Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets

Sucharita Gopal and Curtis Woodcock

Abstract
The use of fuzzy sets in map accuracy assessment expands the amount of information that can be provided regarding the nature, frequency, magnitude, and source of errors in a thematic map. The need for using fuzzy sets arises from the observation that all map locations do not fit unambiguously in a single map category. Fuzzy sets allow for varying levels of set membership for multiple map categories. A linguistic measurement scale allows the kinds of comments commonly made during map evaluations to be used to quantify map accuracy. Four tables result from the use of fuzzy functions, and when taken together they provide more information than traditional confusion matrices. The use of a hypothetical dataset helps illustrate the benefits of the new methods. It is hoped that the enhanced ability to evaluate maps resulting from the use of fuzzy sets will improve our understanding of uncertainty in maps and facilitate improved error modeling.

Introduction: The Assessment of Map Accuracy
Thematic (categorical) maps are made for a specific purpose and portray information using some system of classification for the landscape (soil taxonomy, vegetation classes, land use, etc.). They are increasingly being used by a variety of people involved in the management of land and resources. The increase in their use is largely due to remote sensing and geographic information systems (GIS). Remote sensing is now used widely to produce thematic maps. GIS technology both employs thematic maps as input data sources and creates new thematic maps utilizing the analysis of existing maps and data.

The purpose of this paper is to present new methods for assessing the accuracy of thematic maps based on fuzzy sets. The intent of the new methods is to allow explicitly for the possibility of ambiguity regarding the appropriate map label at any location. In addition, expanded possibilities for error analysis result from the use of fuzzy sets. The focus of this paper is on the theory behind the approach and on a thorough description of the methods. In another paper, these methods are applied to assess the accuracy of a map of forest vegetation, which helps illustrate the utility of the proposed approach (Woodcock and Gopal, 1992).

Background
The difficulties associated with assessing the accuracy of thematic maps are the result of the nature of thematic maps. In thematic maps, each location on the ground has to be assigned to a category (or class). In essence, the continuum of variation found in the landscape has to be divided into a finite set of categories. Typically, the categories are easily differentiable in their pure states, and become less readily separable near the dividing lines between the categories. For example, consider the difference between the vegetation categories *conifer forest* and *hardwood forest.* At their extremes there is no question regarding the appropriate category. However, all degrees of mixing of coniferous and hardwood trees may be found. When coniferous trees dominate, the appropriate label may still be *coniferous forest,* but what happens as the mix approaches 50 percent of each? At that point the decision becomes arbitrary and neither category is either entirely right or entirely wrong. One solution is to add another category to the map that is *mixed forest.* This new category solves the problem in one case (the 50-50 mix), but now there is a problem defining the break between *mixed forest* and both *hardwood forest* and *conifer forest.*

One can argue that these problems of ambiguity concerning the appropriate map label for a given location can be solved by a precise set of specifications for the definitions of each of the map categories, and for these categories to cover all possible situations found on the ground, and for the categories to be mutually exclusive. In fact, these kinds of definitions for map categories would minimize the ambiguity. Most mapping projects, however, are not conducted with an extensive system of breakpoints defining the differences between categories. Accurate and precise measurements of the required parameters on the ground are often either difficult or impossible. Take, for example, the difference between a *meadow* category and a *grassland* category, which may be primarily a function of annual water balance and may not be measurable at a single visit to a site, or even from measurements from a single year. Beyond the problem of category definitions and measurement difficulties, map categories are frequently constructs that do not lend themselves to physical measurement. Examples might be maps of land or residential cost or habitat suitability. There are often descriptions for these categories, but rarely definitions that can be used to make unambiguous decisions for individual sites. The problem that makes accuracy assessment difficult is that there is ambiguity regarding the appropriate map label for some locations. The situation of one category being exactly right and all other categories being equally and exactly wrong often does not exist.

The representation of map categories also depends on scale, as scale generally implies a degree of spatial generali-

Department of Geography, Boston University, Boston, MA 02215.

Photogrammetric Engineering & Remote Sensing,
Vol. 60, No. 2, February 1994, pp. 181–188.

0099-1112/94/6002–181$03.00/0

zation and a minimum mapping unit (Woodcock and Harward, 1992). The minimum mapping unit is defined as the size of the smallest object to be represented on the map. If an object is below the minimum mapping unit, then it is merged with its surroundings. However, if the size of the minimum mapping unit changes, the object may be represented on the map.

There are two primary motivations for assessing the accuracy of a map. The first one concerns trying to understand the errors in the map. Both producers and users of thematic maps are interested in this kind of information. Producers can improve methods of making the maps and of presenting the information on accuracy and errors to the end user. Information about the errors in the map, in turn, can help map users to interpret and use the map more effectively. The second motivation for assessing the accuracy of maps is to provide an overall assessment that can be used as an indication of the general reliability of a map. For example, one use of such an overall assessment would be to compare two maps in order to determine which is better.

The four kinds of information that are desired from an accuracy assessment about the errors in a map are their *nature, frequency, magnitude*, and *source.*The *nature* of the errors concerns which categories are confused in the map, meaning between which categories are there mismatches between the category assigned in the map and the reality at the ground location. In the simple case discussed previously, are there situations where conifer trees have been assigned map labels of *hardwood forest*, and vice versa? *Frequency* refers to how often these mismatched situations between the map and the ground occur. A more subtle concept concerns the *magnitude of errors.* At one level, certain kinds of mismatches can be considered greater in error than others. For example, assigning a map label of *water* to an area of coniferous trees might be considered a larger error than assigning it a map label of *hardwood forest.* At a second level, assigning an area of 100 percent coniferous trees a *hardwood forest* label might be a more serious error than assigning the same label to an area that is 60 percent coniferous trees and 40 percent hardwood trees. The last kind of information desired is the *source of errors*, which is less well defined and primarily of concern to the people who make maps. Essentially, what is desired is any information about the conditions in which errors tend to be made that can help reduce the likelihood of the same kinds of errors being made in the future.

Traditional Methods of Assessing Accuracy

Traditionally, the accuracy of a thematic map is determined empirically by comparing the map with corresponding reference or ground data. The results are tabulated in the form of a square matrix whose columns usually represent the ground data (i.e., assumed correct) and rows indicate the mapped data. Each element in the matrix gives the number of areas on the map assigned to a particular category relative to the ground data. The elements of the principal diagonal of the matrix represent the correct matches and the remaining elements, mismatches. This matrix is popularly known as an *error* or *confusion matrix* (Card, 1982) and forms the basis for a series of descriptive and analytical statistical techniques (Congalton, 1991; Hoffer and Fleming, 1978; Rosenfield and Fitzpatrick-Lins, 1986).

The ideal situation in map accuracy is represented by a diagonal matrix where only principal diagonal elements have non-zero values; all areas on the map have been correctly classified (van Genderen and Lock, 1977; Mead and Szajgin, 1982; Congalton *et al*, 1983). This situation is rarely the case. A distinction is often made between producer's and user's accuracy. Producer's accuracy indicates the probability of a test or reference location known to belong to category *C* is accurately labeled as category *C*. It is a measure of *omission error.* This accuracy measure is obtained by dividing the total number of correct entries in a category (i.e., diagonal element) by the total number of entries of that category as derived from the reference data, i.e., the column total. On the other hand, user's accuracy indicates the probability that a test or reference location labeled as category *C* belongs to category *C* (Story and Congalton, 1986). This is a measure of *commission error.* It is estimated by dividing the total number of correct entries in a category by the total number of entries that were classified in that category, i.e., the row total.

One advantage of these methods is that they yield a single overall map accuracy index, usually presented as a percent correct. This measure can simply be the sum of the major diagonal divided by the total number of samples, or the accuracies of each category in the map can be weighted by its area in the map and these values summed (Card, 1982). Similarly, a statistic, *kappa*, can be calculated that provides a measure of difference between the observed agreement between two maps and agreement that is contributed by chance (Congalton and Mead, 1983; Rosenfield and Fitzpatrick-Lin, 1986; Rosenfield, 1981). Conditional kappa coefficients can also be calculated for individual map categories. For those interested readers, Congalton has recently provided a thorough review of methods for assessing accuracy of thematic maps and the relevant current research (Congalton, 1991).

The traditional methods of assessing accuracy, outlined above, suffer from a number of limitations:

(1) It is assumed that each area in the map can be unambiguously assigned to a single map category. In assessing ground truth for map accuracy, the expert has to resolve this issue by selecting a single category for each ground location and matching this against the map value.
(2) Information on the magnitude of errors is limited to noting the pattern of mismatches between categories in the map. Data concerning the magnitude or seriousness of these mismatches as indicated by the conditions of the ground site cannot be used.
(3) Third, the user needs to be provided with more complete and interpretable information about the map than is currently practiced (Aronoff, 1982a; 1982b). Detailed information on errors will help the users to check if the map can be used for a particular purpose.

The methods presented in this paper provide information about thematic maps and the errors in those maps pertinent to the three issues outline above. The intention is to improve the utility of thematic maps through a better understanding of their errors.

Fuzzy Sets

This section provides some essential background on fuzzy sets and a description of the use of fuzzy sets in the context of map accuracy assessment. The concept of a fuzzy set was introduced by Zadeh (1965, 1973) to describe imprecision that is characteristic of much of human reasoning, particularly in domains such as pattern recognition, communication of information, and abstraction. Zadeh (1965; 1973) and others (Dubois and Prade, 1980; Kaufmann, 1975; Goguen, 1969) have outlined quantitative techniques for dealing with vagueness in complex systems.

Formally, a fuzzy set can be defined as follows:

Let X be a universe of points (or objects) with a generic element of X being denoted by x. A *fuzzy set* of X, labeled A, is characterized by a *membership* or *characteristic function*, μ_A, which associates with each point in X a real number in the closed interval (0,1). The value of $\mu_A(x)$ at x represents the grade of membership of x in A. This can be designated as

$$A = \{(x,\mu_A(x)) \mid x \in X\}.$$

Note that the nearer the value of $\mu_A(x)$ to 1.0, the higher the grade of membership of x in A.

The assumption underlying fuzzy set theory is that the transition from membership to nonmembership is seldom a step function. Rather, there is a gradual but specifiable change from membership to nonmembership. In (classical) set theory, a membership function, $\mu_A(x)$ has only two values 0, 1. In particular,

$$\mu_A(x) = \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{if } x \notin X \end{cases}$$

Fuzzy sets have been used in remote sensing for image interpretation and image classification. While interpreting an image, experts express their assessment using qualitative linguistic values. Fuzzy sets in this context can be used to mathematize these linguistic values and to obtain a consensus if the experts provide different (linguistic) values. Thus, they provide a consistent way to measure and model qualitative values that is useful for subsequent decision-making (Hadipriono *et al.*,1991). Current methods of classification employed in remote sensing generally assume that each pixel belongs to one class only. A classification algorithm based on fuzzy representation (such as Fuzzy *C*-Means classifier) has been shown to be useful in extracting more information from remotely sensed images than conventional algorithms (Trivedi and Bezdek, 1986; Wang, 1990) as well as in generating "hard classification" (Kent and Mardia, 1988). The amount of land cover within a pixel is reflected in the fuzzy membership value derived from the fuzzy classifier (Fisher and Pathirana, 1990).

Fuzzy set theory can also be used in GIS applications. More specifically, it can be used to represent uncertainty in spatial databases and manipulate uncertainty as data are transformed by various GIS functions (Robinson and Strahler, 1984; Leung, 1988; Robinson 1988). For example, a map overlay function results in the creation of a composite map that contains uncertainty and errors. Using fuzzy sets, it is possible to compute the uncertainty in the composite map if the membership values of sites in each cover class were known for each map layer (Veregin, 1989). Fuzzy sets can also be used to estimate similarity and other relationships. Fuzzy relational databases, like FRSIS (Kollias and Voliotis, 1991), can handle imprecision in data representation and manipulation and allow for individualization of data. There are also other examples from land survey (Wang et al., 1990), soil analysis (Burrough, 1988), and representation of viewsheds derived from a DEM (Fisher, 1991; 1992).

Fuzzy Sets in the Context of Map Accuracy Assessment

We define some terms and notations in order to describe how fuzzy sets can be used to analyze the accuracy of thematic maps. Let X be the finite *universe* of discourse, which in the present context is a set of sites (or polygons) represented in the map. Let $\mathcal{C}$ denote the (finite) set of *classes* (or *categories*) assigned to sites in X; let m be the number of categories $|\mathcal{C}| = m$. For each site $x \in X$, we define $\mathcal{X}(x)$ to be the class assigned to x in the map. The set

$$\mathcal{M} = \{(x,\chi(x)) \mid x \in X\}$$

defines the map data.

For the evaluation of map accuracy, usually a subset $S \subset X$ of n sample sites is used. A fuzzy set

$$A_C = \{(x,\mu_C(x)) \mid x \in S\}$$

is associated with each class $C \in \mathcal{C}$ where $\mu_C(x)$ is the *characteristic* or *membership* function of C. Membership functions are derived from the linguistic values provided by the experts in the process of evaluating map accuracy (described below). Thus, the data set used for the fuzzy analysis is an n by m matrix of membership functions, denoted as $\mathcal{A}$. The test sites are the rows of $\mathcal{A}$ and the classes are the columns of $\mathcal{A}$. Abusing the notation slightly, we will also use A_C to denote the column of $\mathcal{A}$ corresponding to the class C. See Table 1 for a hypothetical example using a 40 by 4 matrix. This table also includes two additional columns representing sample site numbers, $x = 1, ..., 40$, or S and the label assigned to the polygon in question, or $\chi(x)$. The different classes A, B, C, and D are in the set $\mathcal{C}$; $\chi(x)$ denotes the map label; and $\mu_A(x)$ denotes the expert evaluation for category A at site (x), $\mu_B(x)$ for a category B, and so on.

Construction of a Linguistic Scale

The methodology presented here was developed based on the observation that experts most often use linguistic constructs to describe map accuracy. It has been the observation of the authors over 10 years of evaluating the accuracy of maps that qualitative, linguistic terms are generally used to evaluate the quality of various map labels for individual sites when visited in the field. Verbal reports of a group of experts during map accuracy assessment were collected and analyzed using the protocol analysis techniques described in Ericsson and Simon (1984). It revealed that experts are not only able to distinguish "absolutely right" from "absolutely wrong" values but are also able to identify intermediate values between these two extremes. At least three different intermediate values are clearly distinguished by all of the experts. Based on this analysis, a five-point membership scale ranging from "absolutely right" to "absolutely wrong" values was developed. The *linguistic values* and the descriptions used by the experts to evaluate a map class at a site are:

(1) *Absolutely Wrong*: This answer is absolutely unacceptable. Very Wrong.
(2) *Understandable but Wrong*: Not a good answer. There is something about the site that makes the answer understandable but there is clearly a better answer. This answer would pose a problem for users of the map. Not Right.
(3) *Reasonable or Acceptable Answer*: Maybe not the best possible answer but it is acceptable; this answer does not pose a problem to the user if it is seen on the map. Right.
(4) *Good Answer*: Would be happy to find this answer given on the map. Very Right.
(5) *Absolutely Right*: No doubt about the match. Perfect.

The following procedure is used to obtain the linguistic membership scale. The expert's job is to evaluate each land-use class at each site and then choose the most suitable linguistic value to describe his/her perception of the nature of match between each map category and the ground truth. The expert does not know what the map class is prior to or during the evaluation procedure. Linguistic values obtained

TABLE 1. A HYPOTHETICAL EXAMPLE; $n = 40$, $m = 4$, AND $C = \{A, B, C, D\}$. THE COLUMNS ($\mu_A(x)$, $\mu_B(x)$, $\mu_C(x)$, $\mu_D(x)$) FORM THE 40×4 MATRIX A. TABLES 2 TO 5 ARE DERIVED FROM THE DATA PRESENTED IN THIS TABLE.

Site No.	Map Label	Expert Evaluation (A)				Site No.	Map Label	Expert Evaluation (A)			
x	$\chi(x)$	$\mu_A(x)$	$\mu_B(x)$	$\mu_C(x)$	$\mu_D(x)$	x	$\chi(x)$	$\mu_A(x)$	$\mu_B(x)$	$\mu_C(x)$	$\mu_D(x)$
1	D	1	4	1	4	21	B	1	1	4	1
2	A	5	1	1	1	22	A	5	1	1	1
3	C	1	1	3	4	23	D	1	1	3	1
4	B	1	5	1	1	24	C	1	1	3	4
5	D	1	4	1	4	25	D	1	1	4	3
6	A	5	1	1	1	26	B	1	2	3	1
7	D	1	4	2	4	27	A	5	1	1	1
8	A	5	1	1	1	28	C	1	1	4	1
9	B	3	4	1	1	29	D	1	2	1	5
10	C	1	3	4	2	30	A	5	1	1	1
11	A	5	1	1	1	31	B	1	2	3	1
12	B	1	3	3	1	32	C	1	5	3	1
13	C	1	3	2	3	33	A	5	2	1	1
14	D	1	3	3	2	34	D	1	1	1	4
15	B	1	4	4	1	35	B	1	1	5	1
16	A	5	1	1	1	36	C	1	3	4	1
17	B	1	4	1	1	37	B	2	5	1	1
18	C	1	1	5	3	38	D	1	4	2	4
19	C	1	2	3	4	39	A	4	3	1	1
20	D	1	4	1	3	40	C	1	3	1	3

from the expert for each class at every site form the input data to Table 1.

For the sake of convenience, the linguistic values are converted into a numerical scale ranging from 1 to 5. Thus, instead of a continuously measured membership function ranging between the usual $0 - 1$ used in fuzzy sets, the membership scale used in the present research is discrete and is based on the five linguistic levels. Each membership function A_C in Table 1 is in the numerical format ranging between 1 and 5.

One of the differences between the traditional and proposed procedure is worth noting. Because each map category is evaluated at each test site, the experts can recognize the heterogeneity of ground cover and ambiguity of map classes and use it to describe the degree of match between each map class and ground data. For example, they might use linguistic values such as "absolutely right" or "reasonably right" to describe the nature of the match. The expert is not limited to a single match for a site or bound to the existence of a perfect match for each site. On the other hand, the traditional method allows only for a "yes" (match) or "no" (mismatch) condition.

Accuracy Assessment Using Fuzzy Operators

A series of measures have been developed based on fuzzy functions that provide indications about the quality of a map and its categories (Zadeh, 1965; 1973; 1975; Goguen, 1969; Dubois and Prade, 1980; Kaufman, 1975). The results of these functions are four tables which, when taken together, give more overall information than is possible from a traditional confusion matrix. The use of these functions and the tables they produce are presented in the context of an example using the linguistic scale previously described. The data used for the example is hypothetical, but is designed to include some of the properties we have found in our initial tests using this approach for map accuracy assessment (Woodcock and Gopal, 1992). Each table is evaluated in terms of its contribution to understanding the nature, frequency, magnitude, and source of errors in a thematic map.

Frequency of Matches and Mismatches

In this subsection a procedure is described that measures the accuracy of the map in terms of the frequency of matches and mismatches between the sample map data and the expert data, i.e., between $\mathcal{M}$ and $\mathcal{A}$. The procedure divides the traditional question of "how accurate is the map?" into the following two more precise questions:

- How frequently is the category assigned in the map the best choice for the site?
- How frequently is the category assigned in the map acceptable?

Define a Boolean function σ that returns a result of zero or one based on whether x belongs to the class C with respect to the matrix $\mathcal{A}$. This evaluation of sample site x from the set S ($x \in S$) and map category $\mathcal{C}$ depends on the criteria used to define σ. That is, $\sigma(x,C) = 1$ if the x "belongs" to C, and $\sigma(x,C) = 0$ if x does not "belong" to C. The definition of "belong" depends on σ. We will consider two different definitions of σ (see below for details). If $\sigma(x,\chi(x)) = 1$, then we say that there is a *match* between the map data and the expert data at site x, and if $\sigma(x, \chi(x)) = 0$, then there is a *mismatch*.

For each map category $C \in \mathcal{C}$, we compute two quantities

$$\omega_C = |\{x \in S \mid \chi(x) = C \text{ and } \sigma(x,C) = 1\}|$$
$$\overline{\omega}_C = |\{x \in S \mid \chi(x) = C \text{ and } \sigma(x,C) = 0\}|.$$

That is, we count the number of sites x with $\chi(x) = C$ in terms of matches (ω_C) and mismatches ($\overline{\omega}_C$).

We now define two functions which can be used as σ. One of them, called MAX, is defined as follows:

$$\text{MAX}(x,C) = \begin{cases} 1 & \text{if } \mu_C(x) \geq \mu_{C'}(x) \text{ for all } C' \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases}$$

That is, MAX (x,C) is 1 if the value of the membership function for x in category C ($\mu_C(x)$), is maximum among all map categories ($\mu_{C'}(x)$).

The other function that we define is called RIGHT. It is

TABLE 2. NATURE AND DISTRIBUTION OF ERRORS USING MAX AND RIGHT

Map Label	Sites	Expert Evaluation MAX(M) Match(ω_C)	Mismatch($\overline{\omega}_C$)	RIGHT(R) Match(ω_C)	Mismatch($\overline{\omega}_C$)	Improvement (R − M)
Class A	10	10 (100.00%)	0 (0.00%)	10 (100.00%)	0 (0.00%)	0 (0.00%)
Class B	10	6 (60.00%)	4 (40.00%)	6 (60.00%)	4 (40.00%)	0 (0.00%)
Class C	10	4 (40.00%)	6 (60.00%)	8 (80.00%)	2 (20.00%)	4 (40.00%)
Class D	10	6 (60.00%)	4 (40.00%)	8 (80.00%)	2 (20.00%)	2 (20.00%)
Total	40	26 (65.00%)	14 (35.00%)	32 (80.00%)	8 (20.00%)	6 (15.00%)

defined with respect to some prespecified *threshold* value τ, where τ is one value in the linguistic scale or a specified value in the membership function. A site x "belongs" to a category C if its membership function, $\mu_C(x) \geq \tau$. Formally,

$$\text{RIGHT}(x,C) = \begin{cases} 1 & \text{if } \mu_C(x) \geq \tau \\ 0 & \text{otherwise.} \end{cases}$$

The threshold value τ in the present research equals any degree of right that is a value ≥ 3 in the linguistic scale described before.

Table 2 is derived from the hypothetical data set shown in Table 1, using the MAX and the RIGHT functions. The first column shows the map category or label for purposes of comparison and the second column shows the total number of sites under each map category. The matches (ω_C) and mismatches ($\overline{\omega}_C$) using the two functions are given both in counts and in percentages in columns 3 to 6 of the table. The last column of the table shows the improvement in accuracy associated with using the RIGHT function instead of the MAX function.

Typically the MAX function is more conservative than the RIGHT function, but that is not always the case. It is possible for a site not to have any map categories that fit the RIGHT function, but still have a highest score. Using our linguistic scale, this situation would occur for a site whose highest score was 2.

Traditional confusion matrices usually follow the approach of either the MAX or the RIGHT function. The MAX function would apply when the confusion matrix comes from blind samples (i.e., when prior knowledge of the map label is not used at the time of the evaluation of the site), and the RIGHT function when the map label is known during its evaluation. Often, information about the level of knowledge of the map label at the time of its evaluation is not provided with a confusion matrix. By explicitly separating the

TABLE 3. RESULTS OF THE DIFFERENCE OPERATOR, WHICH SHOWS BOTH THE MAGNITUDE AND FREQUENCY OF ERRORS ACROSS CATEGORIES

Map Label	Sites	Mismatches −4	−3	−2	−1	Matches 0	1	2	3	4	Arithmetic Mean
Class A	10	0	0	0	0	0	1	0	1	8	3.60
Class B	10	1	1	0	2	2	1	0	2	1	0.20
Class C	10	0	0	2	4	0	2	1	1	0	−0.10
Class D	10	0	0	1	3	4	0	0	2	0	0.10
Total	40	1	1	3	9	6	4	1	6	9	0.95

results for the MAX and RIGHT function, no confusion in this regard is possible.

Data from the MAX and RIGHT functions taken together are more useful than either taken separately. In this regard, the results in Table 2 for Class C are interesting. If just the MAX function were used, it would indicate Class C to be a major problem for the map. However, the results for the RIGHT function indicate that the impact on the user might not be as severe as the MAX function indicates. Class C is clearly a poor class, where the best answer is usually not provided, but an acceptable answer is given 80 percent of the time. Similarly, the strength of the result for Class A indicated by the RIGHT function is enhanced by the knowledge of its results for the MAX function.

Magnitude of Errors

One of the main benefits of the methods presented here based on fuzzy sets is the ability to evaluate the magnitude or seriousness of errors. In order to measure the magnitude of error, a function $\Delta\ S \rightarrow \mathbb{Z}$ is introduced. For a given site x, $\Delta(x)$ measures the difference between the score of the map category $\chi(x)$ of x and the highest score given to x among all other categories of $\mathcal{C}$. Formally, $\Delta(x)$ is defined as

$$\Delta(x) = \mu_{\chi(x)}(x) - \max_{\substack{C \in \mathcal{C} \\ C \neq \chi(x)}} \mu_C(x). \tag{1}$$

In the present study the linguistic values range between 1 and 5, so $-4 \leq \Delta(x) \leq 4$. For the ideal case, where the mapped category is perfectly right (score 5) and all other categories are absolutely wrong (score 1), the above defined function, called the DIFFERENCE function, yields 4. All sites that are matches using the MAX function have DIFFERENCE values greater than or equal to 0 and all mismatches are negative. A mismatch with a DIFFERENCE value of -1 would correspond to a case where the map label received a score only one less than the highest score given. Clearly, this kind of error is not as troublesome as those where a -4 is found. The results of the use of this function are dependent on the manner in which values are assigned for the membership functions. In our example, we have used a simple linear function to convert linguistic answers to values in a membership function, but there are many other possibilities. This approach lends to the situation where using this function, a difference of -2, produced when the mapped category received a score of 3 and the maximum score is 5 at a site, is equivalent to a situation in which the mapped category received a value of 2 and the maximum score is 4.

For each category $C \in \mathcal{C}$ and integer $-4 \leq i \leq 4$, we compute the quantity D_C^i the number of sites x such that $\chi(x)$

TABLE 4. SET MEMBERSHIPS ACROSS CATEGORIES USING MAX (SEE TEXT)

Map Label	Sites	MEMBERSHIP (χ) 0 T	M	N	1 T	M	N	2 T	M	N
Class A	10	0	0	0	9	9	0	1	1	0
Class B	10	0	0	0	6	3	3	4	3	1
Class C	10	0	0	0	1	1	0	9	3	6
Class D	10	0	0	0	3	2	1	7	4	3
Total	40	0	0	0	19	15	4	21	11	10

$= C$ and $\Delta(x) = i$. Table 3 illustrates the results of the DIFFERENCE function for the hypothetical test dataset of Table 1. The computed values D_c^i are shown under matches and mismatches (using the MAX function) columns in Table 3. Several trends are interesting to note. First, Classes B and D, which have similar frequencies of errors as shown by the MAX function, have quite different magnitudes for their errors. Class B generally has errors of higher magnitude than those in Class D. But Class D has a larger number of zero differences (sites where there is no difference between the map category and other categories). Also, as one might expect, Class A tends to have high positive DIFFERENCE values, while Class C rarely has large positive DIFFERENCE values. As the results from Table 3 imply, Class C has frequent negative difference values.

The last column shows the arithmetic mean of all scores for each class. It can be used as an indicator of the quality of a class. For example, the most accurate class, Class A, shows the higher positive arithmetic mean. Class B and D have lower arithmetic means. Class C has a negative arithmetic mean, indicating the larger number of errors in this class. The use of composite values for map categories based on these DIFFERENCE values may provide useful indices. It may even be appropriate to use a weighting scheme of some kind to accentuate the effects of the negative values. One possibility might be to average only the negative values to produce something like an error magnitude index for each category. The use of the DIFFERENCE function is dependent on the link between the linguistic scale used in evaluating map sites and the set membership function. In this paper, a simple discrete membership function is used, with unit differences between each level in the linguistic scale. Clearly, other possibilities exist for relating linguistic scales to set membership functions (Zadeh *et al.*, 1975; Lakoff, 1973) and future research in this area is warranted.

Source of Errors

Categorical errors in thematic maps frequently arise due to the heterogeneous nature of ground composition. This aspect of errors can be examined using a MEMBERSHIP function $\lambda : N \rightarrow \mathbb{Z}$ which measures the representation of multiple cover classes at each site as evaluated by the expert. The function selects those categories whose $\mu_C(x)$ is $\geq \tau$. Formally, it is defined as follows:

$$\lambda(x) = |\{C \mid C \in \mathcal{C} \text{ and } \mu_C(x) \geq \tau\}|, \qquad (2)$$

where τ is the threshold value defined in the section on Frequency of Matches and Mismatches. In the example described in Table 1, $|\mathcal{C}| = 4$, so $\lambda(x)$ ranges between 0 and 4. (Note that $\lambda(x)$ can equal 0 if $\tau > \mu_c(x)$ for all $\mathcal{C}$.)

The MEMBERSHIP function provides data on the frequency of set memberships at each site. One of the strengths of fuzzy sets is the ability to recognize multiple set memberships and grades of set membership. The five linguistic levels given represent degrees of set membership. The MEMBERSHIP function calculates the frequency of set membership for individual sites. In this case, some degree of right (i.e., values greater than or equal to 3) constitutes set membership.

For each category $C \in \mathcal{C}$ and integer $1 \leq i \leq 4$, we compute the number of sites such that $\chi(x) = C$ and $\lambda(x) = i$. These are estimated for the hypothetical data of Table 1 and shown in Table 4. A one-member site has a membership value greater than τ for only one category while a multiple-member site has membership values for two or more categories. The total number of sites (T) in each group is further distinguished into matched (M) and mismatched sites (N) using MAX. Nine out of a total of ten sites in Class A are single-member sites. All nine single-member sites are correctly matched as shown in the second data column marked "M" in the table, as is the two-member site. On the other hand, the least correct category, Class C, shows the following pattern. It has the least single-member sites. Its only single-member site is correctly matched (see column "M" under set membership 1). Class C has nine two-member sites. Six of these sites are mismatched, meaning that some other class has been given a higher score than C in these sites. The remaining cover classes, Class B and D, fall between the two extremes A and C (see Table 4). Neither zero membership nor three or four membership sites occur in this dataset.

The intent of the MEMBERSHIP function is to explore the possible sources of error in maps, which can be indicated in the frequency of matches and mismatches for the different numbers of set memberships. In the example presented, the proportion of matches is higher for the single membership sites than the multiple membership sites. This pattern is the opposite of what would be expected due to random effects. While the dataset used is hypothetical, the results mirror those found in our initial tests of this approach (Woodcock and Gopal, 1992). The use of the table is in giving some indication of the environmental conditions in which errors are being made. If there are significant numbers of mismatches (errors) in the single membership sites, then the mapping process is breaking down in the unambiguous locations. If the mismatches are concentrated in the multiple membership sites, efforts to improve the mapping methods should focus on the areas of heterogeneous cover and relative ambiguity between categories. Also, notice that the frequency of multiple memberships is not the same between categories. The high frequency of occurrence of multiple set memberships or the lack thereof can be an indication of the character of a map category to the users of a map.

If all sites had single set memberships, there would not be any need for the use of fuzzy sets and the methods presented in this paper. For those sites there is one unambiguous right answer. Thus, the frequency of multiple membership sites gives some indication of the need for the use of fuzzy sets.

Nature of Errors

One important kind of information is the categorical nature of the errors, or between which categories is their confusion. In a traditional analysis, this information is contained in the off-diagonal elements of an error or confusion matrix. Similar tables can be constructed using the CONFUSION and AMBIGUITY functions and fuzzy sets. The CONFUSION function $\zeta : N \rightarrow 2^{\mathcal{C}}$ identifies categories with a rating greater than the mapped category. In particular, for a site x, $\zeta(x)$ computes the set of categories whose membership values are greater than the membership value assigned to the map category $\chi(x)$.

$$\zeta(x) = \{C \mid C \in \mathcal{C} \text{ and } \mu_C(x) > \mu_{\chi(x)}(x)\}$$

It is identical to a traditional confusion matrix except for the fact that more than one category can have a rating (μ_C) higher than the mapped category ($\mu_{\chi(x)}(x)$) at a single site (x). For each (distinct) pair of categories $C, C' \in \mathcal{C}$ we compute the quantity $\zeta_{CC'}$, the number of sites x such that $\chi(x) = C$ and $\mu_{C'}(x) > \mu_C(x)$. The values of this CONFUSION function are the first values given in each cell of Table 5.

The AMBIGUITY function $\eta : N \rightarrow 2^{\mathcal{C}}$ identifies categories with the same rating as the mapped category. Formally:

$$\eta(x) = \{C \mid C \in \mathcal{C} \text{ and } \mu_C(x) = \mu_{\chi(x)}(x)\}$$

For each pair of (distinct) categories $C, C' \in \mathcal{C}$ we compute the quantity $\eta_{CC'}$, the number of sites x such that $\xi(x) = C$ and $\mu_{C'}(x) = \mu_C(x)$. The values for the AMBIGUITY function are given in the second column of each category in Table 5. The results are presented only where the membership values are some degree of right (values greater or equal to 3). The information on the equally rated or ambiguous categories will be particularly interesting for users of the maps if the matrix is not symmetric. The asymmetries indicate which categories more frequently include sites whose membership is ambiguous. From our example, Class D includes many sites rated equally with Class B, but Class B does not include any sites rated equally with Class D. The conclusion for a user of the map would be to expect sites in the ambiguous areas between Classes B and D to have been mapped as Class D.

TABLE 5. CONFUSION AND AMBIGUITY FUNCTIONS (SEE TEXT FOR DETAILS)

Map Label	Expert Evaluation								No. of mismatches	
	Class A		Class B		Class C		Class D			
	$\zeta_{CC'}$	$\eta_{CC'}$	$\zeta_{CC'}$	$\eta_{CC'}$	$\zeta_{CC'}$	$\eta_{CC'}$	$\zeta_{CC'}$	$\eta_{CC'}$	$\zeta_{CC'}$	$\eta_{CC'}$
Class A	X	X	0	0	0	0	0	0	0	0
Class B	0	0	X	X	4	2	0	0	4	2
Class C	0	0	3	0	X	X	5	0	8	0
Class D	0	0	2	4	3	0	X	X	5	4
Total	0	0	5	4	7	2	5	0	17	6

Discussion

To summarize, the proposed methods have several advantages over the traditional approach in assessing the accuracy of thematic maps. Some of the improvements are in the form of additional information about errors of the kind provided from traditional analyses. For frequency of errors, the combined use of the MAX and RIGHT functions constitutes an improvement. Similarly, using the CONFUSION and AMBIGUITY functions provides additional information on the nature of errors. There are also benefits obtained in the form of new kinds of information about errors. The DIFFERENCE measure yields information on the magnitude of errors, of a kind unobtainable from traditional approaches. The benefits associated with the use of the DIFFERENCE function may prove valuable in understanding the distribution of errors. The MEMBERSHIP function also provides information outside the domain possible using classical set theory. The MEMBERSHIP function provides indications of the kinds of environments in which errors are concentrated, which may help isolate the sources of errors in maps.

There are many issues and questions that remain to be addressed regarding the use of fuzzy sets in map accuracy assessment. Our hope is that this paper will stimulate interest in the use of fuzzy sets and that others will expand on our ideas. One area that needs attention concerns the ability to use these methods for the purposes of map comparison. Most of the information in the tables presented in this paper relates to the problem of understanding the errors in a single map. For comparing maps, additional research is needed on a couple of topics. First, a method of distilling the results to a single measure or index of map accuracy may be required as well as a way to produce a kappa-like statistic. A related concern is a way to standardize results from different experts. Some experts will be more lenient in the number of multiple set memberships given, and this factor will need to be taken into account. Another dimension of needed research concerns the ability to derive global estimates for map categories along the lines of the approaches given by Hay (1988) and Jupp (1989). Their methods could be used by simply reconstructing a traditional error matrix using the fuzzy data, but a method that took advantage of the additional information would be preferable.

The issue of errors and inaccuracy in a spatial database is an important research topic in GIS. Decisions made using the end products of GIS analysis are dependent on how errors propagate through the system. This research is concerned with one type of error — categorical or thematic error. The methods outlined in this paper describe how magnitude, nature, and source of these errors can be identified and measured using fuzzy sets. It fits into Level I and II of the "hierarchy of needs" model of GIS proposed by Veregin (1989). Until recently, error modeling has not received much attention (Goodchild *et al.*, 1992; Lanter and Veregin, 1992). Understanding the nature and distribution of errors in a map is a step in this direction. The paper has described various measures that can be used to analyze errors in greater detail. This research has implications in the area of error modeling and developing a model of error propagation in the GIS context, as its results can be easily incorporated to produce better and more predictive error models.

Spatial data is often imprecise and cannot be captured by boolean logic. The proposed methodology can cope with imprecise or ambiguous data. It permits subjective evaluation of experts to be incorporated into the design. The results of this research can be incorporated into error modeling and conveyed to the map user.

Conclusion

Categorical maps are commonly produced to represent complex geographical patterns. As such, it is essential that greater efforts be made to deal explicitly with the measurement of errors in categorical maps. This paper describes a new approach to assess the accuracy of a thematic map based on fuzzy sets. The suitability of fuzzy sets in the map accuracy context is demonstrated by deriving a set of measures to analyze the nature, frequency, source, and magnitude of errors. The approach is illustrated using a simple example, and the results obtained from it are compared with those of the traditional approach. The feasibility of the approach is further discussed using empirical data (Woodcock and Gopal, 1992).

The results of the analysis provide useful information to both the producer and user of thematic maps. Producers can improve methods of making the maps and presenting the information on accuracy and errors to the end user. Use of thematic maps will be enhanced by an improved understanding of their reliability.

Acknowledgments

The authors thank Vida Jakabhazy, Scott Macomber, Soren Ryherd, Ken Casey, Ralph Warbington, and Jack Levitan for their valuable assistance in collecting data. This research was supported in part by a grant from the Timber Management Office of Region 5 of the U.S. Forest Service. The authors also thank the anonymous reviewers, as the paper has been significantly improved as a result of their comments.

References

Aronoff, S., 1982a. Classification accuracy: A user approach. *Photogrammetric Engineering & Remote Sensing,* 48(8):1299–1307.

———, 1982b. The map accuracy report: A user's view. *Photogrammetric Engineering & Remote Sensing,* 48(8):1309–1312.

Burrough, P. A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science,* 40:477.

Card, A., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering & Remote Sensing,* 48(3):431–439.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of the Environment,* 37:35–46.

Congalton, R., R. Oderwald, and R. Mead, 1983. Assessing landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering & Remote Sensing,* 49(12):1671–1678.

Dubois, D., and H. Prade, 1980. *Fuzzy Sets and Systems: Theory and Applications. Academic Press,* New York.

Ericcson, K. A., and H. Simon, 1984. *Protocol Analysis. MIT Press,* Cambridge, Massachusetts.

Fisher, P., 1991. First experiments in viewshed uncertainty: The accuracy of the viewshed area. *Photogrammetric Engineering & Remote Sensing,* 57(10):1321–1327.

———, 1992. First experiments in viewshed uncertainty: Simulating fuzzy viewsheds. *Photogrammetric Engineering & Remote Sensing,* 58(3):345–352.

Fisher, P., and S. Pathirana, 1990. An evaluation of fuzzy membership of landcover classes in the suburban zone. *Remote Sensing of the Environment,* 34:121–132.

Goguen, J., 1969. The logic of inexact concepts. *Synthese,* 19:325–373.

Goodchild, M., S. Guoqing, and Y. Shiren, 1992. Development and test of an error model for categorical data. *International Journal of Geographical Information,* 6(2):87–104.

Hadipriono, F. C., J. G. Lyon, and L. Thomas, 1991. Expert opinion in satellite data interpretation. *Photogrammetric Engineering & Remote Sensing,* 57(1):75–78.

Hay, A.M., 1988. The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing,* 9:1392–1398.

Hoffer, R., and M. Fleming, 1978. *Mapping Vegetative Cover by Computer Aided Analysis of Satellite Data.* Purdue University, LARS Technical Report 011178.

Jupp, D.L.B., 1989. The stability of global estimates from confusion matrices. *International Journal of Remote Sensing,* 10(9):1563–1569.

Kaufmann, A., 1975. *Introduction to the Theory of Fuzzy Subsets* (translated from French by D. L. Swanson). Academic Press, New York.

Kent, J.T., and K.V. Mardia, 1988. Spatial classification using fuzzy membership models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 10(5):659–671.

Kollias, V.J., and A. Voliotis, 1991. Fuzzy reasoning in the development of geographical information systems. FRSIS: a prototype soil information system with fuzzy retrieval capabilities. *International Journal of Geographical Information,* 5(2):209–215.

Lakoff, G., 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic,* 2:458–508.

Lanter, D.P., and H. Veregin, 1992. A research paradigm for propagating error in layer-based GIS. *Photogrammetric Engineering & Remote Sensing,* 58(6):825–835.

Leung, Y. C., 1988. *Spatial Analysis and Planning Under Imprecision.* North Holland, Amsterdam.

Mead, R., and J. Szajgin, 1982. Landsat classification accuracy assessment procedures. *Photogrammetric Engineering & Remote Sensing,* 48(1):139–141.

Robinson, V. B., 1988. Some implications of fuzzy set theory applied to geographic databases. *Computers, Environment, and Urban Systems,* 12:89–97.

Robinson, V. B., and A. Strahler, 1984. Issues in designing geographic information systems under conditions of inexactness. *Machine Processing of Remotely Sensed Data Symposium,* pp. 198–204.

Rosenfield, G., 1981. Analysis of variance of thematic mapping experiment data. *Photogrammetric Engineering & Remote Sensing,* 47(12):1685–1692.

Rosenfield, G., and K. Fitzpatrick-Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering & Remote Sensing,* 52(2):223–227.

Story, M., and R. Congalton, 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering & Remote Sensing,* 52(3):377–399.

Trivedi, M.M., and J.C. Bezdek, 1986. Low-level segmentation of aerial images with fuzzy clustering. *IEEE Transactions on Systems, Man and Cybernetics,* 16(4):589–598.

van Genderen, J., and B. Lock, 1977. Testing land-use map accuracy. *Photogrammetric Engineering & Remote Sensing,* 43(9):1135–1137.

Veregin, H., 1989. Error modeling for the map overlay operation. *The Accuracy of Spatial Databases* (M. Goodchild and S. Gopal, editors), Taylor Francis, London, pp. 3–18.

Wang, F., 1990. Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering & Remote Sensing,* 56(8):1163–1169.

Wang, F., G. B. Hall, Subaryono, 1990. Fuzzy information representation and processing in conventional GIS software: database design and application. *International Journal of Geographical Information Systems,* 4:261–283.

Woodcock, C., and S. Gopal, 1992. Accuracy Assessment of the Stanislaus Forest Vegetation Map using Fuzzy Sets, *Proceeding of the Fourth Biennial Remote Sensing Applications Conference,* Orlando, Florida, April, 17 p.

Woodcock, C., and V. Harward, 1992. Nested hierarchical scene models and image segmentation. *International Journal of Remote Sensing,* 13(16):3167–3189.

Zadeh, L., 1965. Fuzzy sets. *Information and Control,* 8:338–353.

———, 1968. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications,* 10:421–427.

———, 1973. Outline of a new approach to the analysis of complex or imprecise concepts. *IEEE Transactions: Systems, Man and Cybernetics,* SMC-3:28–44.

Zadeh, L., K. Fu, K. Tanaka, and M. Shimura (editors), 1975. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes.* Academic Press, New York.

(Received 12 March 1992; revised and accepted 15 December 1992; revised 26 January 1993)