B. K. Quirk
F. L. Scarpace
Environmental Remote Sensing Center
University of Wisconsin-Madison
Madison, WI 53706

A Method of Assessing Accuracy of a Digital Classification

The procedure consists of a selection of pixels, marking those pixels on the original digitized photograph, making a film product, and manually photo interpreting the marked pixels in a color additive viewer.

Abstract: *A procedure was developed for verifying the accuracy of land-cover classifications produced by computer. The procedure involved marking pixels in a computer file, creating a film product which could be placed in a color additive viewer, and performing a manual photo interpretation of the pixels. A program was then used to compare the results of the manual photo interpretation of the pixels with their computer classified counterparts, which resulted in a confusion matrix.*

Introduction

Until recently, one of the overlooked areas in the field of remote sensing has been determination of the accuracy of land-cover maps produced by either manual or digital interpretation of remote sensing products (Benson *et al.*, 1971; Hord and Brooner, 1976).

Essentially, the question is: "How accurately does the map produced by analysis of a remote sensing product represent the actual ground phenomena?" This is a very perplexing question because there is no easy method of determining the accuracy. The question becomes increasingly important in computer analysis of Landsat scenes or digitized aerial photography where the computed areas of the various land-cover categories are needed for planning or management decisions.

The novelty of remote sensing has left researchers to improvise their own techniques of accuracy verification. Many researchers report accuracies of 85 to 95 percent for their land-cover maps but fail to mention how they arrive at this figure. Other researchers simply make visual comparisons between the original and computer generated maps to verify the accuracy. Still others have actually tried to locate a sample of cells or pixels on the Earth's surface. This might be practical for a small number of Landsat pixels, but becomes prohibitively expensive and time consuming when attempting to locate a statistically significant number of sample pixels. Ground verification becomes even more difficult when the pixel size ranges from 5 to 10 metres as it frequently does with digitized aerial photography. Consequently, this aspect of remote sensing commonly requiring "ground truth" has always been one of the most difficult to quantify.

The problem of accuracy determination can be divided into two areas, one theoretical and the other practical. The theoretical component is based on sound statistical sampling procedure. The first step in an accuracy assessment is deciding on a sampling strategy. Typical questions to be asked at this stage are: What type of sampling technique will produce a 95 percent confidence interval or what are the probabilities of a Type I or Type II error? Beckett (1974) discusses sampling theory and provides an excellant background discussion on the subject.

The practical component of accuracy sampling in remote sensing can be divided into four sections:

Photogrammetric Engineering and Remote Sensing,
Vol. 46, No. 11, November 1980, pp. 1427-1431.

- selection of a sampling method,
- determination of the number of points or elements in a sample and point shape and size,
- the actual conducting of the sample(s),
- the statistical analysis of the sample(s)

These four sections are very interdependent, e.g., the statistical analysis of the sample(s) is heavily dependent on the assumptions of the selected sampling technique.

The sampling methods most often cited are the stratified random or stratified systematic unaligned sample procedures. Berry and Baker (1968) conclude that the stratified systematic unaligned sample procedure is ideal in verifying land-use maps produced by means of remote sensing. This method provides the user with a good sampling procedure when little is known about the shape of the autocorrelation function or if linear trends or periodicities occur. Most sampling procedures also depend on the "randomness" which may or may not occur in land-cover maps.

The selection of the number of points or elements in each category of the sample(s) has recently been discussed. Van Genderen *et al.* (1978) determined that 30 to 50 elements/category would produce a statistically valid sample. Hay (1979) and Ginevan (1979) have indicated the specific number of points/categories necessary to achieve a desired confidence level of the classification. The size and shape of the points or elements depends primarily on the type of sensor and purpose of the study. Additionally, Simonett and Coiner (1971) studied the effect of size versus information interpretability and concluded that the information content of photo interpretated Landsat images was low. We feel that, in any accuracy assessment attempt, the point or element size and shape should be compatible to the cell size used in the interpretation, whether it is manual or computer.

Once the sample has been gathered, the task of determining which statistics to analyze the sample(s) becomes important. Most sample(s) generally are compiled into a confusion or correlation matrix: a two-dimensional matrix where each axis pertains to an interpretation and the elements of the matrix indicate the amount of agreement/disagreement between the interpretation methods (Turk 1979). Marshall *et al.* (1969) based his analysis of the sample(s) on computing both the fractional correct and wrong for each category. In order to do this, however, some *a priori* knowledge must be assumed. Rudd (1971) used a chi-squared test for his analysis. More recently, analysis of variance of the resulting confusion matrix has been presented (Mead *et al.*, 1980; Rosenfield, 1980). Consequently, there is some disagreement as to the optimum method or procedure for analyzing the sample(s).

This leads into the practical problem of how to locate and interpret these sample(s), a problem which has been glossed over by researchers in their procedures section. As stated earlier, researchers have used a stratified random or stratified systematic unaligned sample based on dividing the area into a grid of equal area and taking a set of random samples within each grid. Fitzpatrick (1976) typifies these researchers; she chose cells or samples by randomly selecting the corner coordinates of the grid and performing photo interpretations of these sites.

Our research has been directed at developing a method for verifying the accuracy of land-cover maps with small pixel sizes (5 to 10 m) at a *minimum* cost, and thus avoiding the trap of spending more time and money on the accuracy assessment than on the actual classification. The Environmental Remote Sensing Center (ERSC) at the University of Wisconsin-Madison has been investigating this problem for several years with mixed results. Recently ERSC has developed a method which allows a researcher to verify more rigorously the accuracy of the land-cover maps produced from the digital analysis of aerial photography.

Procedures

The accuracy assessment methodology is a small part of the on going remote sensing program at ERSC. The techniques developed have thus far been used in conjunction with computer analysis of digitized aerial photography. The procedures and theories behind conversion of aerial photography into a digital format have been discussed in depth elsewhere (Scarpace, 1978; Scarpace and Quirk, 1980) and will not be discussed to any great detail in this paper.

In the digital analysis of aerial photography the first step is to convert the photograph to a digital format. This is done at ERSC with an Optronics

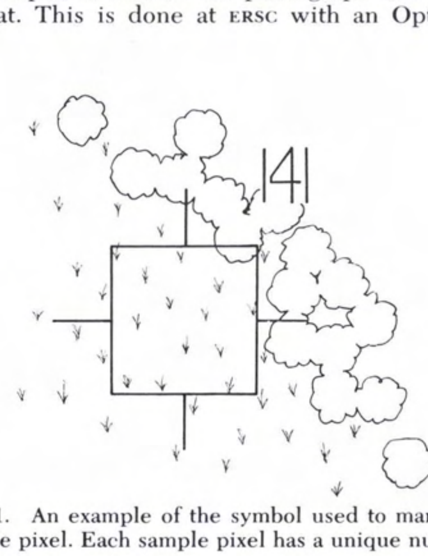

Fig. 1. An example of the symbol used to mark each sample pixel. Each sample pixel has a unique number.

P-1700 scanning microdensitometer. The aerial photograph is placed on the Optronics drum and scanned three times (once for each emulsion layer) through narrow band filters centered at 450, 550, and 650 nm. This digital file is placed on a computer system where a computer software package is used to transform the film densities into log exposures (Scarpace, 1978) and then to locate training sets from density slices of the file. The training set statistics (mean vector, etc.) are then used in classifying the digitized imagery into a land-cover map. Usually the preliminary supervised computer classifications are visually checked against the original aerial photograph by viewing film images of the classification in a color additive viewer. The computer classifications are made into three color separations, photowritten on the Optronics system, and overlayed in the color additive viewer. After several iterations any unclassified or misclassified areas have been identified and the computer generated land-cover map visually appears to represent the land cover imaged on the original aerial photograph.

In order to evaluate this computer generated land-cover map, a procedure has been developed that marks, with a numbered square and tick marks (Figure 1), specific pixels in each scanned emulsion layer of the original scanned photograph. The researcher first determines how many pixels will be marked within the study area. The marked pixels are either randomly distributed over the entire digitized aerial photograph or stratified based on the number of pixels per land-cover category of the computer classification. Once the pixels are marked in the computer file, each color separation is photowritten onto film with the Optronics system. These three images are then viewed in a color additive viewer and a manual photo interpretation is done of the marked pixels. In all cases the manual photo interpretation is based upon field work done in the study area. Figures 2 and 3 are examples of marked images. The

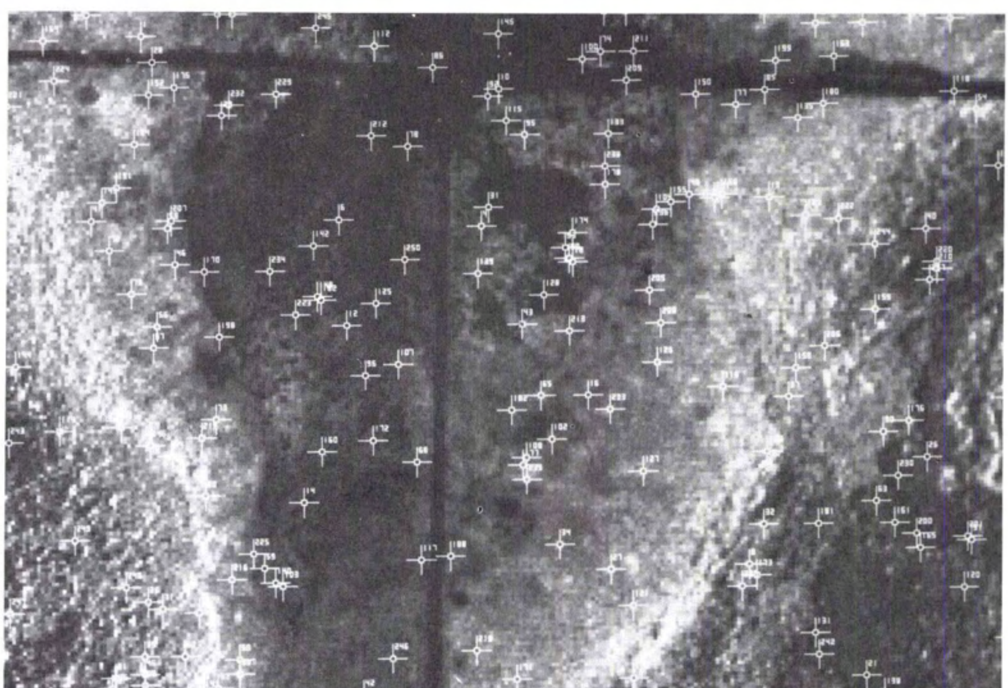

FIG. 2. A marked image of Sheboygan Marsh located in north central Wisconsin. Original image was a color infrared image taken on 31 July 1974 by a NASA RB-57 aircraft at a scale of 1:120,000. A area of 253 hectares (625 acres) is represented on this image.

FIG. 3. A marked image of a suburban development located north of Milwaukee, Wisconsin. The original image was a color infrared image taken on 31 July 1974 at a scale of 1:120,000. An area of approximately 138 hectares (342 acres) is represented on this image.

results of the manual photo interpretation of the marked pixels are then compared to their computer classified counterparts with a program which results in a confusion matrix (Table 1). This method allows a comparison to be made between a manual photo interpretation and a computer generated land-cover map of the same area and, therefore, supplies a more direct measurement of

TABLE 1. COMPARISON OF MANUAL VS. COMPUTER CLASSIFICATION FOR THE AREA WITHIN SCHOONMAKER CREEK. THE NUMBERS OF PIXELS AND THE PERCENTAGE CLASSIFIED CORRECTLY OUT OF 151. THE LEFT AXIS REPRESENTS THE LAND-COVER CATEGORIES CLASSIFIED BY THE COMPUTER FOR THE SAMPLE POINTS IN THE SCHOONMAKER CREEK SITE. THE TOP AXIS REPRESENTS THE MANUAL PHOTO INTERPRETATION USING THE SAME LAND-COVER CATEGORIES OF THE IDENTICAL POINTS OR PIXELS.

COMPUTER CLASSIFICATION	MANUAL PHOTO INTERPRETATION				
	Unclassified	Impervious	Vegetation	Transition	% Accurate/Row
Unclassified	0	1			
Impervious		81	4		95.3
Vegetation		6	49		89.1
Transition		3	2	5	50.0
% Accurate/Column		89.0	89.1	100.0	89.4

the accuracy of the computer generated land-cover classification. We are presently studing procedures for analyzing the resulting confusion matrix.

Conclusions

The methodology described provides an inexpensive, straightforward method of verifying the accuracy of land-cover maps produced by digital analysis of either Landsat or digitized aerial photography. The procedure consists of random or stratified selection of pixels throughout the entire study area, marking the original digitized photograph, making a film product, and manually photo interpreting the marked pixels in a color additive viewer. The subsequent results can be compared with the original computer generated land-cover classification of the area or any smoothed or generalized computer classification. The result of this comparison is a confusion matrix. The cost to mark, photowrite, do the manual photo interpretation, and run the computer comparison is approximately $75.00 per image and requires 9 man hours. The computer cost is based on research rates which are half the commercial rates. In addition to verifying the accuracy of computer generated land-cover maps, this method has been used to determine the accuracy of maps produced by manual photo interpretations of aerial photography. This method provides the user with an inexpensive method for obtaining an estimate of the accuracy of resulting land-cover maps. It also would be ideal for determining the preliminary accuracy of maps produced by means of remote sensing. Once the interpretaters are satisfied with the resulting land-cover maps, a more detailed ground study could be carried out, time and money permitting.

References

Beckett, P. H., 1974. The Statistical Assessment of Resource Surveys by Remote Sensors, *Environmental Remote Sensing*, Edward Arnold, London, England, pp. 9–27.

Benson, A. S., *et al.*, 1971. Ground Data Collection and Use, *Photogrammetric Engineering*, Vol. 37, No. 11, pp. 1159–1166.

Berry, B. J., and A. M. Baker, 1968. Geographic Sampling, in *Spatial analysis: a reader in statistical geography*, Prentice Hall, pp. 91–100.

Fitzpatrick, K. A., 1976. The Strategy and Methods For Determining Accuracy of Small and Intermediate Scale Land Use and Land Cover Maps, *Proceedings of the 2nd Annual Pecora Symposium*, pp. 339–361.

Ginevan, M. E., 1979. Testing Land-Use Map Accuracy: Another Look, *Photogrammetric Engineering and Remote Sensing*, Vol. 45, No. 10, pp. 1371–1377.

Hay, A. M., 1979. Sampling Designs to Test Land-Use Map Accuracy, *Photogrammetric Engineering and Remote Sensing*, Vol. 45, No. 4, pp. 529–533.

Hord, R. M., and W. Brooner, 1976. Land-Use Map Accuracy Criteria, *Photogrammetric Engineering and Remote Sensing*, Vol. 42, No. 5, pp 671–677.

Marshall, R. E., *et al.*, 1969. Use of Multispectral Recognition Techniques For Conducting Rapid, Wide-Area Wheat Surveys, *Proceedings of the 6th International Symposium on Remote Sensing of Environment*, pp. 3–20.

Mead, R. A., *et al.*, 1980. Statistical Techniques For Analysis Of Landsat Classification Accuracy Data, 46th Annual Meeting of the American Society of Photogrammetry, St. Louis, Mo.

Rosenfield, G., 1980. Analysis of Variance for Classification Accuracy Assessment, 46th Annual Meeting of the American Society of Photogrammetry, St. Louis, Mo.

Rudd, R. D., 1971. Macro Land-Use Mapping With Simulated Space Photos, *Photogrammetric Engineering*, Vol. 37, No. 4, pp. 365–372.

Scarpace, F. L., 1978. Densitometry on Multi-Emulsion Imagery, *Photogrammetric Engineering and Remote Sensing*, Vol. 44, No. 10, pp. 1279–1292.

Scarpace, F. L., and B. K. Quirk, 1980. Land-Cover Classification Using Digital Processing of Aerial Imagery, *Photogrammetric Engineering and Remote Sensing*, Vol. 46, No. 8. pp. 1059–1065.

Simonett, D. S., and J. C. Coiner, 1971. Susceptibility of Environments to Low Resolution Imaging for Land-Use Mapping, *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, pp. 373–393.

Turk, G., 1979. GT Index: A Measure of the Success of Prediction, *Remote Sensing of Environment*, 8: pp. 65–75.

Van Genderen, J. L., and B. F. Lock, 1976. *A Methodology For Producing Small Scale Land Use Maps in Semi-arid Developing Counties Using Orbital Imagery*, NASA Final Report, 313 pgs.

(Received 9 November 1979; revised and accepted 27 May 1980)

COVER PHOTOS NEEDED