4 Thematic Accuracy

The major focus of this book is thematic accuracy assessment. Chapter 3 presented a summary of positional accuracy assessment and the standard measure for reporting it, RMSE. This chapter introduces the most widely accepted measure for representing thematic accuracy, the error matrix. The chapter also documents the evolution of thematic accuracy assessment, beginning with a discussion of early non-site-specific assessments. Next, site-specific assessment techniques employing the error matrix are presented, followed by the mathematical representation of the error matrix.

NON-SITE-SPECIFIC ASSESSMENTS

In a non-site-specific accuracy assessment, only total areas for each category mapped are computed, without regard to the location of these areas. In other words, a comparison between the number of acres or hectares of each category on the map generated from remotely sensed data and the reference data is performed. In this way, the errors of omission and commission tend to compensate for one another and the totals compare favorably. However, nothing is known about any specific location on the map or how it agrees or disagrees with the reference data.

A simple example quickly demonstrates the shortcomings of the non-site-specific approach. Figure 4.1 shows the distribution of the forest category on both a reference image and two different classifications generated from remotely sensed data. Classification #1 was generated using one type of classification algorithm (e.g., supervised, unsupervised, or nonparametric, etc.), while classification #2 employed a different algorithm. In this example, only the forest category is being compared. The reference data shows a total of 2,435 acres of forest, while classification #1 shows 2,322 acres and classification #2 shows 2,635 acres. In a non-site-specific assessment, you would conclude that classification #1 is better for the forest category because the total number of forest acres for classification #1 more closely agrees with the number of acres of forest on the reference image (2,435 acres − 2,322 acres = 113 acres difference for classification #1, while classification #2 differs by 200 acres). However, a visual comparison (see Figure 4.2) between the forest polygons on classification #1 and the reference data demonstrates little locational correspondence. Classification #2, despite being judged inferior by the non-site-specific assessment, appears to locationally agree much better with the reference data forest polygons (see Figure 4.2). Therefore, the use of non-site-specific accuracy assessment can be quite misleading. In the example shown here, the non-site-specific assessment actually recommends the use of the inferior classification algorithm.

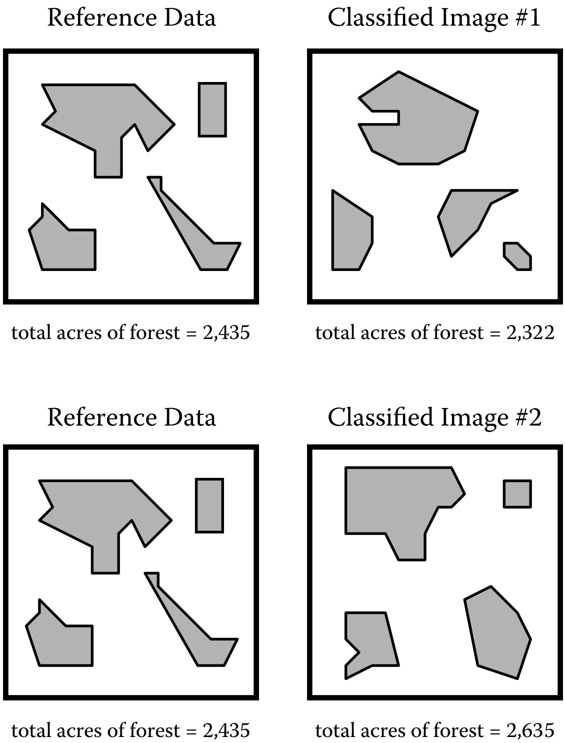

FIGURE 4.1 Example of non-site-specific accuracy assessment.

SITE-SPECIFIC ASSESSMENTS

Given the obvious limitations of non-site-specific accuracy assessment, there was a need to know how the map generated from the remotely sensed data compared to the reference data on a locational basis. Therefore, site-specific assessments were instituted. Initially, a single value representing the accuracy of the entire classification (i.e., overall accuracy) was presented. This computation was performed by comparing a sample of locations on the map with the same locations on the reference data and keeping track of the number of times there was agreement.

An overall accuracy level of 85% was adopted as representing the cutoff between acceptable and unacceptable results. This standard was first described in Anderson et. al (1976) and seems to be almost universally accepted despite there being nothing magical or even especially significant about the 85% correct accuracy level. Obviously, the accuracy of a map depends on a great many factors including the amount of effort, the level of detail (i.e., classification scheme), and the variability of the categories to be mapped. In some applications an overall accuracy of 85% is more than sufficient and in other cases it would not be accurate enough.

While having a single number to measure overall thematic map accuracy was an improvement over the non-site-specific assessment method, it was soon realized

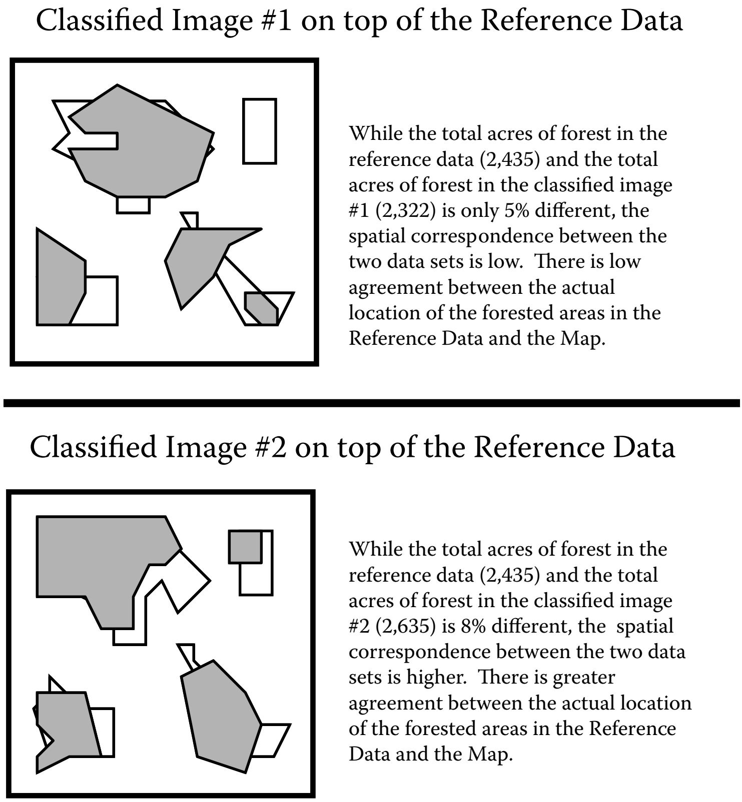

FIGURE 4.2 Spatial correspondence for the non-site-specific accuracy assessment example.

that this single number was not enough. The need to evaluate individual categories within the classification scheme was recognized, and thus began the use of the error matrix to represent map accuracy.

The Error Matrix

As previously introduced, an error matrix is a square array of numbers set out in rows and columns that expresses the number of sample units assigned to a particular category in one classification relative to the number of sample units assigned to a particular category in another classification (Table 4.1). In most cases, one of the classifications is considered to be correct (i.e., the reference data) and may be generated from aerial photography, airborne video, ground observation, or ground measurement. The columns usually represent this reference data, while the rows indicate the classification generated from the remotely sensed data (i.e., the map). It should be noted that the reference data has often been referred to as the "ground truth" data. Now, while it is true that the reference data are assumed to be more correct than the map it is being used to assess, it is by no means true that these data are perfect

TABLE 4.1

Example Error Matrix (Same as that Presented in Figure 2.6)

		Reference Data D	C	AG	SB	Row Total
Classified Data	D	65	4	22	24	115
	C	6	81	5	8	100
	AG	0	11	85	19	115
	SB	4	7	3	90	104
	Column Total	75	103	115	141	434

Land Cover Categories

D = deciduous
C = conifer
AG = agriculture
SB = shrub

OVERALL ACCURACY = (65 + 81 + 85 + 90)/434 = 321/434 = 74%

PRODUCER'S ACCURACY

D = 65/75 = 87%
C = 81/103 = 79%
AG = 85/115 = 74%
SB = 90/141 = 64%

USER'S ACCURACY

D = 65/115 = 57%
C = 81/100 = 81%
AG = 85/115 = 74%
SB = 90/104 = 87%

or represent "the truth." Therefore, the term "ground truth" is inappropriate and, in some cases, very misleading. Throughout this book, the authors will use the term *reference data* to identify the data being used to compare to the map generated from remotely sensed data (i.e., the map).

An error matrix is a very effective way to represent map accuracy in that the individual accuracies of each category are plainly described along with both the errors of inclusion (commission errors) and errors of exclusion (omission errors) present in the classification. A commission error is simply defined as including an area in a category when it does not belong to that category. An omission error is excluding an area from the category to which it belongs. Each and every error is an omission from the correct category and a commission to a wrong category.

For example, in the error matrix in Table 4.1, there are four areas that were classified as deciduous when the reference data shows that they were actually conifer. Therefore, four areas were omitted from the correct coniferous category and committed to the incorrect deciduous category. In addition to clearly showing errors of omission and commission, the error matrix can be used to compute other accuracy measures such as overall accuracy, producer's accuracy, and user's accuracy (Story and Congalton, 1986). Overall accuracy is simply the sum of the major diagonal (i.e., the correctly classified sample units) divided by the total number of sample units in the entire error matrix. This value is the most commonly reported accuracy assessment statistic and is probably most familiar to the reader. However, just presenting the overall accuracy is not enough. It is important to present the entire matrix so that other accuracy measures can be computed as needed and confusion between map classes is clearly presented and understood.

Producer's and user's accuracies are ways of representing individual category accuracies instead of just the overall classification accuracy, and were introduced by Story and Congalton (1986). Before error matrices became the standard accuracy reporting mechanism, it was common to report the overall accuracy and either only the producer's or user's accuracy. Sometimes, only the higher of the two accuracies (between the producer's and user's accuracies) was selected to be reported, resulting in misleading information about the map accuracy. A quick example will demonstrate the need to publish the entire matrix so that all three accuracy measures can be computed.

Studying the error matrix shown in Table 4.1 reveals an overall map accuracy of 74%. However, suppose we are most interested in the ability to classify hardwood forests so we calculate a "producer's accuracy" for this category. This calculation is performed by dividing the total number of correct sample units in the deciduous category (i.e., 65) by the total number of deciduous sample units as indicated by the reference data (i.e., 75 or the column total). This division results in a "producer's accuracy" of 87%, which is quite good. If we stopped here, one might conclude that although this classification appears to be average overall, it is more than adequate for the deciduous category. Drawing such a conclusion could be a very serious mistake. A quick calculation of the "user's accuracy" computed by dividing the total number of correct sample units in the deciduous category (i.e., 65) by the total number of sample units classified as deciduous (i.e., 115 or the row total) reveals a value of 57%. In other words, although 87% of the deciduous areas have been correctly identified as deciduous, only 57% of the areas called deciduous on the map are actually deciduous on the ground. The high producer's accuracy occurs because too much of the map is labeled deciduous. A more careful look at the error matrix reveals that there is significant confusion in discriminating deciduous from barren and shrub. Therefore, although the producer of this map can claim that 87% of the time an area that was deciduous on the ground was identified as such on the map, a user of this map will find that only 57% of the time that the map says an area is deciduous will it actually be deciduous on the ground.

Mathematical Representation of the Error Matrix

This subsection presents the error matrix in mathematical terms necessary to perform the analysis techniques described in the Chapter 7. The error matrix was presented previously in descriptive terms, including an example (Table 4.1) that should help the reader make this transition to equations and mathematical notation easier to understand.

Assume that n samples are distributed into k^2 cells, here each sample is assigned to one of k categories in the map (usually the rows), and independently to one of the same k categories in the reference data set (usually the columns). Let n_{ij} denote the number of samples classified into category i ($i = 1, 2, \ldots, k$) in the map and category j ($j = 1, 2, \ldots, k$) in the reference data set (Table 4.2).

TABLE 4.2
Mathematical Example of an Error Matrix

		j = Columns (Reference) 1	2	k	Row Total n_{i+}
i = Rows (Classification)	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{k1}	n_{k2}	n_{kk}	n_{k+}
Column Total n_{+j}		n_{+1}	n_{+2}	n_{+k}	n

Let

$$n_{i+} = \sum_{j=1}^{k} n_{ij}$$

be the number of samples classified into category i in the remotely sensed classification, and

$$n_{+j} = \sum_{i=1}^{k} n_{ij}$$

be the number of samples classified into category j in the reference data set.

Overall accuracy between remotely sensed classification and the reference data can then be computed as follows:

$$\text{overall accuracy} = \frac{\sum_{i=1}^{k} n_{ii}}{n}.$$

Producer's accuracy can be computed by

$$\text{producer's accuracy } j = \frac{n_{jj}}{n_{+j}}$$

and the user's accuracy can be computed by

$$\text{user's accuracy}_i = \frac{n_{ii}}{n_{i+}}$$

Finally, let p_{ij} denote the proportion of samples in the i, jth cell, corresponding to n_{ij}. In other words, $p_{ij} = n_{ij}/n$.

Then let p_{i+} and p_{+j} be defined by

$$p_{i+} = \sum_{j=1}^{k} p_{ij}$$

and

$$p_{+j} = \sum_{i=1}^{k} p_{ij}$$

This mathematical representation of the error matrix takes a little practice to get used to. Actually, understanding an error matrix the very first time can take a little effort. However, given the importance of the error matrix in thematic accuracy assessment and the need for the mathematical representation for some of the analysis techniques, readers are encouraged to spend a little time here until they feel comfortable. Many examples will be provided throughout the book, as well as some case studies, to aid every reader in becoming an error matrix expert.