9 Fuzzy Accuracy Assessment

As our use of remotely sensed data and maps has grown in complexity, so have the classification schemes associated with these efforts. The classification scheme then becomes an even more important factor influencing the accuracy of the entire project. A review of the accuracy assessment literature points out some of the limitations of using only the traditional error matrix approach to accuracy assessment of a map with a complex classification scheme. Congalton and Green (1993) recommend the error matrix as a jumping-off point for identifying sources of confusion (i.e., differences between the map created from remotely sensed data and the reference data) and not simply the "error." For example, variation in human interpretation can have a significant impact on what is considered correct. If photo interpretation is used as the source of the reference data and that interpretation is flawed, then the results of the accuracy assessment could be very misleading. This is true even for observations made in the field. As classification schemes become more complex, more variation in human interpretation is introduced. In addition, in situations where the breaks (i.e., divisions between classes) in the classification system represent artificial distinctions along a continuum, variation in human interpretation is often very difficult to control and, while unavoidable, can have profound effects on accuracy assessment results (Congalton, 1991; Congalton and Green, 1993). Several researchers have noted the impact of the variation in human interpretation on map results and accuracy assessment (Gong and Chen, 1992; Lowell, 1992; McGwire, 1992; Congalton and Biging, 1992).

Gopal and Woodcock (1994) proposed the use of fuzzy sets to "allow for explicit recognition of the possibility that ambiguity might exist regarding the appropriate map label for some locations on the map. The situation of one category being exactly right and all other categories being equally and exactly wrong often does not exist." In this fuzzy set approach, it is recognized that instead of a simple system of correct (agreement) and incorrect (disagreement), there can be a variety of responses such as "absolutely right," "good answer," "acceptable," "understandable but wrong," and "absolutely wrong."

Fuzzy set theory or fuzzy logic is a form of set theory. Although initially introduced in the 1920s, fuzzy logic gained its name and its algebra in the 1960s and 1970s from Zadeh (1965), who developed fuzzy set theory as a way to characterize the ability of the human brain to deal with vague relationships. The key concept is that membership in a class is a matter of degree. Fuzzy logic recognizes that, on the margins of classes that divide a continuum, an item may belong to both classes. As Woodcock and Gopal (1992) state, "The assumption underlying fuzzy set theory is that the transition from membership to non-membership is seldom a step function." Therefore, while a 100% hardwood stand can be labeled hardwood, and a 100%

conifer stand may be labeled conifer, a 49% hardwood and 51% conifer stand may be acceptable if labeled either conifer or hardwood.

Lowell (1992) calls for "a new model of space which shows transition zones for boundaries, and polygon attributes as indefinite." As Congalton and Biging (1992) conclude in their study of the validation of photo interpreted stand type maps, "The differences in how interpreters delineated stand boundaries was most surprising. We were expecting some shifts in position, but nothing to the extent that we witnessed. This result again demonstrates just how variable forests are and is a strong indicator of human variation in photo interpretation."

There are a number of techniques that have been proposed to incorporate ambiguity or fuzziness into the accuracy assessment process. Three methods are presented in this chapter including (1) expanding the major diagonal of the error matrix, (2) measuring map class variability, and (3) using a fuzzy error matrix approach.

EXPANDING THE MAJOR DIAGONAL OF THE ERROR MATRIX

The simplest and most straightforward method for incorporating fuzziness into the accuracy assessment process is to expand the major diagonal of the error matrix. Remember that the major diagonal of the error matrix represents agreement between the reference data and the map, and is represented by a single cell in the matrix for each map class. By acknowledging some fuzziness in the classification, the class boundaries may be expanded to accept as correct plus or minus one class of the actual class. In other words, the major diagonal is no longer just a single cell for each map class, but rather wider. This method works well if the classification scheme is continuous, such as elevation or tree size class or forest crown closure. If the classification scheme is discrete, such as in a vegetation or land cover mapping project, then this method probably cannot be used.

Table 9.1 presents the traditional error matrix for a classification of forest crown closure (a continuous classification scheme divided into 6 discrete classes). Only exact matches are considered correct; they are tallied along the major diagonal. The overall accuracy of this classification is 40%. Table 9.2 presents the same error matrix with the major diagonal expanded to include plus or minus one crown closure class. In other words, for crown closure class 3 both crown closure classes 2 and 4 are also accepted as correct. This revised major diagonal then results in a tremendous increase in overall accuracy to 75%.

The advantage of using this method of accounting for fuzzy class boundaries is obvious; the accuracy of the classification can increase dramatically. The disadvantage is that if the reason for accepting plus or minus one class cannot be adequately justified or does not meet the map user's requirements, then it may be thought that you are cheating to try to get higher accuracies. Therefore, although this method is very simple to apply, it should only be used when agreement exists that it is a reasonable course of action. The other techniques described next may be more difficult to apply, but are easier to justify.

TABLE 9.1
Error Matrix Showing the Ground Reference Data versus the Image Classification for Forest Crown Closure

Image Classification	Ground Reference 1	2	3	4	5	6	Row Total
1	2	9	1	2	1	1	16
2	2	8	3	6	1	1	21
3	0	3	3	4	9	1	20
4	0	0	2	8	7	10	27
5	0	1	2	1	6	16	26
6	0	0	0	0	3	31	34
Column Total	4	21	11	21	27	60	144

Crown Closure Categories

Class 1 = 0% CC
Class 2 = 1–10% CC
Class 3 = 11–30% CC
Class 4 = 31–50% CC
Class 5 = 51–70% CC
Class 6 = 71–100% CC

OVERALL ACCURACY = 58/144 = 40%

PRODUCER'S ACCURACY

Class 1 = 2/4 = 50%
Class 2 = 8/21 = 38%
Class 3 = 3/11 = 27%
Class 4 = 8/21 = 38%
Class 5 = 6/27 = 22%
Class 6 = 31/60 = 52%

USER'S ACCURACY

Class 1 = 2/16 = 13%
Class 2 = 8/21 = 38%
Class 3 = 3/20 = 15%
Class 4 = 8/27 = 30%
Class 5 = 6/26 = 23%
Class 6 = 31/34 = 91%

MEASURING MAP CLASS VARIABILITY

The second method for incorporating fuzziness into the accuracy assessment process is not as simple as expanding the major diagonal of the error matrix. While it is difficult to control variation in human interpretation, it is possible to measure the variation, and to use these measurements to compensate for differences between reference and map data that are caused not by map error but by variation in interpretation. There are two options available to control the variation in human interpretation to reduce the impact of this variation on map accuracy. One is to measure each reference site with great precision to minimize the variance in the reference site labels. This method can be prohibitively expensive, usually requiring extensive field sampling and detailed measurements. The second option is to measure the variance and use these measurements to compensate for nonerror differences between reference and map data. Measuring the variance requires having multiple analysts assess each reference site. This assessment

TABLE 9.2
Error Matrix Showing the Ground Reference Data versus the Image Classification for Forest Crown Closure within Plus or Minus One Tolerance Class

Image Classification	Ground Reference 1	2	3	4	5	6	Row Total
1	2	9	1	2	1	1	16
2	2	8	3	6	1	1	21
3	0	3	3	4	9	1	20
4	0	0	2	8	7	10	27
5	0	1	2	1	6	16	26
6	0	0	0	0	3	31	34
Column Total	4	20	11	21	27	60	144

Crown Closure Categories

Class 1 = 0% CC
Class 2 = 1–10% CC
Class 3 = 11–30% CC
Class 4 = 31–50% CC
Class 5 = 51–70% CC
Class 6 = 71–100% CC

OVERALL ACCURACY = 108/144 = 75%

PRODUCER'S ACCURACY

Class 1 = 4/4 = 100%
Class 2 = 20/21 = 95%
Class 3 = 8/11 = 73%
Class 4 = 13/21 = 62%
Class 5 = 16/27 = 59%
Class 6 = 47/60 = 78%

USER'S ACCURACY

Class 1 = 11/16 = 69%
Class 2 = 13/21 = 62%
Class 3 = 10/20 = 50%
Class 4 = 17/27 = 63%
Class 5 = 23/26 = 88%
Class 6 = 34/34 = 100%

could be done by field visitation or using photo interpretation, and requires an objective and repeatable method of capturing the impacts of human variation. The collection of reference data for accuracy assessment is an expensive component of any mapping project; multiple visits to every reference site to capture variation may be prohibitively expensive. Therefore, while theoretically possible, measuring map class variability is not a viable component of most remotely sensed data mapping projects.

THE FUZZY ERROR MATRIX APPROACH

The previous approaches of expanding the major diagonal to incorporate fuzziness in the accuracy assessment process may be hard to justify, and the effort needed to measure the variability may be cost-prohibitive. Therefore, another method is required to incorporate fuzziness into the map accuracy assessment process. As mentioned earlier, the difficult task in using fuzzy logic is the development of the specific rules for its application. Fuzzy systems often rely on experts for development of these rules. Hill (1993) developed an arbitrary but practical fuzzy set rule that determined "sliding class widths" for assessing the accuracy of maps produced for The California Department of Forestry and Fire Protection of the Klamath Province in Northwestern California. Woodcock and Gopal (1992) relied on experts in their application of fuzzy sets to assess the accuracy of maps generated for Region 5 of the U.S. Forest Service. While both of their methods incorporated fuzziness into the accuracy assessment process, neither used an error matrix approach. Instead, a number of other metrics to represent map accuracy and agreement were computed.

The Fuzzy Error Matrix

Given the wide acceptance of the error matrix as the standard for reporting the accuracy of thematic maps, it would be far better to employ some approach that combines both the error matrix and some measure of fuzziness. Such a technique, called the *fuzzy error matrix approach*, was introduced by Green and Congalton (2004) and is described here. The use of the fuzzy error matrix is a very powerful tool in the accuracy assessment process because the fuzzy error matrix allows the analyst to compensate for situations in which the classification scheme breaks represent artificial distinctions along a continuum of landcover and/or where observer variability is often difficult to control. While one of the assumptions of the traditional or deterministic error matrix used in the rest of this book is that a reference data sample site can have only one label, this is not the case with the fuzzy error matrix approach.

Let us continue with the example used so far in this chapter. Table 9.3 presents a fuzzy error matrix generated from a set of fuzzy rules applied to the same classification that was used to generate the deterministic (i.e., nonfuzzy) error matrix that was presented in Table 9.1. In this case, the classification was defined using the following fuzzy rules:

- Class 1 was defined as always 0% crown closure. If the reference data indicated a value of 0%, then only a map classification of 0% was accepted.
- Class 2 was defined as acceptable if the reference data was within 5% of that of the map classification. In other words, if the reference data indicates that a sample has 15% crown closure and the map classification put it in Class 2, the answer would not be absolutely correct, but would be considered acceptable.
- Classes 3–6 were defined as acceptable if the reference data were within 10% of that of the map classification. In other words, a sample classified as Class 4 on the image, but found to be 55% crown closure on the reference data would be considered acceptable.

As a result of these fuzzy rules, off-diagonal elements in the matrix contain two separate values. The first value in the off-diagonal represents those labels that, although not absolutely correct, are considered acceptable labels within the fuzzy rules. The second value indicates those labels that are still unacceptable (i.e., wrong). The major diagonal still only tallies those labels considered to be absolutely correct. Therefore, in order to compute the accuracies (overall, producer's, and user's), the values along

TABLE 9.3
Error Matrix Showing the Ground Reference Data versus the Image Classification for Forest Crown Closure Using the Fuzzy Logic Rules

Image Classification \ Ground Reference	1	2	3	4	5	6	Row Total
1	2	6,3	1	2	1	1	16
2	0,2	8	2,1	6	1	1	21
3	0	2,1	3	4,0	9	1	20
4	0	0	0,2	8	5,2	10	27
5	0	1	2	1,0	6	12,4	26
6	0	0	0	0	2,1	31	34
Column Total	4	21	11	21	27	60	144

Crown Closure Categories

Class 1 = 0% CC
Class 2 = 1–10% CC
Class 3 = 11–30% CC
Class 4 = 31–50% CC
Class 5 = 51–70% CC
Class 6 = 71–100% CC

OVERALL ACCURACY = 92/144 = 64%

PRODUCER'S ACCURACY

Class 1 = 2/4 = 50%
Class 2 = 16/21 = 76%
Class 3 = 5/11 = 45%
Class 4 = 13/21 = 62%
Class 5 = 13/27 = 48%
Class 6 = 43/60 = 72%

USER'S ACCURACY

Class 1 = 8/16 = 50%
Class 2 = 10/21 = 48%
Class 3 = 9/20 = 45%
Class 4 = 13/27 = 48%
Class 5 = 19/26 = 73%
Class 6 = 33/34 = 97%

the major diagonal (i.e., absolutely correct) and those deemed acceptable (i.e., those in the first value) in the off-diagonal elements are combined. In Table 9.3, this combination of absolutely correct and acceptable answers results in an overall accuracy of 64%. This overall accuracy is significantly higher than the original error matrix (Table 9.1), but not as high as Table 9.2.

It is much easier to justify the fuzzy rules used in generating Table 9.3 than it is to simply extend the major diagonal to plus or minus one whole class as was done in Table 9.2. For crown closure, it is recognized that mapping typically varies by plus or minus 10% (Spurr, 1948). Therefore, it is reasonable to define as acceptable a range within 10% for classes 3–6. Class 1 and class 2 take an even more conservative approach and are therefore even easier to justify.

In addition to this fuzzy set theory working for continuous variables such as crown closure, it also applies to more categorical data. All that is required is a set of fuzzy rules to explain or capture the variation. For example, in the hardwood range area of California, many land cover types differ only by which hardwood species is dominant. In many cases, the same species are present and the specific land cover type is determined by which species is most abundant. Also, in some of these situations, the species look very much alike on aerial photography and on the ground. It is clear that there is a great deal of acceptable and unavoidable variation in mapping the hardwood range.

In a worldwide mapping effort funded by the National Geospatial-Intelligence Agency (NGA) using Landsat imagery, no ground visitation was possible for collecting the reference data. In some areas of the world, the imagery used for the reference data was of such low resolution as to make interpretation of the individual classes very difficult. The use of this fuzzy error matrix approach was the only viable solution in this case (Green and Congalton, 2004). A traditional, deterministic accuracy assessment conducted with such highly variable reference data would have unfairly represented the accuracy of this mapping effort.

Therefore, in many situations, the use of these fuzzy rules, which allow for the incorporation of acceptable reference labels in addition to the absolutely correct reference labels into the construction of the error matrix, makes a great deal of sense. Using fuzzy rules in an error matrix approach combines all the established descriptive and analytical power of the error matrix while incorporating variation into the assessment.

Implementation of the Fuzzy Error Matrix

Implementation of the fuzzy error matrix approach is greatly simplified with the use of a special reference data collection form (Figure 9.1). Each reference site can be evaluated for the likelihood of being identified as each of the possible map classes

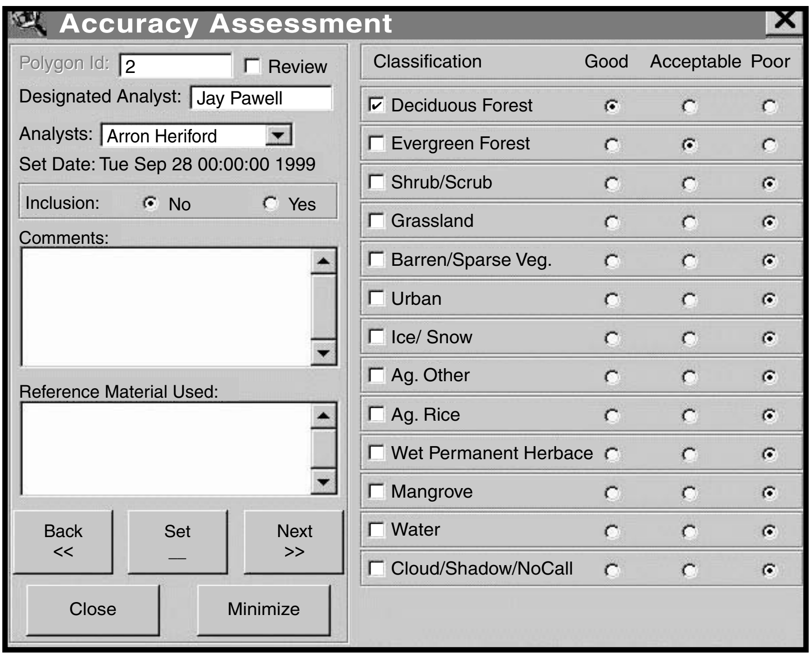

FIGURE 9.1 Form for labeling accuracy assessment reference sites.

given the fuzzy rules for that project. First, the analyst determines the most appropriate ("good") label for the site and enters this label in the appropriate box under the "classification" column on the form. This label determines which row of the matrix the site will be tallied in and is also the value used for calculation of the deterministic error matrix. After assigning the most appropriate label for the site, the remaining possible map labels are evaluated as either "acceptable" or "poor" candidates for the site's label, again as indicated by the fuzzy rules. For example, a site might fall near the classification scheme margin between deciduous forest and evergreen forest because of the exact mix of species and/or the difficulty interpreting the exact mixture on the reference data imagery. In this instance, the analyst might rate deciduous forest as the most appropriate label (i.e., "good"), but also rate evergreen forest as "acceptable" (see Figure 9.1). In this case, no other map classes would be acceptable; all the others would be rated as "poor."

Using this fuzzy error matrix approach allows for the analyst to compensate for interpreter variability and difficulty in determining only a single label for each reference data sample site. While this method can be used for any assessment, it works best when (1) there are issues in collecting good reference data because of limitations in the reference data collection methods, (2) when interpreter variability cannot be controlled, or (3) when the ecosystem being mapped is highly heterogeneous. If there is little variation or fuzziness in the classification scheme or if detailed measurements can be taken to minimize the variation, then there may be little need for this approach. However, in most projects creating maps from remotely sensed imagery, the use of the fuzzy error matrix approach can significantly help to incorporate variation inherent in the project.

Another Fuzzy Error Matrix Example

Table 9.4 shows a fuzzy error matrix for a categorical classification scheme (i.e., a land cover mapping project). Again, the power of this approach lies in the ability to compute the same descriptive metrics as in the traditional deterministic error matrix. Computation of the overall, producer's and user's accuracy statistics for the fuzzy error matrix follows the same methodology as the traditional deterministic error matrix with the following additions. Nondiagonal cells in the matrix contain two tallies, which can be used to distinguish class labels that are uncertain or that fall on class margins, from class labels that are most probably in error. The first number represents those sites in which the map label matched an "acceptable" reference label in the fuzzy assessment (Table 9.4). Therefore, even though the label was not considered the most appropriate, it was considered acceptable given the fuzziness of the classification system and/or the minimal quality of some of the reference data. These sites are considered a "match" for estimating fuzzy accuracy. The second number in the cell represents those sites where the map label was considered poor (i.e., an error).

The fuzzy assessment overall accuracy is estimated as the percentage of sites where the "good" and "acceptable" reference labels matched the map label. Producer's and user's accuracies are computed in the traditional way, but again instead of just using

TABLE 9.4
Example of Fuzzy Error Matrix Showing Both Deterministic and Fuzzy Accuracy Assessment

	REFERENCE DATA								User's Accuracies			
MAP	Decid. Forest	EG Forest	Scrub/ Shrub	Grass	Barren/ Sparse	Urban	Agric.	Water	Deterministic Totals	Percent Deterministic	Fuzzy Totals	Percent Fuzzy
Deciduous Forest	**48**	24,7	0,1	0,3	0,0	0,1	0,11	0,18	48/113	42.5%	72/113	63.7%
Evergreen Forest	4,0	**17**	0,1	0,0	0,0	0,0	0,1	0,3	17/26	65.4%	21/26	80.8%
Shrub/Scrub	2,0	0,1	**15**	8,1	0,0	0,0	2,2	0,0	15/31	48.4%	27/31	87.1%
Grassland	0,1	0,0	5,1	**14**	0,0	0,0	3,0	0,0	14/24	58.3%	22/24	91.7%
Barren/Sparse Veg.	0,0	0,0	0,2	0,0	**0**	0,0	0,1	0,0	0/3	0.0%	0/3	0.0%
Urban	0,0	0,0	0,0	0,0	0,0	**20**	2,0	0,0	20/22	90.9%	22/22	100.0%
Agriculture	0,1	0,1	7,15	18,6	0,0	2,0	**29**	1,2	29/82	35.4%	57/82	69.5%
Water	0,0	0,0	0,0	0,0	0,0	0,0	0,0	**8**	8/8	100.0%	8/8	100.0%
Producer's Accuracies												
Deterministic Totals	48/56	17/50	15/47	14/50	NA	20/24	29/51	8/33	**Overall Accuracies**			
Percent Deterministic	85.7%	34.0%	31.9%	28.0%	NA	83.3%	56.9%	24.2%	**Deterministic**		**Fuzzy**	
Fuzzy Totals	54/56	41/50	27/47	40/50	NA	22/24	36/51	10/33	151/311	48.6%	230/311	74.0%
Percent Fuzzy	96.4%	82.0%	57.4%	80.0%	NA	91.7%	70.6%	30.3%				

the value on the major diagonal ("good"), the value in the first off-diagonal position ("acceptable") is also included (Table 9.4).

SUMMARY

While three methods are presented in this chapter for dealing with variation or fuzziness in the accuracy assessment process, the fuzzy error matrix approach is by far the most useful and operational. The elegance of this approach is that it combines all of the power of the error matrix, including computing overall, producer's, and user's accuracies, with the ability to incorporate the variation inherent in many classification schemes or resulting from the reference data collection process. Given that the matrix contains the information to compute both the traditional deterministic accuracy measures and fuzzy accuracy measures, there is strong impetus to use this approach. It is highly recommended that this approach be considered whenever map class variation or variation in the reference data collection process is a significant issue in the mapping project.