Anais XI SBSR, Belo Horizonte, Brasil, 05 - 10 abril 2003, INPE, p. 2877 - 2879.

# SOURCES OF ERROR IN A VIDEO-DERIVED REFERENCE DATA SET FOR ACCURACY ASSESSMENT OF LAND-COVER MAPS IN THE AMAZON

REBECCA POWELL[1], NICHOLAS MATZKE[1], CARLOS DE SOUZA, JR.[1,2]
MATTHEW CLARK[1], IZAYA NUMATA[1], LAURA HESS[3], DAR ROBERTS[1]

[1]Department of Geography—University of California, Santa Barbara
Santa Barbara, California 93106, USA
[becky,matzke,carlos,mclark,numata,dar]@geog.ucsb.edu

[2]Instituto do Homem e Meio Ambiente da Amazônia—Imazon
Caixa Postal 5101, Belém, PA, Brasil • CEP 66613-397
souzajr@imazon.org.br

[3]Institute for Computational Earth System Science—ICESS
University of California, Santa Barbara, CA 92106, USA
lola@icess.ucsb.edu

**Abstract**: Using a reference data set of high-resolution geocoded images generated from mosaicked digital videography, we have examined the sources of errors introduced during the accuracy assessment of a regional land-cover map classified from Landsat TM imagery of central Rondônia. We compared the reference data produced by five trained interpreters to quantify the degree of subjectivity in labeling of reference data and to identify which classes are most likely to be confused by interpreters. Errors were attributed to sources including geolocation difference between reference data and classified map, change between dates, mixed pixels, and disagreement between interpreters. Interpreters disagreed on the classification of almost 30% of the total samples; mixed pixels accounted for a large proportion of the disagreement. Agreement between any two interpreters in labeling of 30x30 m reference samples averaged less than 50% for the second growth forest class. Approximately 10% of the error in agreement between map and reference data could be attributed to geolocation error. These findings suggest that 1) definitions of land-cover types such as "second growth" are more subjective and variable than is commonly assumed; 2) validation data sets that include only non-mixed, non-edge samples are likely to result in overly optimistic accuracy estimates.

**Keywords**: regional land-cover mapping, accuracy assessment, error analysis, aerial videography.

## 1. Introduction

Thematic maps derived from remotely sensed data are used in many applications, including as input to scientific models, as the source of regionally extensive environmental data, or as the basis of policy analysis. The results of such applications may significantly depend on the accuracy of the thematic map, yet documenting map accuracy is not a straight forward task. Though measures of map accuracy are well established in the literature (e.g. Congalton and Mead,1983; Stehman, 1997), there remain "many uncertainties with the meaning and interpretation of map quality that make it a difficult variable to consider objectively" (Foody, 2002). Estimation of the thematic accuracy of land-cover maps by means of a set of labeled reference samples assumes that the reference data set is a statistically valid sample of the mapped area, that the reference samples are accurately co-registered with the map, that the samples can be consistently and unambiguously labeled as one of the map classes, and, if time has elapsed between acquisition of the map and reference datasets, that land cover has not changed in the interim. This study examines the validity of these assumptions for the accuracy assesssment of a regional land-cover map derived from Landsat TM imagery, using digital aerial videography as the source of reference data.

## 2. Study area and land-cover map

The study area is covered by three Landsat TM scenes over central Rondônia: Ariquemes (P232, R67), Ji-Paraná (P231, R67), Luiza (P231, R68). The region includes a range of pasture and second-growth ages, as well as a range in degree of pasture maintenance. Elevation varies between 100 and 1000 m, including some areas of steep terrain. Two of the three Landsat scenes (Ji-Paraná and Luiza) were collected in August 1999; for the third scene (Ariquemes), scenes from July 1998 and October 1999 were combined due to cloud cover.

Land cover was mapped using a multi-stage process based on spectral mixture analysis (Roberts et al. 1998; Roberts et al. 2002). Endmember fractions and root-mean square error values were used to train a binary decision tree classifier. The rules generated by the decision tree were used to classify the entire image into seven classes: rock/savanna, primary forest, pasture, green pasture, secondary forest, water, and urban/soil. For the purpose of accuracy assessment, we combined the pasture and green pasture classes, and did not include the rock/savannah class because of its scarcity on the landscape.

## 3. Accuracy Assessment

Reference data were derived from digital aerial videography collected over Rondônia in June 1999, as part of the Validation Overflights for Amazon Mosaics (Hess et al., 2002). Wide-angle and zoom data were collected simultaneously, with average swath widths of 550m and 55m, and average pixel dimensions of 0.75m and 7.5 cm, respectively. GPS location and time code, aircraft attitude, and aircraft height data were encoded on each video frame, and used to automatically generate geocoded video mosaics with an estimated absolute geolocation error of 5-10m along the center third of the video swath.

We randomly selected image frames based on the recorded time code along several flight lines that crossed the Ariquemes, Ji-Paraná, and Luiza scenes. Georectified mosaics were constructed from wide-angle video frames covering 1 second (30 frames) of aircraft flying time. A 5x5 grid of 30m pixels was superimposed and centered on each mosaic. The center and four corner pixels were collected as reference samples, for a total of five samples per mosaic.

Five graduate students were trained in photointerpretation of this region. Three of the students had prior experience working in the Amazon region of Brazil; the other two had extensive experience working in tropical forests in other regions of the world. An effort was made to increase consistency between interpreters. First, all interpreters agreed on the criteria for each class based on simple bio-physical descriptions; second, the interpreters practiced classifying the videography as a group on mosaics not included in the accuracy assessment. After the training period, each interpreter independently analyzed the samples on all mosaics included in the analysis. A total of 158 mosaics, generating 790 sample points, were included.

Each set of reference data (one per interpreter) was compared to the classified map, as well as to each other. Error matrices, user's and producer's accuracies, and the kappa coefficient of agreement were generated for each combination based on Congalton and Mead (1983) and Hudson and Ramm (1987).

## 4. Analysis of error sources

Before an accurate report of the classified map error can be made, a realistic assessment of error present in the reference data must be conducted. In most accuracy assessment reports, "ground reference data are assumed to be an accurate representation of reality" (Foody, 2002). We question this assumption and investigate the following sources of error in

the reference data (or between the reference and map data): georectification, change between collection of videography and collection of the TM image, mixed pixels, human interpretation, and difficulty of discriminating between classes. Once these categories of error are accounted for, we assume the remaining disagreement between the reference data and the map is due to misclassification error.

As a final step, the interpreters will reassemble and as a group produce the reference data set used for the final accuracy assessment analysis of the classified map. Reasons for disagreement in the independent analysis are being recorded and analyzed. In cases where the group continues to disagree or remained uncertain, zoom video and laser altimetry data will be accessed to clarify the vegetation structure on the ground.

Among the significant findings are as follows: Agreement between individual interpreters for the second growth class ranged between 32% and 68%, with an average agreement of 49%. Interpreters disagreed on the classification of almost 30% of the total samples; over 50% of all interpreter disagreement was between pasture and second growth classes, and almost 30% of disagreement was between second growth and primary forest classes. Almost half of the disagreement between interpreters occurred when the sample was a mixed pixel, and disagreement was most common when 50% or less of the pixel was covered by a single land-cover class. Finally, approximately 10% of the original error between the reference data and the classified map was due to georectification error.

These findings suggest at least two significant points: 1) definitions of some land-cover types such as "pasture" and "second growth forest" are more subjective and variable than is commonly assumed, and 2) validation data sets that include only non-mixed, non-edge samples are likely to result in overly optimistic accuracy estimates.

## References

Congalton, R. G.; Mead, R. A. A Quantitative Method to Test for Consistency and Correctness in Photointerpretation. *Photogrammetric Engineering and Remote Sensing*, v. 49, n. 1, p. 69-74, 1983.

Foody, G. M. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, v. 80, p. 185-201, 2002.

Hess, L. L.; Novo, E. M. L. M.; et al. Geocoded digital videography for validation of land cover mapping in the Amazon basin. *International Journal of Remote Sensing*, v. 23, n. 7, p. 1527-1556, 2002.

Hudson, W. D.; Ramm, C. W. Correct Formulation of the Kappa Coefficient of Agreement. *Photogrammetric Engineering and Remote Sensing*, v. 53, n. 4, p. 421-422, 1987.

Roberts, D. A.; Batista, G. T.; et al. Change identification using multitemporal spectral mixture analysis: applications in Eastern Amazonia. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. R. S. Lunetta; C. D. Elvidge, eds. Chelsea, Michigan, Ann Arbor Press, p. 137-161, 1998.

Roberts, D. A.; Numata, I.; et al. Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers. *Journal of Geophysical Research*, in press, 2002.

Stehman, S. V. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, v. 62, p. 77-89, 1997.