C VERIFICATION OF LAND COVER MAPS FROM LANDSAT DATA

David S. Linden & John Szajgin (Technical Director - Technicolor Graphics Services - BLM Operations - Denver, CO)

Introduction

The US Geological Survey's EROS Data Center has pioneered the application of cluster sampling to verifying the accuracy of land cover maps derived from digital Landsat data. This approach was applied by Rohde (Rohde 1976), as part of a pilot project in a 65,000 hectare area in the Denali region of Alaska. Analysis of the pilot study data indicated that cluster sampling was an efficient technique for accuracy assessment. Based on this result, variants of the cluster sampling technique have been used in large scale accuracy assessments for areas in excess of one million hectares in Alaska, Oklahoma and Arizona.

Cluster sampling can be an efficient means of sampling in wildland environments. The largest cost incurred in the field effort is traveling to, and locating, the sample pixels. Data collection procedures on the pixel represent a small proportion of the total cost. For this reason, once a specific pixel is found, it is more efficient to collect data from a number of pixels in close proximity, than to travel to, and locate, widely scattered individual pixels. In this way, more pixels are visited with a corresponding decrease in the sampling cost per pixel.

There is, however, a point of diminishing returns. Sampling adjacent pixels yields less information about the overall population than does sampling the same number of spatially separated pixels. This is because adjacent pixels tend to be similar to each other, and redundant information may exist within a sample cluster. The amount of redundant information is related to the statistical parameter rho known as the intracluster correlation coefficient (Sukhatme et al., 1970). Rho is a measure of the homogeniety of the population. Values of rho close to 1.0 indicate very small clusters should be used. Rho has averaged about .3 in the accuracy assessments discussed herein.

There are three types of classification errors which may be of interest, commission, omission and overall error classification. Commission errors for a particular cover type occur when pixels are classified as that cover type but are found to be some other cover type when field checked. Omission errors for a particular cover type occur when pixels, field visited and known to be that cover type, are classified as some other cover type. Overall error is the proportion of pixels incorrectly classified, without regard to omission or commission.

Since the classified image represents the sampling frame, sampling for accuracy assessment is designed to estimate commission error. However, the sample can also provide useful estimates of omission and overall error. The appropriate use of a particular paired observation, Landsat classification and corresponding ground classification, enables one to utilize that observation for each of these estimates.

Alaska

The Alaskan accuracy assessment was conducted during the summer of 1979 as part of a cooperative project between EROS and the US Department of Interior's Bureau of Land Management (BLM). The Landsat classification verified, was a Level IV classification of a one million hectare site in the Denali region. The objective of the accuracy assessment was to estimate the commission error at Level IV for each of six resource class strata with a precision of plus or minus 10% at the 90% confidence level (10/90). Overall error as well as individual stratum omission errors were also to be estimated. However, no precision levels were specified for these estimates.

The original sample was designed as a stratified two phase cluster sample. The first phase consisted of the photointerpretation of all allocated clusters using 9" x 9" true color stereo triplets at a scale of 1:3,000. The second phase was on the ground visitation of a subsample of the clusters where a classification was made for each pixel in the cluster. The clusters were to consist of 25 pixels laid out in 5 x 5 square grid. The cluster size of 5 x 5 was selected as being the largest cluster that a field crew could locate and sample within half a day. In order to obtain the required 10/90 precision for each stratum, an independent allocation was made for each. The entire digital image was first gridded into 5 x 5 clusters. For each stratum, clusters were selected with probability proportional to the number of pixels in the cluster classified as the resource class contained in the stratum under allocation. This sample allocation is called probability proportional to cluster size (referred to as PPCS sampling).

PPCS allocation was used to insure that the clusters selected for a given strata would contain as many pixels as possible from that strata while still being statistically sound. The effectiveness of the PPCS allocation is shown in the table below where the number of pixels expected in a randomly selected cluster as estimated by area proportion is compared to the average number actually obtained in the PPCS sample.

Strata	Expected Pixels In Randomly Selected Cluster	Average Pixels In Actual PPCS Sample Clusters
1	0.5	10.3
2	1.6	13.0
3	2.0	9.0
4	7.4	19.0
5	9.3	17.8
6	4.0	16.8

In determining the sample size required to obtain the required precision, the following assumptions were made —

1 The classification accuracy was similar to that of the classification developed by Rohde in 1976 (pers. comm.). Therefore, the sample variation found in the accuracy assessment of the 1976 product could be used to determine the sample size required for the product presently under evaluation.

2 Personnel could perform the required photointerpretation with a photo-ground correlation of at least .85.

3 Lack of suitable weather conditions during the preferred part of the growing season and monetary constraints would only allow aerial photography to be obtained for approximately 150 clusters.

4 The availability of personnel would only allow 72 clusters to be sampled in the field.

Based on these assumptions, a precision level of 10/90 could be achieved for each stratum if 24 clusters were photointerpreted, 12 of which were ground visited. This required a total of 144 photo clusters and 72 ground clusters. To allow for inaccessible clusters and other contingencies, 168 prospective clusters were actually allocated and plotted on 1:63,360 scale quadrangle sheets. Black/white stereo triplets at a scale of 1:6,000 were acquired over all the ground clusters to be sampled. The cluster outlines were then plotted on the photos. These photos were used by the field crews to locate the clusters on the ground. Seventy clusters were actually visited.

All 168 clusters were photographed at 1:3,000 scale using true color film. There were 154 acceptable clusters which were interpreted. Preliminary data analysis indicated that the ground and photointerpretations for individual pixels agreed for only 47% of the pixels. Further investigation indicated that the ground data adquately indicated vegetation association, but was inadequate in determining percent cover. The photo data adequately indicated percent cover while inadequately determining vegetation association. Based on these findings, it was decided that only those clusters with both photo and ground data could be used in the accuracy assessment. A single classification was made for each pixel in the 70 clusters based on both the photo and ground data. The remaining 98 photo clusters were discarded.

It was decided that Level IV was too detailed a classification to be workable. Ground crews had great difficulty in reaching agreement on Level IV calls. Therefore, it was decided that the accuracy of the individual strata as well as overall accuracy would be evaluated at Level III.

The results of the modified design are presented below. The interested reader may refer to Appendix A for a detailed treatment of the statistical formulae used in the analysis.

Level III

STRATA	COMMISSION		OMISSION	
	Percent Correct	Confidence Interval (90%)	Percent Correct	Confidence Interval (90%)
Tall Shrub	15	± 12	22	± 19
Low Shrub	70	± 7	57	± 3
Woody Tundra	26	± 9	57	± 5
All Other	33	± 12	11	± 7

Although the first phase of the original design had to be dropped, the effort still provided useful estimates of classification errors. The objective of estimating commission errors plus/minus 10% at the 90% confidence level was nearly met even though half of the planned sample plots could not be used. The stratified PPCS sampling was proved to be an effective means of controlling the sample allocation. However, the statistical formulae and resulting analysis are quite complicated as a glimpse at Appendix A will show. The authors recommend that PPCS sampling only be used when a statistician is available for all phases of the assessment.

Oklahoma

The Oklahoma accuracy assessment was conducted during the early spring of 1980 as part of a cooperative project between EROS and the US Fish & Wildlife Service. The objective of the project was to identify potential prairie chicken habitat based on cover type information developed from digital Landsat data. The area mapped consisted of 7 subscenes of approximately 8,300 hectares each. Two representative subscenes were chosen to be verified. The objective of the accuracy assessment was to estimate the overall accuracy of the combined subscenes plus/minus 10% at the 90% confidence level. No estimates of individual class commission or omission errors were required.

The subscenes were relatively small and irregular in shape. If clusters were chosen randomly, there was a high probability that selected sample clusters would cross over the subscene boundaries into areas of image fill and therefore, contain no classification data. PPCS sampling was used to minimize the chance of sampling boundary areas while maintaining the unbiasedness of the estimators. Based on the Alaska experience, the desired precision level, and the available resources, a sample size of 30 - 4 pixel x 4 pixel clusters was used.

The entire project area in Oklahoma was readily accessible by automobile. There was an extensive network of roads and fences throughout. The clusters were plotted on 7.5 minute 1:24,000 scale topographic maps. The field crews were able to locate the clusters on the ground by using these maps along with a staff compass and a tape measure. It was not necessary to acquire any aerial photography of the area.

The overall accuracy of the classification was determined to be 86% plus or minus 4.4% at the 90% confidence level. The appropriate statistical forumulae can be found in Appendix B. The specified precision level was easily attained.

The PPCS estimators for overall error are unbiased and relatively simple when compared to the PPCS estimators for individual class commission errors which are slightly biased and very complex. The authors do recommend PPCS sampling for estimating overall error.

Arizona

The Arizona accuracy assessment was conducted during the summer of 1980 as part of a cooperative project between EROS and the BLM (Rohde/Miller 1980). The area classified is comprised of 8 Level II cover types. The objective of the accuracy assessment was to evaluate the commission error of each of the 8 cover types within plus/minus 10% at the 90% confidence level.

The sample design was a stratified two-phase cluster sample with equal probabilities of selection within strata. The strata corresponded to the eight Level II resource cover types represented in the classification. The digital image was gridded into mutually exclusive sample clusters. The cluster size used was 5 pixels by 3 pixels. To control the precision of the estimate for each cover type, and to assure frequent occurrence of the cover type of interest within the corresponding stratum, the image was stratified based upon class plurality within the sample clusters. This will be referred to as stratified plurality sampling (SPS). This established the 8 strata and defined the sampling frame. The effectiveness of the SPS allocation is shown in the table below, where the number of pixels of a given class expected in a randomly selected cluster is compared to the average number obtained in the stratified plurality sample.

Strata	Expected Pixels In Randomly Selected Cluster	Average Pixels In Actual SPS Clusters
1	0.1	8.9
2	0.1	11.8
3	2.6	12.4
4	0.1	7.6
5	2.5	14.0
6	9.2	13.3
7	0.1	9.9
8	0.5	9.3

The sample size was determined based on the desired level of precision and confidence, previous experience with the Alaska and Oklahoma accuracy assessments and available resources. Taking these factors into consideration, a sample size of 160 clusters was chosen. Twenty clusters were allocated to each strata. The sample design called for 20 in each strata to be photointerpreted and a subsample of 10 in each strata to be ground visited.

All clusters were plotted into 7.5 minute and 15 minute USGS topographic maps and orthophoto quads. These were used to plot flightlines for acquiring 1:3,000 scale black/white stereo triplets over all the clusters to be ground sampled. The cluster outlines were then plotted onto the photos. These photos were used by the field crews to locate the clusters on the ground.

Due to unusually poor weather and mechanical problems, completion of the photo acquisition over all sample clusters for the first phase of this design was considerably delayed. The photos were recently delivered but are not yet interpreted. A preliminary data analysis based on the ground data has been completed, using the statistical formulae described in Appendix C. Based on these preliminary results and an expected photo-ground correlation of .8, the specified precision of the estimates should be met. The SPS sampling appears to be as efficient as the PPCS sampling used in Alaska. The authors strongly recommend the SPS approach over the PPCS approach when estimating commission errors for individuals classes. The SPS estimates are unbiased and far less complex than the PPCS estimates. The SPS approach could be applied by anyone well versed in statistics but not necessarily a statistician.