Remote Sensing Brief

Accuracy Assessment: A User's Perspective

Michael Story
Science Applications Research, Lanham, MD 20706
Russell G. Congalton
Department of Forestry and Resource Management, University of California, Berkeley, CA 94720

MUCH HAS recently been written about accuracies of images and maps derived from remotely sensed data. These studies have addressed errors caused by preprocessing (Smith and Kovalick, 1985), by interpretive techniques both manual (Congalton and Mead, 1983) and automated (Story *et al.*, 1984; Congalton and Rekas, 1985), by the imaging system (Williams *et al.*, 1983), and by techniques for sampling, calculating accuracy, and comparing results (Hord and Brooner, 1976; van Genderen and Lock, 1977; Ginevan, 1979; Hay, 1979; Aronoff, 1982; Congalton *et al.*, 1983). The most common way to express the accuracy of such images/maps is by a statement of the percentage of the map area that has been correctly classified when compared with reference data or "ground truth." This statement is usually derived from a tally of the correctness of the classification generated by sampling the classified data, and expressed in the form of an error matrix (sometimes called a confusion matrix or contingency table) (Table 1). In this kind of tally, the reference data (usually represented by the columns of the matrix) are compared to the classified data (usually represented by the rows). The major diagonal indicates the agreement between these two data sets. Overall accuracy for a particular classified image/map is then calculated by dividing the sum of the entries that form the major diagonal (i.e., the number of correct classifications) by the total number of samples taken.

More detailed statements of accuracy are often derived from the error matrix in the form of individual land-use/land-cover category accuracies. The reason for this additional assessment is obvious. If a classified image/map is stated to have an overall accuracy of 73 percent, the value represents the accuracy of the entire product. It does not indicate how the accuracy is distributed across the individual categories. The categories could, and frequently do, exhibit drastically differing accuracies, and yet combine for equivalent or similar overall accuracies. Individual category accuracies are, therefore, needed in order to completely assess the value of the classified image/map for a specific application.

An examination of the error matrix suggests at least two methods for determining individual category accuracies. The most common and accepted method is to divide the number of correctly classified samples of category *X* by the number of category *X* samples in the reference data (column total

TABLE 1. AN EXAMPLE ERROR MATRIX SHOWING ROW, COLUMN, AND GRAND TOTALS.

		Reference Data X	Reference Data Y	Reference Data Z	Row Total	
Classified Data	X	15	2	4	21	Sum of the major diagonal = 41
Classified Data	Y	3	12	2	17	Overall Accuracy = 41/56 = 73%
Classified Data	Z	1	3	14	18	
Column Total		19	17	20	56	

0099-1112/86/5203-397$02.25/0

TABLE 2. A NUMERICAL EXAMPLE SHOWING PRODUCER'S AND USER'S ACCURACIES

		Reference Data			
		F	W	U	Row Total
Classified Data	F	28	14	15	57
	W	1	15	5	21
	U	1	1	20	22
Column Total		30	30	40	100

Sum of the major diagonal = 63

Overall Accuracy = 63/100 = 63%

Producer's Accuracy	User's Accuracy
F = 28/30 = 93%	F = 28/57 = 49%
W = 15/30 = 50%	W = 15/21 = 71%
U = 20/40 = 50%	U = 20/22 = 91%

for category *X*). An alternate method is to divide the number of correctly classified samples of category *X* by the total number of samples classified as category *X* (row total for category *X*). It is important to understand that these two methods can result in very different assessments of the accuracy of category *X*. It is also important to understand the interpretation of each value.

In the traditional accuracy calculation, the number of correctly classified samples of category *X* is divided by the total number of reference samples of category *X* (column total). The resulting percentage accuracy indicates the probability that a reference (ground) sample will be correctly classified. What is really being measured using this method are errors of omission. In other words, samples that have not been correctly classified as category *X* have been omitted from the correct category. This accuracy value may be referred to as the "producers accuracy," because the producer of the classified image/map is interested in how well a specific area on the Earth can be mapped.

However, an important, but often overlooked, point is that a misclassification error is not only an omission from the correct category but also a commission into another category. Unless the classified image/map is 100 percent correct, all samples that are classified as category *X* are not actually category *X*. When the number of correctly classified samples of category *X* are divided by the total number of samples that were classified in category *X* (row total), the resulting percentage accuracy is indicative of the probability that a sample from the classified image/map actually represents that category on the ground. What is really being measured in this case are errors of commission. In fact, a better name for this value may be "reliability" (Congalton and Rekas, 1985) or "user's accuracy" because a map user is interested in the reliability of the map, or how well the map represents what is really on the ground.

An example of what might happen if one does not understand the use of these accuracy calculations follows. Suppose that an area is composed of three land-use/land-cover categories: forest (F), water (W), and urban (U). A classified image/map is produced, sampling performed, and an error matrix (Table 2) generated to assess the accuracy of the product.

An examination of the error matrix in Table 2 shows that the overall map accuracy is 63 percent. The traditional producer's accuracy for the individual land-use/land-cover categories shows that the forest classification is 93 percent accurate. This high value could lead a resource manager to conclude that this classified image/map is sufficiently accurate for his needs. However, upon identification of specific forest sites on the classified image/map for use in the field, the forester will be disappointed to find that only 49 percent of the sites identified as forests on the classified image/map are actually forested. In other words, 93 percent of the forest has been correctly identified as such, but only 49 percent of those areas identified as forests are actually forests while 51 percent of those areas identified as forests are either water or urban. Another way to view this difference is to consider the image/map producer standing in a forested site in this hypothetical area. The probability that this forested site was identified on his image/map as a forested site is 93 percent. However, consider the view of the forester (i.e., the user) who has chosen a forested site on the image/map for possible timber sales. The probability that this site, which was identified on the image/map as a forest, actually is a forest is only 49 percent.

Although these measures of accuracy may seem very simple, it is critical that they both be considered when assessing the accuracy of a classified image/map. All too often, only one measure of accuracy is reported. As was demonstrated in the example above, using only a single value can be extremely misleading. Given the optimal situation, error matrices should appear in the literature whenever accuracy is assessed so that the users can compute and interpret these values for themselves.

REFERENCES

Aronoff, S., 1982. Classification accuracy: A user approach. *Photogrammetric Engineering and Remote Sensing*. Vol. 48, No. 8, pp. 1299–1307.

Congalton, R., R. Oderwald, and R. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing*. Vol. 49, No. 12, pp. 1671–1678.

Congalton, R., and R. Mead, 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*. Vol. 49, No. 1, pp. 69–74.

Congalton R., and A. Rekas, 1985. COMPAR: A computerized technique for the in-depth comparison of remotely sensed data. *Proceedings of the 51st Annual Meeting of the American Society of Photogrammetry*, Washington, D.C., pp. 98–106.

Ginevan, M., 1979. Testing land use map accuracy: another look. *Photogrammetric Engineering and Remote Sensing*. Vol. 45, No. 10, pp. 1371–1377.

Hay, A., 1979. Sampling designs to test land use map accuracy. *Photogrammetric Engineering and Remote Sensing*. Vol. 45, No. 4, pp. 529–533.

Hord, R., and W. Brooner. 1976. Land use map accuracy criteria. *Photogrammetric Engineering and Remote Sensing*. Vol. 42, No. 5, pp. 671–677.

Smith, J., and W. Kovalick, 1985. A comparison of the effects of resampling before and after classification on the accuracy of a Landsat derived cover type map. *Proceedings of the International Conference on Advanced Technology for Monitoring and Processing Global Environmental Information*. University of London, London, England. 6 p.

Story, M., J. Campbell, and G. Best, 1984. An evaluation of the accuracies of five algorithms for machine processing of remotely sensed data. *Proceedings of the Ninth Pecora Remote Sensing Symposium*. Sioux Falls, S.D. pp. 399–405.

van Genderen, J., and B. Lock, 1977. Testing land use map accuracy. *Photogrammetric Engineering and Remote Sensing*. Vol. 43, No. 9, pp. 1135–1137.

Williams, D., J. Irons, R. Latty, B. Markham, R. Nelson, M. Stauffer, and D. Toll, 1983. Impact of TM sensor characteristics on classification accuracy. *Proceedings of the International Geoscience and Remote Sensing Symposium*. IEEE, New York, New York. Vol. I, Sec. PS–1. pp. 5.1–5.9.

(Received 6 August 1985; accepted 22 August 1985; revised 24 September 1985)

Forthcoming Articles

Walter H. Carnahan and *Guoping Zhou*, Fourier Transform Techniques for the Evaluation of the Thematic Mapper Line Spread Function.

J. D. Curlis, V. S. Frost, and *L. F. Dellwig*, Geological Mapping Potential of Computer-Enhanced Images from the Shuttle Imaging Radar: Lisbon Valley Anticline, Utah.

Ralph O. Dubayah and *Jeff Dozier*, Orthographic Terrain Views Using Data Derived from Digital Elevation Models.

Alan F. Gregory and *Harold D. Moore*, Economical Maintenance of a National Topographic Data Base Using Landsat Images.

John A. Harrington, Jr., Kevin F. Cartin, and *Ray Lougeay*, The Digital Image Analysis System (DIAS): Microcomputer Software for Remote Sensing Education.

Richard G. Lathrop, Jr., and *Thomas M. Lillesand*, Use of Thematic Mapper Data to Assess Water Quality in Green Bay and Central Lake Michigan.

Donald L. Light, Planning for Optical Disk Technology with Digital Cartography.

Thomas H. C. Lo, Frank L. Scarpace, and *Thomas M. Lillesand*, Use of Multitemporal Spectral Profiles in Agricultural Land-Cover Classification.

Uğuz Müftüoğlu, Parallelism of the Stereometric Camera Base to the Datum Plane in Close-Range Photogrammetry.

K. J. Ranson, C. S. T. Daughtry, and *L. L. Biehl*, Sun Angle, View Angle, and Background Effects on Spectral Response of Simulated Balsam Fir Canopies.

Arthur Roberts and *Lori Griswold*, Practical Photogrammetry from 35-mm Aerial Photography.

Paul H. Salamonowicz, Satellite Orientation and Position for Geometric Correction of Scanner Imagery.