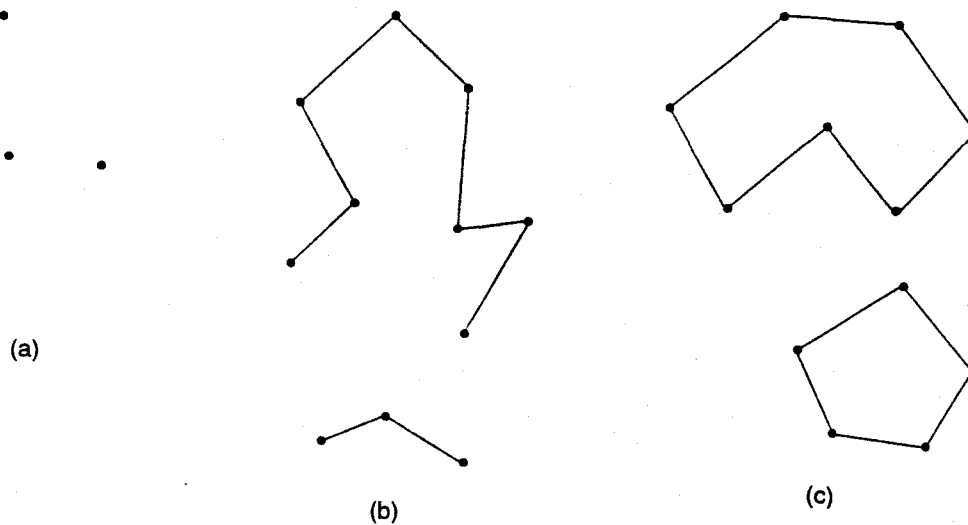


Fig. 1. Digital representations of (a) points, (b) lines as polylines, and (c) areas as polygons.

elevations would yield such topographic objects as mountain, valley, watershed, ridge, pass, or saddle. Interpretation of a raw field of atmospheric pressure might yield high, low, trough, or front. Many models address the behavior of these discrete entities, rather than the continuous fields from which they may have been derived.

In some cases, the derivation of entities may be well-defined. However, not all maxima in a continuous field of ground elevation are significant enough to be called summits - the rule base of interpretation is likely to be complex. Definitions of entities often fail to give them precise footprints (the summit of a mountain is often better defined than its geographic limits; there are no precisely defined limits to the Atlantic Ocean) and entities of the same class may overlap (the Gulf of Mexico may be considered part of the Atlantic Ocean) or have hierarchical relationships with each other.

The field and entity views are mental models of geographical variation that allow the mind to comprehend what is potentially an infinitely complex world. In everyday life the entity view is dominant, and reflected in the construction of topographic maps, or the giving of navigational directions. The State of Minnesota is described as having 10 000 lakes, not a continuous field of 'lakeness' that can be evaluated objectively at any point on the surface of the state. In science, on the other hand, both views are in common use; atmospheric pressure is captured and recorded by sampling a field, and the field view is reflected in the finite difference simulation models used to predict future conditions, but weather forecasts provide the public with predictions of the behavior of highs, lows, and fronts as discrete entities

moving in an otherwise empty space. In vegetation mapping, the field view is clearly dominant, vegetation being conceived as having some unique class or description at every point in the plane. On the other hand, modeling in vegetation science may be more concerned with the behavior of discrete objects distributed in an otherwise empty space, and thus more consistent with the entity view.

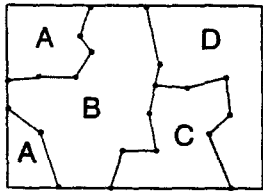
#### *Digital representation*

Although fields and entities are useful mental models of real geography, they both contain a potentially infinite amount of information. The digital representation of entities is relatively straightforward (Fig. 1). A finite set of point entities (Fig. 1a) can be represented with no information loss as a series of tuples  $\langle x, y, z \rangle$ , and the numerical precision of a digital representation normally exceeds both the positional accuracy of the coordinates and the measurement accuracy of  $z$ . A line is conventionally represented as a polyline (Fig. 1b), an ordered set of points assumed to be connected by straight line segments. Although accuracy can be improved by increasing the density of points, the polyline convention has two significant disadvantages: the length of a polyline is normally less than the curve it approximates; and it is difficult to estimate tangents or perpendiculars from a polyline representation. Area entities are conventionally represented as polygons (Fig. 1c), with similar problems regarding the measurement of perimeters, perpendiculars, and tangents.

The representation of fields offers more alternatives,

A	A	A	B	D	D	D
A	A	B	B	D	D	D
B	B	B	B	C	C	D
A	B	B	B	C	C	D
A	B	B	C	C	C	D

(a)



(b)

Fig. 2. The (a) raster and (b) polygon data models for a nominal-scaled field such as vegetation class. Note the largely independent approximations inherent in each representation.

and at the same time raises more problems. Six methods are in common use (Goodchild 1992):

- randomly located sample points;
- sample points in a regular rectangular array;
- a regular rectangular array of cells, with some aggregate value of the field recorded for each cell [the 'raster' model in subsequent discussion (see Fig. 2a)];
- a polyline representation of selected isolines (contours);
- an irregular array of triangles, with the value of the field recorded at every vertex, and the field assumed to vary linearly within each triangle (the TIN, or triangulated irregular network model);
- an irregular array of polygons, with the value of the field assumed constant within each polygon [the 'polygon' model in subsequent discussion (see Fig. 2b)].

Although many other representations are possible, such as arrays of hexagonal cells, they have not been exploited to any great extent in geographical data modeling to date.

Each of the six field models creates sets of points, polylines, or polygons, and thus at this superficial level the distinction between field and entity becomes moot within a digital representation. The term 'vector' is used to refer ambiguously to all three types of entity models, and all but two of the field models, while the term 'raster' refers to the second and third field models. Moreover, any set of non-overlapping entities can be

represented as a field whose value is the entity identifier at locations inside an entity, and zero elsewhere.

However, the distinction between entities and fields is critical when the behavior or processing of points, polylines, or polygons is considered. The polygons of a field representation must collectively exhaust the space, and cannot overlap (Fig. 2b); one polygon cannot be dragged over another during editing, and common boundaries can be moved only as long as they do not intersect other common boundaries. The polylines of a contour representation cannot cross, and adjacent contours must have adjacent values. At a more sophisticated level, the common process of interpolation between a sample of points representing a field, exemplified by the contouring operation, makes no sense whatever if the points are conceived as entities littering an empty space.

Each of the six field models has its own characteristics, and some of these are important in any effort to integrate GIS and remote sensing. First, only two of the field models are appropriate for discrete (categorical, multinomial) fields: the regular array of cells, termed the 'raster' model for the purposes of this paper, and the polygon model (Fig. 2). In both cases the value assigned to each object in the representation, respectively cell and polygon, may be assumed to be the most common or modal value of the field within the limits of the object.

Second, only three of the field models include a representation of the value of the field at every point in the plane: the regular array of cells, the polygon model, and the TIN (triangulated irregular network). In the remaining cases (irregular points, regular array of points, and digitized contours) the model must be coupled with some set of rules about interpolation if values are to be determinable at any point in the plane. From an object oriented perspective, it would be desirable to encapsulate the method of interpolation with the data model, but in practice this is rarely done; instead, selection of an appropriate rule or rules is left to the user.

The next major section of the paper discusses the traditions of remote sensing and GIS within this conceptual framework. Because vegetation is traditionally mapped as a nominal-scaled (multinomial) field of vegetation class, the emphasis will be on the raster and polygon models of fields, illustrated in Fig. 2. Note, however, that ratio-scaled fields of vegetation occur in remote sensing in examples such as the Normalized Difference Vegetation Index (NDVI).

## Remote sensing and GIS

### *Remote sensing*

Although the technical problems of platform, instrument, and preprocessing are important, it is assumed for the purposes of this paper that remote sensing begins when they have been overcome, with a so-called corrected image. In its digital version, the image consists of a representation of a multivariate field, each variable capturing the radiation from the Earth's surface in some part of the electromagnetic spectrum. Each field is represented using a raster data model, the value of spectral response in each cell or pixel being approximately the mean value within the cell's footprint on the Earth's surface. In relative terms, the positional accuracy of each pixel is a small fraction of the pixel's area, but in absolute terms, positional accuracy is likely to be at least nine pixels (that is, the true location of the pixel's central point may be anywhere within a 3 by 3 array of pixels surrounding the point). When a scene contains well-defined objects that can be identified on the ground, it may be possible to register the image to a fraction of a pixel; but if no well-defined objects are visible in the scene, it may be difficult to find a pixel's footprint on the ground to better than an area of nine pixels.

When a scene is perfectly registered to the ground, the error inherent in this digital representation is simply the difference between the true value of the field at a point, and the mean value of the containing pixel. However, in practice, with imperfect registration, error is the difference between the true value at a point, and the mean value of the pixel in which the point appears to fall, which may be as much as two pixels away from the point. In both cases, error depends on the local variability of the field; if the field is constant over distances of as much as two pixels, error can be zero.

Much of the research effort in remote sensing over the past three decades (for general reviews of remote sensing and its applications see Colwell 1983; Jensen 1986; Lillesand & Kiefer 1987) has gone into methods of classification - the conversion of a multivariate field of spectral response to a univariate field of categories. Ideally, classification attempts to recover the true class present at every point in the plane; in practice, it recovers the dominant class in each pixel. For example, in agricultural applications it is possible to conceive of the truth as the identity of the crop being grown in every agricultural field. In practice, however, classification is applied in areas where there is no such true class. For example, in vegetation mapping the definition of a class at a point will involve scanning a finite area around the point (see the previous discussion of spatial resolution, and note the implicit assumption that the spatial resolu-

tion of the classification is much less than the pixel size; if it is not, it is in principle impossible to define the class to which a pixel belongs). It may involve substantial subjectivity, such that two randomly chosen but trained observers would not always agree on the class to be assigned at a point. This uncertainty of definition merges with other sources of uncertainty, such as the mixed pixel that overlaps two well-defined classes, or imperfect spectral discrimination between classes.

Even in cases where it is possible to conceive of true classes existing on the ground, as in the agricultural example, there will still be substantial spectral variation within the signals obtained from each class. Discrimination between classes is thus imperfect, and the result may be a 'salt and pepper' effect where individual pixels form inclusions within otherwise uniform patches of the same class. To remove such noise, it is common to apply a filter after classification, often using an array of 3 by 3 pixels; the central value is often replaced by the modal class of the array. In terms of the previous discussion, this amounts to a coarsening of spatial resolution from approximately one to ca. nine pixels (expressed in area measure), although linear features of less than three pixels width will clearly survive the filter.

### *Geographic Information Systems*

Remote sensing is dominated almost exclusively by the raster model of continuous or discrete fields. By contrast, and as noted earlier, GIS uses a much broader range of data models, and some currently available GIS support all of the entity and field models. In this paper the focus is on vegetation cover mapping, conceptualized as a discrete or multinomial field. In GIS, such data may have been obtained by remote sensing, or alternatively by digitizing or scanning an existing map. Thus the field's digital representation may use either the raster or polygon model. The choice may depend on the original source of data (raster is more likely if the data came from a classified scene, or by scanning a map, and polygon is more likely if the data were created by digitizing a map); the processing software used (some GIS require the raster model for analysis and modeling, and some require the polygon model); the analysis and modeling to be performed; or the preferences of the user.

Thus several options for the representation of a vegetation map can be found in GIS practice. A raster model may have been obtained by classifying a scene, or by scanning a map; a polygon model may have been obtained by digitizing a map, or by vectorizing a raster representation. In the raster case the error inherent in the model is attributable to the replacement of the true value at a point by the dominant value in one or a small array

(e.g. 3 by 3) of pixels; in the polygon case, the model imposes an error through the replacement of the true value at a point by the dominant value within a polygon of arbitrary size and shape.

Thus both approaches involve information loss. In the raster case, this is imposed by the pixel size, which is likely a characteristic of the sensor beyond the user's control, and by the perceived need to remove 'salt and pepper'. In the polygon case, the reasons behind the lumping of space into polygons are more complex, and are reviewed here at some length. In each case, the discussion includes an interpretation of the minimum mapping unit, the area of the smallest polygon in the model.

*Cartographic:* The traditions of map making have evolved under the constraints imposed by available technology. It is comparatively easy to draw a map of homogeneous areas separated by sharp boundaries - the boundaries can be drawn by pen, and the areas can be filled using a variety of forms of shading or by applying prepared textures. It is much more difficult to portray fuzzy boundaries, small inclusions within larger areas, ecotones, or spatially continuous change of any kind. The minimum mapping unit may be the smallest area that can be conveniently drawn and labeled at the planned scale of the map. In addition, area boundaries may be smoothed to create a cleaner appearance.

*Management:* Maps of vegetation cover are often made for well-defined purposes, such as forest management, and the areas shown on them may be managed as stands. While vegetation cover may change continuously, management practice must work within well-defined boundaries, and must be applied uniformly over each well-defined area. The minimum mapping unit may be the smallest area that can be conveniently managed.

*Cognitive:* Much mapping tradition is concerned not so much with the accurate measurement of conditions on the ground as with the communication of an impression of geographic variation; in this interpretation, the cartographer plays a distinct role. It was observed earlier that the entity models have more in common with everyday human spatial cognition than the field models: the polygon model, which blocks space into entities of uniform character, can be seen as a response to human patterns of thinking about geographic variation. The minimum mapping unit may reflect a human need to impose a certain level of uniformity on the landscape.

For some mix of the above reasons, it is common for vegetation cover data obtained by remote sensing to be aggregated into polygons with some minimum mapping unit greater than one pixel (or nine pixels if a post-

classification filter has been applied). For example, it is not uncommon to use Landsat Thematic Mapper data (pixels of 0.09 ha) aggregated to a minimum mapping unit as large as 20 ha, an aggregation factor of over 200. Aggregation produces polygons whose value becomes the commonest class of the pixels within the polygon, although many other aggregation rules are available. Pixels of other classes contained within the polygon contribute to the error inherent in the model.

In summary, while both remote sensing and GIS may embed a conception of the vegetation mapping task as the creation of a digital representation of a univariate, nominal (multinomial) field, they differ in several important respects that impede integration, or the effective and integrated use of both tools. Because of the nature of measurement instruments on the one hand, and different disciplinary traditions on the other, remote sensing and GIS have been dominated by two distinct data models, raster and polygon. Since both approximate the vegetation field in different ways, techniques for measuring and expressing data uncertainty, or related effects of scale, resolution, and spatial aggregation, are generally incompatible. The next section reviews the techniques that have been developed for error modeling in remote sensing, GIS, and the related field of cartography, and places them within the integrated conceptual framework discussed earlier.

## Error models

For the reasons discussed in the preceding section, it is impossible to create an exact representation of the spatial distribution of a phenomenon as complex as vegetation within a digital database. At the same time, a digital representation is precise, and capable of giving unambiguous answers to simple queries. It follows that the response to a query, and the results of analysis, depend on the details of the representation. As discussed in the previous section, the representations used in remote sensing and GIS are broadly distinct and incompatible.

This discussion will focus on two particular generic queries, on the assumption that more complex queries can be reduced to them. The two queries of interest are:

- the value of the vegetation cover class field at some specified point ( $x,y$ ); and
- the area of a specified vegetation cover class.

In order not to be misled by the precision of answers to these queries, it is desirable that the database respond with some appropriate measure of uncertainty in both cases. In the case of a point query, the response should include not only the class of the containing pixel or polygon, but also some measure of the variation known

to exist within the spatial object, whatever its source. In the case of an area estimate, the response should include some expression of confidence limits around the value of the estimate.

Databases containing vegetation cover maps are being widely used at this time for purposes such as land use control, and environmental management. If uncertainty is known to exist in the responses to these two generic queries, it is crucial that that information be conveyed as succinctly as possible to the user or decision-maker. It would also be helpful if map displays could also convey some impression of uncertainty, as discussed later in the paper.

For these reasons, there has been increasing interest in accuracy issues in recent years, particularly in the research community. This research follows three distinct threads, which will be discussed here under the headings of classification, cartographic, and spatial statistical. Each offers its own answers to questions of how accuracy should be described and measured, and propagated into estimates of the uncertainty associated with responses to queries. Because the remote sensing community has tended to follow the classification approach, and the GIS community has favored the cartographic approach, there is substantial incompatibility in their approaches to description of uncertainty.

Anderson et al. (1976) proposed that 85 % be an appropriate accuracy standard for land cover mapping. The following sections review the various meanings that have been ascribed to that statement, and associated methods of measurement.

### Classification

Over the past three decades, there has been widespread use of remote sensing imagery to classify the land surface. The spectral response of each pixel in a scene is used to assign the pixel to one of a number of classes, using various classification techniques. In this context, accuracy is logically viewed as a problem of misclassification. A pixel is said to be misclassified if its true class, as determined by ground check or from a source of higher accuracy, is  $j$ , but its assignment by the classifier is  $i$  (in this discussion the terms true or truth should always be taken to imply ground check or a source of higher accuracy; the general term reference data is often used). Pixels assigned to a class other than their true class are termed errors of omission, or false negatives; pixels assigned to class  $i$  that are not truly of that class are termed errors of commission, or false positives. Note that any one misclassified pixel can be regarded as either an error of omission or an error of commission, depending on the perspective taken. Story & Congalton (1986) refer to omission errors as produc-

er's accuracy because the producer's concern is presumably to avoid them; commission errors define user's accuracy because a user is interested in knowing how many pixels that appear to be of class  $i$  are actually of class  $j$ .

Consider an idealized case, an image of a forested area, divided into pure stands of different species. Because of the inherent limitations of remote sensing, it is inevitable that some misclassification or confusion of classes will occur, where pixels falling in stands of species  $j$  are mistakenly assigned to class  $i$ . The likelihood that this will happen depends largely on the difference in spectral response of the two species, and also on the technical efficiency of the classification method used; it is also dependent on the proportion of pixels that intersect the edges of stands. It is helpful to think of misclassification in the form of a matrix, in which the cell in row  $i$ , column  $j$  gives the number of pixels that are truly class  $j$  but have been assigned to class  $i$  (Table 1). This matrix has also been called the error matrix or confusion matrix. The term error matrix will be used in this paper.

If the pixels that have been subject to accuracy assessment can be regarded as a random sample of the entire population of pixels, then the entries in row  $i$  can be divided by the row total to give an estimate of the probability that a pixel of class  $i$  is actually of class  $j$ ,  $p(j|i)$ . For example, in Table 1,  $p(B|A) = 0.265$ . Alternatively, the entries in column  $j$  can be divided by the column total to give an estimate of the probability that a pixel that is actually in class  $j$  has been assigned to class  $i$ ,  $p(i|j)$ . Here, as elsewhere in this approach, it is assumed that the probability of misclassification is constant for a given class over the entire image. In practice, some areas may be more likely to be misclassified than others, because of differences in the growing stage of the trees, and numerous other factors. Several studies (e.g. Campbell 1981; Congalton 1988a) have looked at the extent to which errors tend to cluster in space, indicating that the probabilities of errors are not constant.

In order to estimate the error matrix, it is common to conduct a random sample stratified by class. A random sample of pixels classified as  $i$  is selected and checked against the truth. The proportion that is found to be of true class  $j$  is then used to estimate the probability  $p(j|i)$ . Stratification is used because some classes tend to be more abundant than others; in order to obtain a reasonably sound estimate of the accuracy in each class, it is necessary to assign a greater number of samples to the rarer classes than would occur if sample points were located randomly. There have been several studies comparing the effectiveness of alternative sampling designs, particularly concerning the number of samples to be

allocated to each class in the stratified approach, and the location of samples within areas of each class (e.g. Hay 1979; Fitzpatrick-Lins 1981; Rosenfield et al. 1982; Congalton 1988b).

Much research effort has gone into devising suitable overall measures to summarize the contents of an error matrix. Let  $x_{ij}$  denote the number of cases recorded in row  $i$ , column  $j$  of the matrix, that is, the number of cases where the true class was  $j$  and the assigned class was  $i$ . The percent correctly classified compares the number of correctly classified pixels (those appearing on the diagonal of the matrix) to the total number of pixels (in Table 1,  $247/342 = 72.2\%$ ). In order for this statistic to be meaningful as an estimate of the probability that a randomly chosen pixel will have been assigned the correct class, it is necessary that the proportion of test pixels in each column (the ratio of the column total to the grand total) be the same as the proportion of total area that is truly of that class. Since this is generally unknown, the proportion of total area assigned to that class is used instead. Thus the row totals, rather than the column totals, are in the same proportions as the observed areas. When a stratified sampling scheme has been used, the totals must be adjusted by weighting each case by the inverse of that class's sampling density.

Unfortunately – although it is the most obvious interpretation of Anderson et al.'s (1976) recommended 85% accuracy threshold – the percent correctly classified can be a misleading statistic because a certain number of correctly classified cases is expected to occur by chance, even in the most confused classification. Thus it is often replaced by a statistic which allows for chance, and ranges from 0 in the case of the most confused classification to 1 in the case of the most accurate. Various known as Cohen's Kappa and  $K_{\text{hat}}$ , it is defined as follows:

$$\hat{k} = \frac{\sum_i x_{ii} - \sum_i \frac{x_{i.}x_{.i}}{N}}{N - \sum_i \frac{x_{i.}x_{.i}}{N}} \quad (1)$$

where a subscript replaced by a dot indicates summation over that subscript, and  $N$  is the total sample size (Congalton et al. 1983; Hudson & Ramm 1987). For Table 1, Kappa = 61.6%. Rosenfield & Fitzpatrick-Lins (1986) describe a variant on the Kappa statistic that can be calculated separately for each class; this is especially useful when it is desirable to know the classification accuracy of each class.

In some applications it may be useful to compare error matrices. For example, a study may compare the effectiveness of two image classifiers applied to the

same scene; it is then important to know whether one classifier has outperformed the other (for a study of interpreter variability in remote sensing classification see McGwire 1992). It may also be interesting to compare the effectiveness of a classification to a random assignment of classes, to see if the classifier has done better than chance. For many purposes, a simple comparison of Kappa values is sufficient. But, since any error matrix or Kappa statistic is based on a limited sample, one might want to know whether the apparent difference between two Kappa statistics could have occurred by chance, implying that if other samples had been taken, the answer might have been different.

Congalton et al. [1983; see Hudson & Ramm (1987) for corrected equations] describe methods for conducting inferential tests of Kappa, using the methods for testing discrete contingency tables described by Bishop et al. (1975). In their simplest form, such tests assume that the entries in the error matrices are strictly the number of times a condition was observed. Weighting, or other numerical manipulation of the table, will invalidate this assumption, and must be dealt with using special variants of the test.

The approach just described relies for its success on the validity of its underlying conceptual model, of areas of homogeneous class divided by sharp discontinuities. This model may apply well to agricultural fields or pure stands, but in forestland mapping it may be confused by continuous variation within more or less homogeneous areas, and slow transitions across boundaries along ecotones. Both of these lead to a condition described statistically as non-stationarity, in which the error matrix captures only an average over the entire study area; locally, and for a variety of reasons, the probabilities of misclassification may vary markedly. To deal effectively with non-stationarity, it is necessary to partition the study area into regions of more or less constant error probabilities, and to sample them with as many samples as would have been used to characterize the entire study area under stationarity; otherwise, the estimates of the error matrix will be arbitrarily affected by the limits of the study area. Moreover, the basis of regionalization is unlikely to be known in advance, although Congalton et

Table 1. Example of an error matrix of four classes.

	True class				
	A	B	C	D	
Observed class	A	50	22	3	8
	B	5	85	0	4
	C	3	14	22	15
	D	1	19	1	90



al. (1983) describe discrete multivariate methods that can potentially be used to identify the factors that cause spatial variation in rates of error.

In addition to the problems of misclassification of spectral response, some pixels will fall on the boundaries between homogeneous patches, and will therefore contain more than one class. Such mixed pixels are not misclassifications in the same sense, and it is common to avoid pixels that straddle boundaries in accuracy assessment and the development of error matrices. Recently, there has been much interest in improved image classification methods that deal explicitly with the mixed pixel problem (e.g. Smith et al. 1990).

An alternative approach is provided by fuzzy and probabilistic classification techniques, which seek to assign values to each pixel representing the pixel's degree or probability of membership in each class respectively (Fisher & Pathirana 1990; Foody et al. 1992; Foody 1992). As yet, these methods lack clearly defined concepts of error or accuracy assessment.

#### *Cartographic aspects*

To a cartographer, the problem of map accuracy assessment is perceived very differently. Whether its source was remotely sensed imagery, aerial photography, or ground mapping, a map of forestland cover shows a set of non-overlapping, space-exhausting areas with homogeneous characteristics, bounded by lines of constant width. In one of the final stages of the map-making process, the boundaries will likely have been smoothed, or splined, to create a pleasing, orderly appearance.

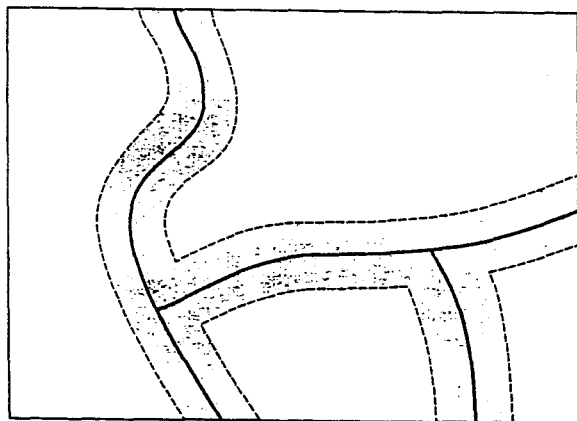


Fig. 3. Epsilon bands of positional uncertainty applied to the boundaries in a vegetation class map.

The accuracy of such maps is often seen in terms of two questions:

- are the boundaries in the correct locations; and
  - have the areas been assigned to the correct classes?
- Hord & Brooner (1976) use this approach in their analysis of land-use map accuracy. The new United States spatial data transfer standard, now known as Federal Information Processing Standard (FIPS) 173 [see Morrison (1992), and other papers in that journal issue], refers to these as positional accuracy and attribute accuracy respectively. Positional accuracy can be measured by the Perkal epsilon band, a band of width epsilon centered on the observed location of the boundary, within which the true position of the boundary is assumed to lie (Blakemore 1984), and Mark & Csillag (1989) have described an alternative probabilistic approach. Fig. 3 illustrates the concept of an epsilon band. Attribute accuracy can be described by a form of error matrix, but note that the matrix now refers to the classification of each area, rather than to each pixel. In this approach, the percent of areas correctly classified is compared to Anderson et al.'s (1976) threshold of 85 %.

This form of accuracy assessment asks only whether each area or polygon has been assigned to the correct class, and thus ignores the almost inevitable heterogeneity that results from the approximation of continuously varying land cover characteristics. If an area is known to be heterogeneous, the 'correct' class is normally assumed to be the commonest class, but no assessment is made of the occurrence of other classes within the polygon. This approach to accuracy assessment is often referred to as per polygon accuracy assessment, to distinguish it from the per pixel or per point approach of the classification school. Unfortunately, this means that an accuracy assessment based on the cartographic approach cannot be transformed into one based on the classification approach, or vice versa.

The cartographic approach has clear advantages if it is reasonable to assume that areas are homogeneous, and that boundaries are sharp and distinct (crisp). On the other hand the classification approach is clearly more appropriate if areas are heterogeneous, with substantial inclusions of unmapped classes, or if boundaries are indistinct (fuzzy). Chrisman (1989) shows how it is possible using the cartographic approach to distinguish between different sources of error, by observing whether they result in small shifts in boundaries, or in reclassification of entire areas (see also Chrisman & Lester 1991; Lester & Chrisman 1991). In reality, some boundaries on land cover maps are more distinct than others, and some areas are more homogeneous than others; the distinction is rarely clearcut.

### *Spatial statistical aspects*

From a statistical perspective, error is variation that cannot be explained. Most forms of measurement are subject to error, and statistical methods have been developed for describing the amount of error present, and for analyzing data despite the presence of error.

In the measurement of a simple quantity like temperature, it is conventional to conceive of a true value, which has been distorted to an unknown degree by error. Repeated measurements using different observers, or different measuring instruments, would produce a range of values characterized by a mean and a standard deviation, and often with a histogram that follows the normal or Gaussian distribution closely. A range of well-developed methods allow such data to be analyzed.

The spatial equivalent of the Gaussian distribution is a range of maps, each showing the same general characteristics but with individual patterns of distortion, and each being equally likely to have been observed. The range might represent the work of different interpreters, or the use of different training sites for image classification. An error model is defined as a statistical process capable of generating such a range of maps given appropriate parameters (the Gaussian distribution is the most commonly used error model for simple measurements). Any one such map is termed a realization of the model. Simple measures of error, such as the percent correctly classified, are termed error descriptors, and may be related to the parameters of an error model. For example, the standard deviation, a commonly used measure of measurement uncertainty, is a parameter of the Gaussian distribution.

The use of an error model has numerous advantages. The effects of uncertainty can be visualized by using the error model to simulate a sample of possible maps. Errors can be propagated by performing analysis on several realizations of the error model, and computing the variation in results across them. The parameters of an error model can be calibrated by adjusting them so that the range of outcomes under the error model matches the range observed in reality.

One of the more straightforward applications of forestland mapping is in the estimation of areas having particular characteristics. For example, one might want to know the total area of mixed conifer forest in Mendocino County, California. Knowing that the database includes misclassifications and errors, it would be useful to know the standard error associated with the estimate of area. Unfortunately, neither the classification approach, with its error matrix and Kappa statistics, nor the cartographic approach with its epsilon bands and polygon error matrix, are capable of providing such an estimate. However, given an error model, an estimate of

standard error can be made by generating a sample of realizations, calculating area on each, and computing the variation between them. Under appropriate circumstances, it might be possible to compute the standard error analytically.

Several recent papers have described error models for maps, and their use in predicting the effects of uncertainty. Of most relevance to this paper are those that deal with land cover maps, generally known as area class maps. Goodchild et al. (1992) propose a general error model, and show how it can be used to obtain estimates of uncertainty in such GIS products as area estimates. In their model, the uncertainty associated with the class at any point is represented by a probability vector in a raster representation. Each pixel  $ij$  is associated with a vector of probabilities  $\{p_{ij1}, p_{ij2}, \dots, p_{ijn}\}$ , giving the probability that the pixel truly belongs to each class 1 through  $n$ . These probabilities can be interpreted as the consequence of mixed pixels, of uncertainty of class definition, as derivatives of fuzzy or probabilistic classification, or as the effect of confusion of spectral signatures - in this spatial statistical approach, the origins of error are not of immediate concern. The classification approach described earlier assumes that the contents of the vector are determined by the class to which the pixel appears to belong.

In this model, the classes allocated to the pixels in the database represent one realization of a stochastic process defined by these vectors of probabilities. For example, if there are two classes and the probabilities for a given pixel are  $\{0.5, 0.5\}$ , then one interpretation might be that in the maps made by two independent observers, one map would assign the pixel to class 1 and the other to class 2. Over a large number of realizations, the proportion of times the pixel is assigned to each class will converge on each class's probability.

In addition, Goodchild et al. (1992) propose that within any one realization, the outcomes in neighboring pixels be correlated, and the model also includes parameters describing the level of spatial dependence. For example, if a large area of many pixels is suspected of being conifer forest, but with a small probability of being hardwood forest, a low level of spatial dependence would imply that hardwood occurs in small inclusions; a high level of spatial dependence corresponds to large inclusions, or to the possibility that the entire patch is hardwood rather than conifer.

The relationship between this model and the error matrix discussed earlier is straightforward: the error matrix assumes that probability vectors are a function only of the class assigned to each pixel in an observed realization, and leaves spatial dependence undefined. The relationship to the cartographic approach requires more detailed explanation.

Consider a fuzzy boundary line between two classes A and B on a map of vegetation cover. In probabilistic terms, the fuzziness can be represented as a rate of change of probability as the boundary is traversed. Far to one side of the boundary, the probability that a pixel belongs to class A is 1, and the probability of class B is 0. As the boundary is approached, the probability of class A falls, at a rate depending on the degree of fuzziness. Each realization of the stochastic process will assign each pixel to one class, allowing the boundary to be inferred in a position that will vary from one realization to another within the general area of the fuzzy boundary. Crisp boundaries are represented by sharp changes in pixel probabilities as the boundary is crossed; every realization will place the inferred boundary in the same position.

A less general model that omits spatial autocorrelation between outcomes in neighboring cells has been described by Fisher (1991) for soil maps. A general model for raster data has been developed by Haining & Arbia (1993), and Fisher (1992) has used a model of spatially autocorrelated errors in digital elevation models.

In summary, while the classification and cartographic approaches are useful to monitor errors in image classification and map making respectively, they are mutually incompatible, and only the spatial statistical approach offers sufficient generality to deal comprehensively with the problem of error propagation into the products of GIS analysis, and to provide a bridge between the distinct traditions of remote sensing and GIS.

### Conclusions

Mapping of vegetation has a long and largely effective history. However, the introduction of remote sensing as a commonly used and efficient means of mapping vegetation has led to two contrasting and potentially conflicting approaches, termed here the classification and cartographic. The two approaches are particularly distinct in their views on uncertainty; since any digital representation of geographic variation must be an approximation, handling and representation of uncertainty is a critical issue in the use of such data for analysis and modeling.

This paper has proposed a reconciliation of these two distinct traditions, through what has been termed the spatial statistical approach. A vegetation map is viewed as a multinomial field, a single realization of a multinomial error model. Uncertainty is described through the parameters of the model, and expresses itself in the variation between realizations of the model. The representation of the infinite information present in a field in the discrete space of a digital computer is one significant source of error, together with a variety of

other forms introduced at various stages in the data collection and interpretation processes.

Recent interest in conservation, and in environmental change, has led to a renewal of efforts to devise better methods of land surface characterization, of which vegetation mapping is a significant element. Such mapping will necessarily always fall short of being a precise exercise in scientific measurement, despite the availability of remote sensing. However, the unification of classification and cartographic approaches, and the traditional perspectives of remote sensing and GIS, under a single framework may help to remove some of the problems that have plagued the field in the past.

**Acknowledgements.** The National Center for Geographic Information and Analysis is supported by the National Science Foundation, grant SES 85-10917.

### References

- Anderson, J.R., Hardy, E.E., Roach, J.T. & Witmer, R.E. 1976. *A land use and land cover classification system for use with remote sensor data*. U.S. Geological Survey, Professional Paper 964. Washington, DC.
- Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.
- Blakemore, M. 1984. Generalization and error in spatial data bases. *Cartographica* 21: 131-139.
- Burrough, P.A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Clarendon Press, Oxford.
- Campbell, J.B. 1981. Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogramm. Eng. Remote Sens.* 47: 355-363.
- Chrisman, N.R. 1989. Modeling error in overlaid categorical maps. In: Goodchild, M.F. & Gopal, S. (eds.) *Accuracy of Spatial Databases*. pp. 21-34. Taylor and Francis, London.
- Chrisman, N.R. & Lester, M.K. 1991. A diagnostic test for categorical maps. *Technical Papers, 1991 ACSM-ASPRS Annual Convention* 6: 330-348. ASPRS/ACSM, Bethesda, MD.
- Colwell, R.N. 1983. *Manual of Remote Sensing*. American Society of Photogrammetry, Falls Church, VA.
- Congalton, R.G. 1988a. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogramm. Eng. Remote Sens.* 54: 587-592.
- Congalton, R.G. 1988b. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogramm. Eng. Remote Sens.* 54: 593-600.
- Congalton, R.G., Oderwald, R.G. & Mead, R.A. 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogramm.*

- Eng. Remote Sens.* 49: 1671-1678.
- Fisher, P.F. 1991. Modeling soil map-unit inclusions by Monte Carlo simulation. *Int. J. Geogr. Inf. Syst.* 5: 193-208.
- Fisher, P.F. 1992. First experiments in viewshed uncertainty - simulating fuzzy viewsheds. *Photogramm. Eng. Remote Sens.* 58: 345-352.
- Fisher, P.F. & Pathirana, S. 1990. The evaluation of fuzzy membership of land cover classes in the suburban zone. *Remote Sens. Environ.* 34: 121-132.
- Fitzpatrick-Lins, K. 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map. *Photogramm. Eng. Remote Sens.* 47: 343-351.
- Foody, G.M. 1992. A fuzzy sets approach to the representation of vegetation continua from remotely sensed data - an example from lowland heath. *Photogramm. Eng. Remote Sens.* 58: 221-225.
- Foody, G.M., Campbell, N.A., Trodd, N.M. & Wood, T.F. 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogramm. Eng. Remote Sens.* 58: 1335-1341.
- Goodchild, M.F. 1992. Geographical data modeling. *Comp. Geosci.* 18: 401-408.
- Goodchild, M.F., Parks, B.O. & Steyaert, L.T. 1993. *Environmental Modeling with GIS*. Oxford University Press, New York, NY.
- Goodchild, M.F., Sun, G.Q. & Yang, S. 1992. Development and test of an error model for categorical data. *Int. J. Geogr. Inf. Syst.* 6: 87-104.
- Haines-Young, R., Green, D.R. & Cousins, S. 1993. *Landscape Ecology and Geographic Information Systems*. Taylor and Francis, London and New York.
- Haining, R.P. & Arbia, G. 1993. Error propagation through map operations. *Technometrics* 35: 293-305.
- Hay, A.M. 1979. Sampling designs to test land-use map accuracy. *Photogramm. Eng. Remote Sens.* 45: 529-533.
- Hord, R.M. & Brooner, W. 1976. Land use map accuracy criteria. *Photogramm. Eng. Remote Sens.* 42: 671-677.
- Hudson, W.D. & Ramm, C.W. 1987. Correct formulation of the Kappa coefficient of agreement. *Photogramm. Eng. Remote Sens.* 53: 421-422.
- Jensen, J.R. 1986. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Prentice-Hall, Englewood Cliffs, NJ.
- Lester, M. & Chrisman, N.R. 1991. *Not all slivers are skinny: a comparison of two methods for detecting positional error in categorical maps*. Proc. GIS/LIS 91, ASPRS/ACSM, pp. 648-658. Bethesda, MD.
- Lillesand, T.M. & Kiefer, R.W. 1987. *Remote Sensing and Image Interpretation*. Wiley, New York, NY.
- Maguire, D.J., Goodchild, M.F. & Rhind, D.W. 1991. *Geographical Information Systems: Principles and Applications*. Longman, London.
- Mark, D.M. & Csillag, F. 1989. The nature of boundaries on 'area-class' maps. *Cartographica* 26(1): 65-78.
- McGwire, K.C. 1992. Analyst variability in labeling of unsupervised classifications. *Photogramm. Eng. Remote Sens.* 58: 1673-1677.
- Morrison, J.L. 1992. Implementing the spatial data transfer standard - introduction. *Cartogr. Geogr. Inf. Syst.* 19: 277.
- Rosenfield, G.H. & Fitzpatrick-Lins, K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogramm. Eng. Remote Sens.* 52: 223-227.
- Rosenfield, G.H., Fitzpatrick-Lins, K. & Ling, H.S. 1982. Sampling for thematic map accuracy testing. *Photogramm. Eng. Remote Sens.* 48: 131-137.
- Smith, M.O., Ustin, S.L., Adams, J.B. & Gillespie, A.R. 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sens. Environ.* 31: 1-26.
- Star, J.L. & Estes, J.E. 1991. *Geographic Information Systems: An Introduction*. Prentice-Hall, Englewood Cliffs, NJ.
- Story, M. & Congalton, R.G. 1986. Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote Sens.* 52: 397-399.

Received 20 November 1993;  
Revision received 21 April 1994;  
Accepted 22 August 1994.