A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data

Russell G. Congalton
Department of Forestry and Resource Management, University of California, Berkeley, CA 94720

ABSTRACT: Sample means and variances, obtained through computer simulation, were compared with the corresponding population means and variances for five sampling schemes typically used in assessing the accuracy of land-cover maps derived from remotely sensed data. The five sampling schemes were simple random sampling, stratified random sampling, cluster sampling, systematic sampling, and stratified systematic unaligned sampling. Three data sets of varying spatial complexity, including an agricultural area, a range area, and a forest area, were investigated. The patterns of error in each data set, as measured by spatial autocorrelation analysis, greatly influenced the appropriate sampling scheme to be used for assessing the map accuracy. The results indicate that simple random sampling always provides adequate estimates of the population parameters, provided the sample size is sufficient. For the less spatially complex agriculture and range areas, systematic sampling and stratified systematic unaligned sampling greatly overestimated the population parameters and, therefore, should be used only with extreme caution. Cluster sampling worked reasonably well. However, clusters should not be taken of size greater than 25 pixels and preferably 10 pixels.

INTRODUCTION

IN RECENT YEARS, accuracy assessment of remotely sensed data has received widespread support in the remote sensing community. The discrete multivariate analysis techniques introduced by Congalton *et al.* (1983) and endorsed by Rosenfield and Fitzpatrick-Lins (1986), Hudson and Ramm (1987), and others have become widely used tools for assessing the accuracy of maps derived from remotely sensed data. The importance of the error matrix is now well recognized both for the information it yields itself (Story and Congalton 1986) and as input into more advanced statistical techniques. The overriding assumption then in the entire accuracy assessment procedure is that the error matrix must be indicative or representative of the entire area mapped from the remotely sensed data. In other words, the proper sampling approach was used in generating the error matrix on which all future analysis will be based. If this assumption is a bad one, then all the results of the accuracy assessment are void. Therefore, it is the objective of this paper to investigate the sampling schemes used in generating the error matrices for accuracy assessment. This objective can only be met by first exploring the inherent nature of errors in remotely sensed data and then relating these patterns to the sampling approaches of interest.

LITERATURE REVIEW

Assessing the accuracy of land-cover maps generated from remotely sensed data is expensive in both time and money. Obviously, a total enumeration of the mapped areas for verification is impossible. Sampling, therefore, becomes the means by which the accuracy of the land-cover map can be derived. Using the wrong sampling approach can be costly and yield poor results. Therefore, it is imperative to understand the patterns of error in the map, the sampling scheme, and the their interaction in order to properly assess the accuracy.

Ginevan (1979) states three criteria for judging a sampling procedure:

- The procedure should have a low probability of accepting a map of low accuracy.
- The procedure should have a high probability of accepting a map of high accuracy.
- The procedure should require some minimum number of ground samples.

As in other disciplines, early work in sampling for accuracy assessment focused on sample size determination (the third criterion above) (e.g., Hord and Brooner, 1976; van Genderen *et al.*, 1978; Hay, 1979). The binomial distribution was recognized as the appropriate mathematical model to use for accuracy assessment. The probability of getting y misclassifications in a sample of n from a certain map is given by the binomial probability density function (pdf) as follows:

$$f(y) = \frac{n!}{(n-y)!y!} p^y \quad (1-p)^{n-y}$$

where n = sample size,

y = number of failures (i.e. misclassifications), and

p = map accuracy (i.e., proportion of incorrectly classified pixels).

Other researchers, notably Rosenfield and Melley (1980) and Rosenfield (1982), used the normal distribution as the large sample approximation to the binomial.

Research in sample size determination for accuracy assessment has been adequate enough to at least set some general guidelines. On the other hand, little research has been performed in the area of sampling schemes. As pointed out by Card (1980), "... there has been no comparative study of systematic versus simple random sampling for map accuracy estimation." In fact, due to the high cost of data collection and time involved, there have been no direct comparisons of any sampling schemes used in assessing the accuracy of maps derived from remotely sensed data.

Many times, sampling schemes are chosen by tradition or because of time or money constraints. Little consideration is given to statistical validity. Also, there seems to be little agreement about which sampling scheme should be used. Hord and

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 54, No. 5, May 1988, pp 593–600.

0099-1112/88/5405–593$02.25/0

Brooner (1976) used simple random sampling (SRS) with replacement while Zonneveld (1974) stated that simple random sampling is impractical because it is too costly in time and money. Also, SRS yields too many samples in large areas and not enough in the small and usually important areas. Instead, Zonneveld advocated the use of stratified random sampling in which the map stratified by land-cover types. Van Genderen *et al.* (1979) and Ginevan (1979) agreed with Zonneveld while Rudd (1971) would stratify by cover type but then place some minimum number of sample points in the smallest category to make sure all categories are adequately sampled. Berry and Baker (1968) would stratify geometrically and not by land-cover type. Todd *et al.* (1980) and Rhode (1978) cited economic considerations as justification for using cluster sampling while Fitzpatrick-Lins (1978a, 1978b) and Ling *et al.* (1979) favored systematic unaligned sampling.

METHODS

The first step in investigating sampling schemes used in accuracy assessment involves exploring the patterns of error in remotely sensed data. To this end, three data sets of varying spatial complexity were collected. A detailed description of these data sets and their analysis is given by Congalton (1984, 1988). A summary description is given here so that the reader can understand the analysis of the various sampling schemes.

DATA

The data sets used in this study consisted of a USGS 7½-minute quadrangle-sized area of a forested environment (most spatially complex), a rangeland environment, and an agricultural environment (least spatially complex). For each quad sized area, a Landsat MSS land-use/land-cover classification and an assumed correct reference classification were available. A difference image, so called because it shows the differences between the two classifications, was generated for each of the three data sets. Each image is a matrix of zeros and ones in which the ones indicate the errors (disagreement) and the zeros indicate the agreement (see Table 1). All further analysis was performed on the difference images. Each difference image could be visually displayed and the patterns of error observed. In addition, spatial autocorrelation analysis could be applied to each difference image to quantify the pattern of error (Congalton, 1988). Finally, each difference image could be viewed as a population and used in investigating different sampling schemes. Because the population is binomially distributed (i.e., correct or error), the parameters of interest are (1) the population size, N, (2) the proportion of incorrect (error) responses or mean, P, and (3) population variance. Once these population parameters were obtained, computer simulation of the various sampling schemes was performed in order to evaluate the effectiveness of each in assessing the accuracy of land-cover maps derived from remotely sensed data. In other words, the sample estimates for the mean and variance from each sampling scheme were compared to the population parameters to see which estimates best approximated the population.

TABLE 1. POPULATION PARAMETERS FOR THE THREE DIFFERENCE IMAGES.

	Difference Image		
	Agriculture	Range	Forest
Image shape (pixels)	245 × 176	240 × 160	241 × 144
Image size (pixels)	43,120	38,400	34,704
Pixel size (metres square)	63.6	50	60
Population mean (proportion incorrect)	0.2411	0.3243	0.2180
Population variance	0.1830	0.2191	0.1705

SAMPLING

Computer simulation was used to compare five sampling schemes on the three difference images of varying spatial complexity. These five sampling schemes were simple random sampling (SRS), cluster sampling (CS), stratified random sampling (STRAT), systematic sampling (SYSTEM), and stratified systematic unaligned sampling (SSUS). For each simulation the following information was recorded and used for comparison: the sample size (n), the mean or proportion incorrect (p), and the variance of the sample. All the sampling schemes except cluster sampling used a single pixel as the sampling unit. However, because the sample unit in cluster sampling is a cluster of pixels, the following additional information was recorded: cluster size, cluster shape, variance between clusters, variance within clusters, relative efficiency, and intracluster correlation.

Each of the five sampling schemes were simulated at various sample sizes and at 400 repetitions using a FORTRAN computer program written by the author. In other words, each simulation was repeated 400 times for each chosen sample size for each sampling scheme and the results were averaged together. Sample sizes were chosen, as in any sampling procedure, to obtain the most information with the least effort. In other words, small sample sizes were investigated to see if they would yield adequate results. It should also be emphasized that the sampling approaches chosen were the ones most commonly used by the remote sensing community. Therefore, this analysis should be of practical significance as well. Because the population parameters for each difference image were known, it was possible to compare the results of each sampling scheme simulation (sample mean and variance) with the actual known values (population mean and variance). This comparison allowed one to determine the best sampling scheme to use for each spatially different data set.

Simple Random Sampling Without Replacement. Simple random sampling is a method of selecting n sample units out of N units in the population, such that every one of the possible distinct samples has an equal chance of being selected (Cochran, 1977). Each sample unit is identified by an x,y coordinate pair which is selected using a random number generator. In the case of map accuracy, simple random sampling is performed without replacement, which means that each unit can be selected only once.

The sample estimates derived from simple random sampling have desirable statistical properties. The estimate of the mean is consistent and unbiased. An estimate is said to be consistent if the estimate equals the population parameter when the entire population is sampled (i.e., $n=N$). An estimate is said to be unbiased if the average value of the estimate at sample size n over all possible samples is equal to the population parameter. Only the mean or proportion is independently estimated in simple random sampling of proportions. The variance is computed from the mean and, therefore, is a function of it. The equations for all the sample estimates can be found in most sampling texts including Kish (1965) and Barrett and Nutt (1979) as well as in Congalton (1984).

Cluster Sampling. Cluster sampling is a method of sampling in which the sample units are not single pixels but, rather, groups (clusters) of pixels (population elements) where each element (pixel) must be unique to only one sampling unit or cluster (Kish 1965). Cluster sampling is performed for many reasons, including convenience and cost savings. It is much easier and cheaper to visit a few large areas than it is to visit many smaller areas. The disadvantages of cluster sampling as compared to simple random sampling are that the variance for

a given sampling effort is greater for cluster sampling due to the homogeneity of elements in the clusters, and the complexity of the statistical analysis is greater for cluster sampling.

A very important factor to consider in cluster sampling is "ROH," the intraclass correlation coefficient. "ROH," rate of homogeneity, is a measure of the homogeneity of the elements within a cluster and is computed as follows (Kish 1965):

$$\text{ROH} = \frac{\frac{1}{\text{RE}} - 1}{\bar{m} - 1}$$

where

$$\text{RE} = \frac{1}{\bar{m}} + \frac{\bar{m} - 1}{\bar{m}} \frac{\bar{S}_W^2}{\bar{m} S_A^2}$$

in which

RE = relative efficiency,

$\bar{S}_W^2$ = average variance within clusters,

S_A^2 = the variance among clusters, and

$\bar{m}$ = the average cluster size.

When the pixels within a cluster are relatively homogeneous, the intraclass correlation coefficient is high. Conversely, when the pixels within the cluster are relatively heterogeneous, the intraclass correlation coefficient is low. Each cluster contains the most information when the pixels in it are most hetereogeneous or when the intraclass correlation is low. With this in mind, cluster size and cluster shape should be chosen to minimize "ROH."

Stratified Random Sampling. Stratified random sampling is a sampling method in which *a priori* knowledge about a population is used to divide or stratify the population into nonoverlapping subpopulations (strata) which are more similar than the original population (Barrett and Nutt, 1979). This sampling approach is used when one wants to know specific information about certain subpopulations, for administrative convenience, and to increase the precision of the estimates for the entire population (Cochran, 1977).

Stratification of the difference images in this study was done geometrically and not by land-use category. The reason for this stratification was not to take advantage of *a priori* information, but rather to assume stratification for administrative convenience so as to not add more information into the sampling method. If additional information had been added, then direct comparison of the stratified sample with the other four sampling approaches would have been confounded. Therefore, the difference image was stratified by dividing it into four equal parts. Simple random sampling was employed within each stratum to derive the sample estimates.

Systematic Sampling. A systematic sample is one in which the sample units (pixels) are selected at some equal interval over time or space (Barrett and Nutt, 1979). The starting point (first sample) is located at random and each successive unit is taken at a specified interval thereafter. The advantages of systematic sampling are the ease or convenience of obtaining the sample and the uniform spread of the sampled observations over the entire population (Cochran, 1977). The major disadvantage of systematic sampling is that the selection procedure implies that each unit in the population does not have an equal chance of being included in the sample. In addition, if the population contains some periodicity, then the regular spacing of the sampling units might result in unrepresentative samples (Berry and Baker, 1968).

Because each sample unit in systematic sampling does not have an equal chance of selection, there is no fully valid estimate of sampling error (i.e., variance). However, there are many ways to approximately estimate the sampling error (Yates, 1981). The simplest and most common method, and the one used in this study, is to assume that the sample is random and use the appropriate equations from simple random sampling. More complicated methods (e.g., Kish (1965) and Quenouille (1949)) are available.

Stratified Systematic Unaligned Sampling. As described by Berry and Baker (1968), stratified systematic unaligned sampling "... combines the advantage of randomization and stratification with the useful aspects of systematic sampling, while avoiding the possibilities of bias due to the presence of periodicities." Figure 1 shows how a stratified systematic unaligned sample is acquired. First, the population is divided into strata with the strata size being determined by the number of samples to be taken. This division into implied strata is the same as that which occurs in systematic sampling when one takes samples at some given fixed interval. Next, point 1 (see Figure 1) is selected at random. The x coordinate at point 1 is then used with a new random y coordinate to locate point 2, and again with a new random y coordinate to locate point 3, and so forth across the row. In the same way, the y coordinate at point 1 is used with a new random x coordinate to locate point 4 and so forth down the column. Point 5 is located using the random x coordinate from point 4 and the random y coordinate from point 2, while point 7 is located using the x coordinate of point 3 and the y coordinate of point 6, and so forth until each stratum contains a sample unit (Berry, 1962).

The sample estimates derived from stratified systematic unaligned sampling are obtained by assuming simple random sampling and using the appropriate equations. Again, the reason for making this assumption is that it is commonly made when applying this sampling approach, and the object of this paper is to compare common approaches.

RESULTS

The objective of this study was to compare sample statistics, obtained through computer simulation, with the corresponding population parameters for five sampling schemes commonly used in assessing the accuracy of maps generated from remotely sensed data. A major factor influencing this accuracy is spatial autocorrelation, the details of which are presented in Congalton (1988). All the analysis, both spatial autocorrelation and sample simulations, was performed on the difference images. These images were generated by comparing, on a pixel by pixel basis, the Landsat classification with the assumed correct reference data for each of the study areas. Table 1 presents the population

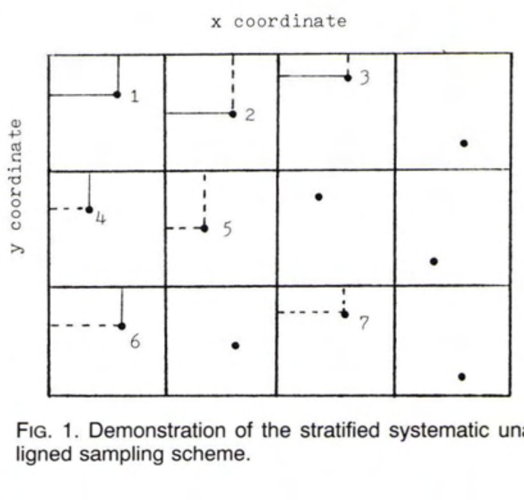

FIG. 1. Demonstration of the stratified systematic unaligned sampling scheme.

parameters for the three difference images. The image sizes vary somewhat because the agriculture data set is slightly larger than the other two data sets and also because the pixel size varies between images.

The results of the spatial autocorrelation analysis indicate that each difference image exhibits positive autocorrelation. In other words, the errors are clumped or clustered together. This pattern of error has a definite affect on the sampling simulations as we will see.

Sampling Simulations

The results of the sampling simulations were graphically displayed with the sample statistic (i.e., sample mean or sample variance) on the y-axis and the number of samples or sample size on the x-axis. For all the graphs, except those for cluster sampling, the plotted values were the average sample statistic for a particular sample size. The average sample statistics plotted were computed by averaging 400 repetitions at each sample size. In other words, the computer was used to simulate the taking of a given sample 400 times, and the average result for each sample size was plotted on the graph. Repeating the sample 400 times and taking the average eliminates the problem of the odd chance of obtaining an unrepresentative sample, which is always possible in any sampling scheme.

In the cluster sampling, the sample size was held constant and the cluster size was tested. Cluster sizes of 2, 3, 4, 5, 8, and 10, which correspond to cluster shapes of 1 by 2, 1 by 3, 2 by 2, 1 by 5, 2 by 4, and 2 by 5, were chosen for reasons that will be discussed later in the section on intracluster correlation. Therefore, in the cluster sampling, the plotted values were the average sample statistic for a particular cluster size. The actual population parameter was also plotted on each graph as a horizontal line orginating from the appropriate place on the y-axis.

The results of the sampling simulations for each of the three difference images were plotted on as few graphs as possible. The results for the sample mean of simple random sampling (SRS), stratified random sampling (STRAT), systematic sampling (SYSTEM), and stratified systematic unaligned sampling (SSUS) are presented on a single graph for each difference image. The results for the sample variance of simple random sampling (SRS), systematic sampling (SYSTEM), and stratified systematic unaligned sampling (SSUS) are presented on a single graph for each difference image. The results for the variance of stratified random sampling (STRAT) require their own graph. These results are not directly comparable to the other schemes because the effect of stratification is to reduce the variance in each stratum. For this reason and for space limitations, the graphs will not be presented in this paper. They are available in Congalton (1984). The results for the cluster sampling also require individual graphs for the mean and variance of each difference image. In addition to the graphs, sample sizes in percent are reported for each difference image. These percentages are reported to provide some input on minimum sample sizes necessary for accuracy assessment.

Figure 2 shows the results for the sample mean on the agricultural difference image using simple random sampling, stratified random sampling, systematic sampling, and stratified systematic unaligned sampling. Note that the sample estimates for stratified systematic unaligned sampling and systematic sampling overestimate the population mean while the sample estimates for the stratified random sampling and the simple random sampling closely estimate the population mean. Figure 2 also shows the results for the sample variance on the agriculture difference image. Remember that stratified random sampling is not included for reasons already discussed. As in the case of the sample mean, the sample estimates for the stratified systematic unaligned sampling and the systematic sampling overestimate the population variance. The sample estimates for simple random sampling closely follow the population variance. The sample sizes for the agricultural difference image ranged from 0.05 percent to 1.40 percent of the entire image.

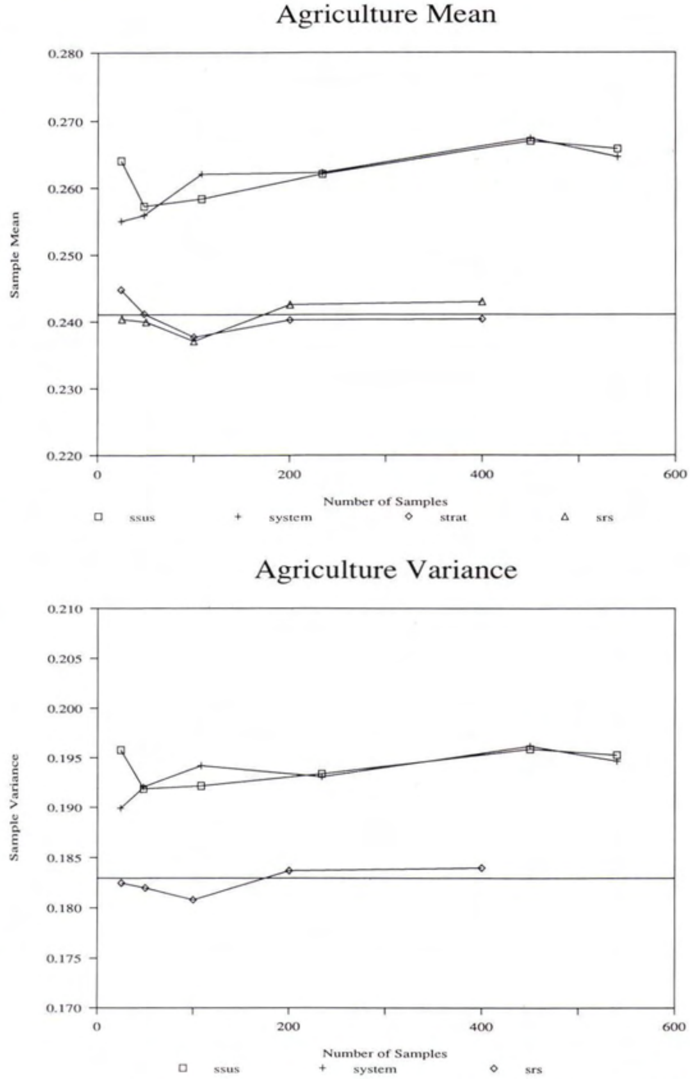

FIG. 2. The average of the sample statistics at each sample size for 400 repetitions using simple random sampling, systematic sampling, stratified random sampling (sample mean only), and stratified systematic unaligned sampling on the agriculture difference image (the horizontal line is the population parameter).

The results for the sample mean and sample variance on the range difference image using simple random sampling, stratified random sampling (sample mean only), systematic sampling, and stratified systematic unaligned sampling are presented in Figure 3. These results look very similar to the results for the agricultural difference image. Again, stratified systematic unaligned sampling and systematic sampling overestimated the population statistics. The only exception to this overestimation is at the largest sample size of 600 pixels. The results for both the stratified random sample and the simple random sample were almost identical and closely estimate the sample mean. Also, simple random sampling clearly provides the best estimate of sample variance. The sample sizes for the range difference image were between 0.07 percent and 1.50 percent of the entire image.

Figure 4 shows the results of the four sampling schemes for estimating sample mean and the results of the three sampling

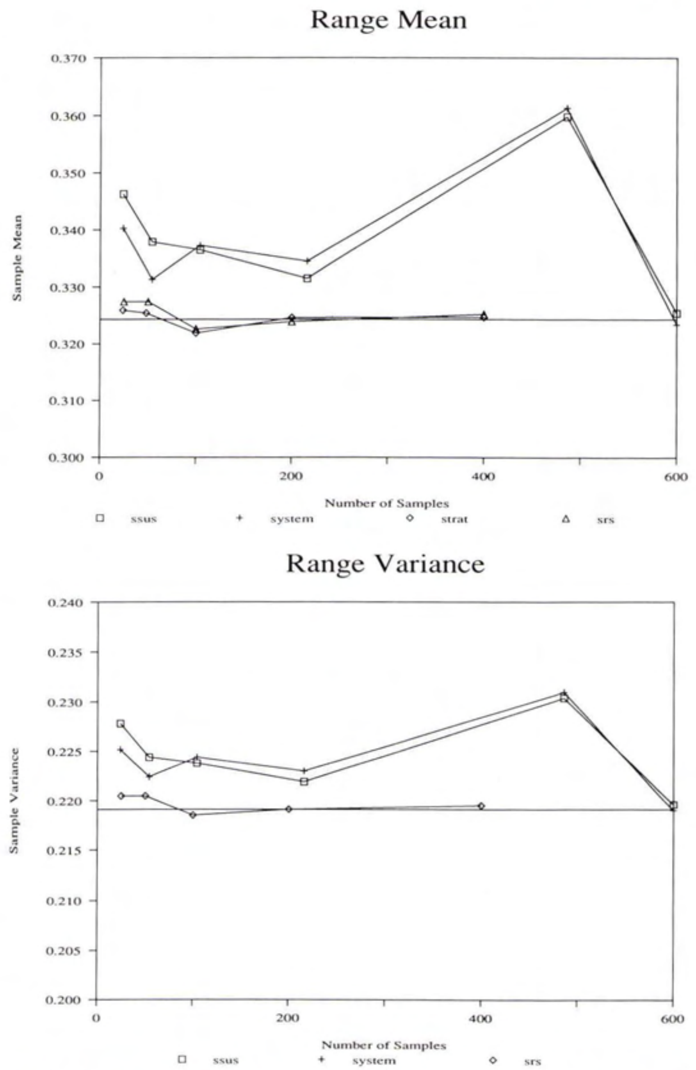

FIG. 3. The average of the sample statistics at each sample size for 400 repetitions using simple random sampling, systematic sampling, stratified random sampling (sample mean only), and stratified systematic unaligned sampling on the range difference image (the horizontal line is the population parameter).

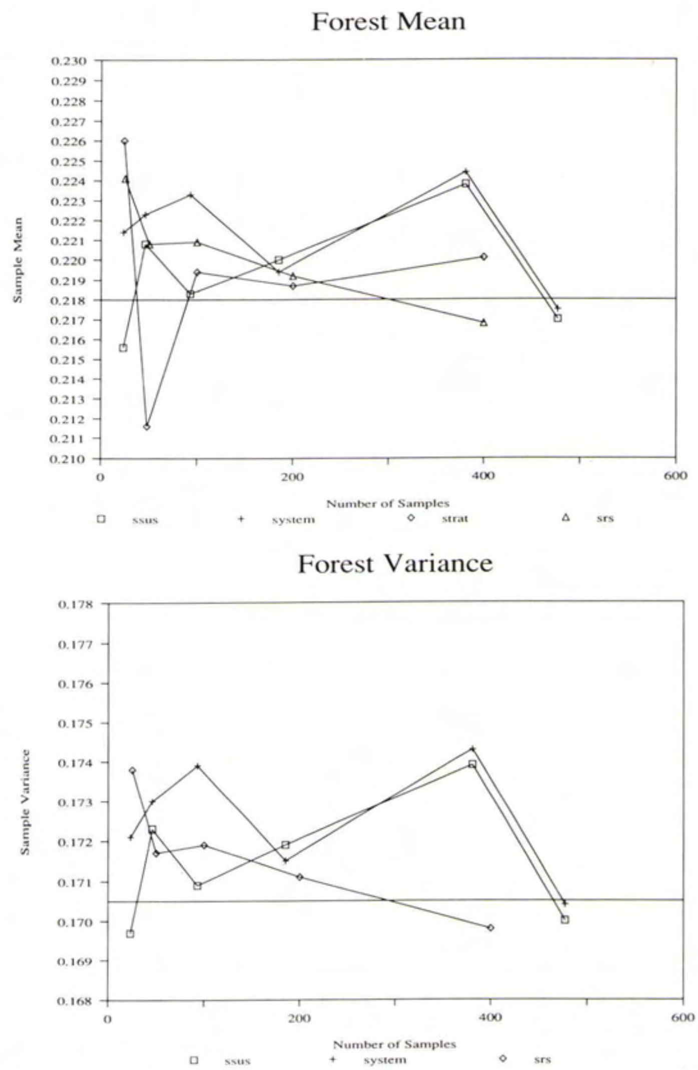

FIG. 4. The average of the sample statistics at each sample size for 400 repetitions using simple random sampling, systematic sampling, stratified random sampling (sample mean only), and stratified systematic unaligned sampling on the forest difference image (the horizontal line is the population parameter).

schemes (not stratified random sampling) for estimating the sample variance on the forest difference image. Note that the scales of the graphs here are different from those of the other graphs. The range of sample means is from 0.211 to 0.226 and the range of sample variances is from 0.169 to 0.174 which are smaller than those of the other graphs. The reason for this range change is obvious from looking at the graphs. All the sampling schemes do a fairly good job of estimating the population mean and variance. However, the simple random sampling scheme and the stratified random sampling scheme best approximate the population mean, especially when the sample size is 100 pixels or more. The same can be said of simple random sampling for the sample variance. At a sample size of less than 100, there is considerable variation within the results for all four schemes. The stratified systematic unaligned sampling and the systematic sampling continue to vary over the entire range of sample sizes. The sampling schemes for the forest difference image used sample sizes that ranged from 0.07 percent to 1.70 percent of the entire image.

Figures 5, 6, and 7 present the results of the clustering sample scheme for the agriculture, range, and forest difference images, respectively. Remember that the results here are not for different sample sizes, but rather for various cluster sizes. The sample size was held constant at 200 pixels while the results are for 400 repetitions at each cluster size. All cluster sizes investigated performed adequately. However, a cluster size of 2 best approximated the population mean for the agriculture and the forest difference images, while a cluster size of 5 was best for the range difference image. As for the sample variance, a cluster of size 2 best approximated the population variance for the agriculture difference image, clusters of sizes 4 and 5 for the range difference image, and a cluster of size 10 for the forest difference image. More important are the results presented in Figure 8. This graph shows the plot of average intracluster correlation coefficients versus cluster size for the three difference images. Note that the intracluster correlation coefficient begins high for small cluster sizes and decreases as the cluster size increases.

DISCUSSION

The results of the sampling simulations for the five sampling schemes on the three difference images were influenced greatly

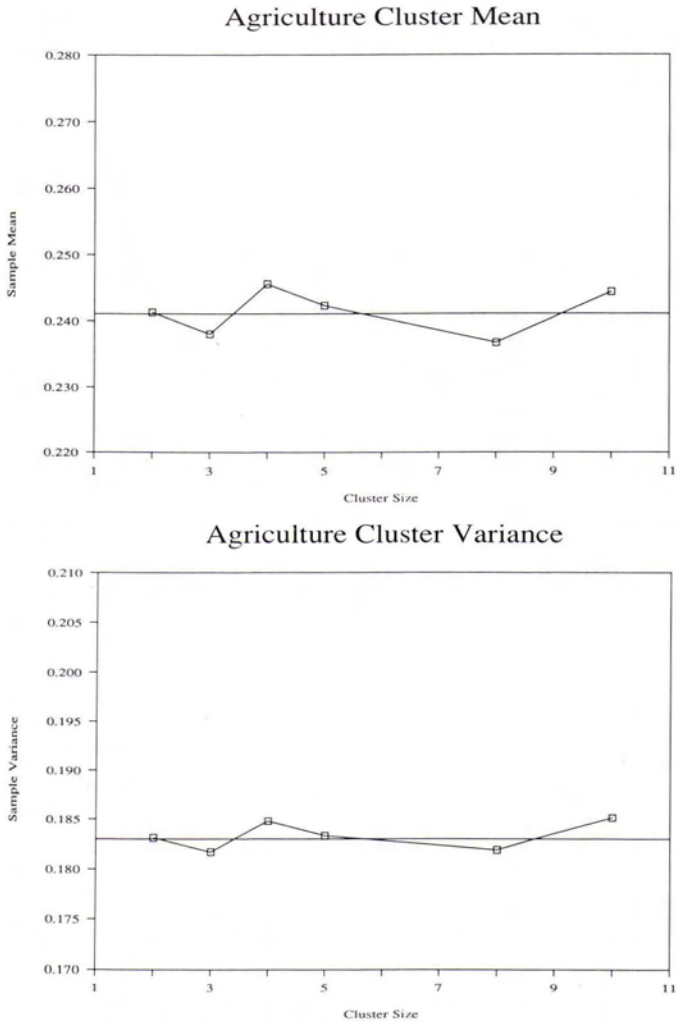

FIG. 5. The average of the sample statistics at each cluster size for 400 repetitions using cluster sampling on the agriculture difference image.

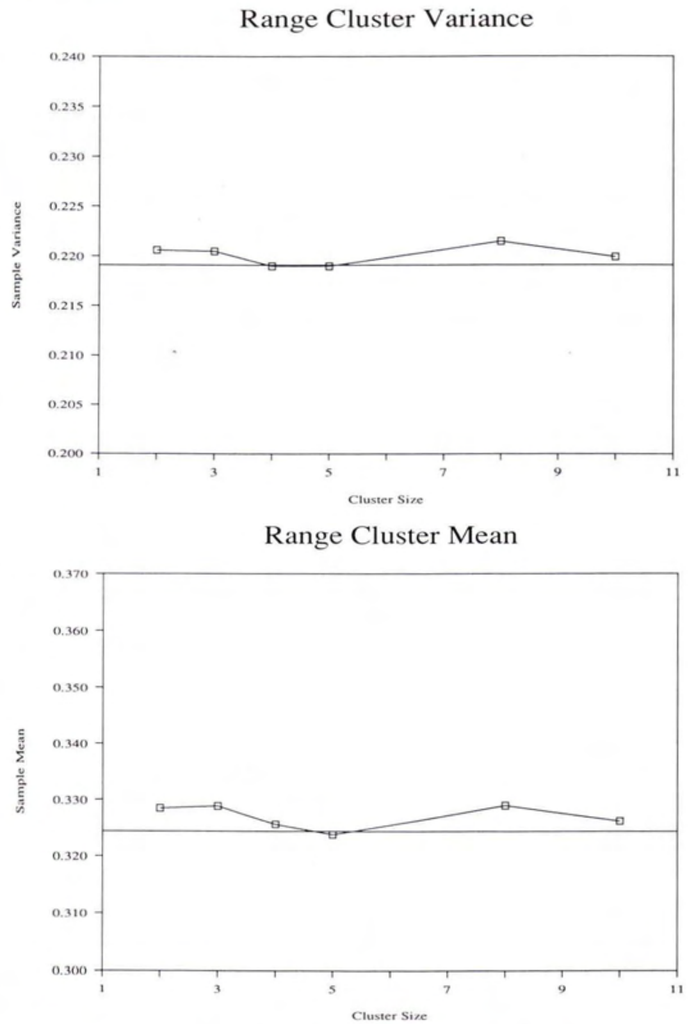

FIG. 6. The average of the sample statistics at each cluster size for 400 repetitions using cluster sampling on the range difference image.

by the pattern of error (spatial autocorrelation) within each difference image. In all three difference images, simple random sampling performed well in estimating the population statistics while the other four sampling schemes yielded different results for each of the three difference images. In the case of the agriculture difference image, stratified random sampling performed almost as well as simple random sampling. In addition, cluster sampling provided adequate estimates so as to be useful in the situation dictates the need for a cluster sample. Such instances might include using a cluster to deal with small misregistration problems in the data or when sampling over tremendously large areas such as in Alaska. According to these results for the agriculture study area, stratified systematic unaligned sampling and systematic sampling should not be used in assessing the errors in remotely sensed data. The results also indicate that sample sizes in the range of a 1 percent sample should be obtained if possible.

The results of the sampling simulations for the range study area indicate that the stratified random sampling scheme and the simple random sampling scheme perform the best. As in the agriculture data, a sample size around 1 percent of the image is required. Neither the stratified systematic unaligned sampling scheme nor the systematic sampling scheme performed well for the range study area except at a sample size of 600 pixels. The reason for this problem is periodicity in the data caused by the spatial autocorrelation between errors (see autocorrelation results in Congalton (1988)). Therefore, neither of these sampling schemes should be used in range environments for assessing the accuracy of remotely sensed data. Again, cluster sampling performed adequately enough, especially at a cluster of size 5, to be used if the situation demands it.

The results of the sampling simulations on the forest study area differ significantly from the other two study areas. All five of the sampling schemes do an adequate job of estimating the population mean. However, the best results are still achieved with simple random sampling followed by stratified random sampling. Unlike their performance in the agriculture and range difference images, stratified systematic unaligned sampling and systematic sampling adequately estimate the population parameter in the forest difference image. The reason for these sampling schemes working here and not in the other difference images has to do with the complexity of the forest study area. The pattern of error in the forest difference image is more complex and linear than the more simple, blocky pattern found in the range and especially the agriculture difference image. This complexity is clearly shown in the graphs of autocorrelation (Congalton, 1988). These graphs also demonstrate pronounced periodicity in the agriculture difference image, slightly less in the range difference image, and considerably less in the forest difference image. Therefore, although some periodicity is evi-

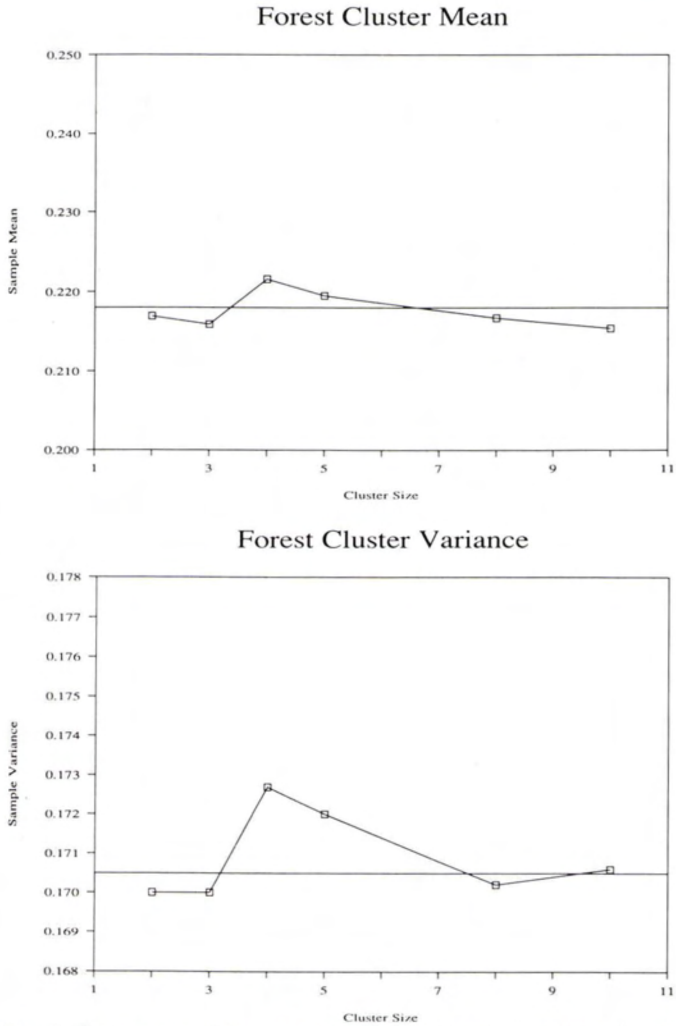

FIG. 7. The average of the sample statistics at each cluster size for 400 repetitions using cluster sampling on the forest difference image.

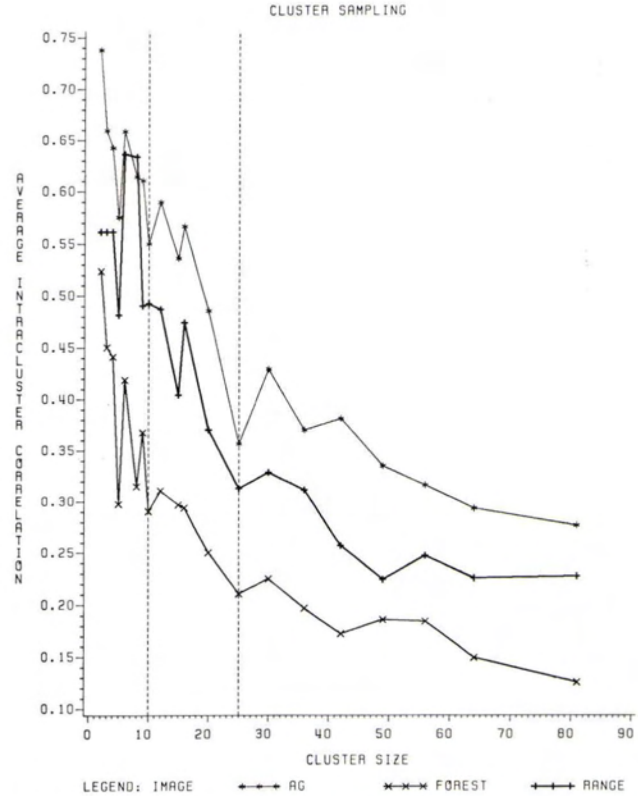

FIG. 8. A plot of the average intracluster correlation coefficient versus cluster size for the three difference images.

dent in the forest difference image, it is not present to the same extent as in the other two images. For this reason, the two sampling schemes more closely approximate the population mean in the forest study area. The results of the cluster sample show that it is adequate in estimating the population statistics. A cluster of size 2 appears to be the best although others could be used if so warranted by the situation. Sample sizes of around 1 percent and maybe even greater are dictated by the results, especially if stratified systematic unaligned sampling or systematic sampling is to be used.

The results of the analysis of average intracluster correlation coefficient are most interesting. Traditionally, cluster sampling has been a popular method of collecting ground data for comparison with remotely sensed data. In some instances, very large clusters up to a size of 100 pixels were taken. The results of this study show that clusters larger than about size 10 add very little new information to the cluster. In other words, the cost of adding another pixel to the cluster outweighs the information obtained from that pixel after the cluster reaches the size of about 10. This concept is well demonstrated by Figure 8. Notice that the value of the average intracluster correlation coefficient drops rapidly for about the first 10 pixels and then begins to level off. After the cluster reaches a size of 25, almost no change in the graph exists at all. Therefore, in order to maximize the information derived from cluster sampling in assessing the errors in remotely sensed data, no cluster should be taken that is larger than 10 or at most 25. In addition, Figure 8 also reveals something about the spatial complexity of the three difference images that was also brought out in the spatial autocorrelation analysis. Notice that the difference image with the highest intracluster correlation curve is the agriculture difference image. The agriculture study area is also the least complex and, therefore, would be expected to have the highest intracluster correlation (rate of homogeneity). Conversely, the most complex environment, forest, has the lowest intracluster correlation curve. Therefore, the spatial autocorrelation in the difference images also confirms the need for small cluster sizes.

CONCLUSIONS

Spatial complexity of a given environment dictates the appropriate sampling scheme(s) to be used for creating error matrices necessary to assess the accuracy of maps generated from remotely sensed data. In the agriculture, range, and forest areas used in this study, simple random sampling always performed adequately and can be used in all situations. Stratified random sampling also performed well and should be used especially when it is necessary to make sure that small, but important, areas are represented in the sample. Cluster sampling also can be used if the situation dictates it. However, small clusters should be taken using no more than 10 pixels per cluster or 25 pixels per cluster, maximum. Stratified systematic unaligned sampling and systematic sampling should be used only with extreme caution. Depending on the complexity of the area as determined by spatial autocorrelation analysis, these two schemes may yield adequate results. However, periodicity in remotely sensed data due to the positive correlation between errors could result in a very poor estimate of the population parameters if these schemes are used inappropriately.

Analysis of this type should be extended to higher spatial resolution Thematic Mapper data to see if these relationships hold. Initial results of accuracy assessments reported in the literature tend towards anticipating that spatial complexity will be a more important measure in Thematic Mapper data than it obviously is in MSS data. Its importance is expected to increase even more when using SPOT data.

ACKNOWLEDGMENTS

I would like to thank the Applications Branch scientists at the EROS Data Center for their help in this study. Special thanks go to Mr. Wayne Miller, Mr. Thomas Loveland, Mr. Mark Shasby, Mr. Jeff Newcomer, and Mr. John Szajgin. I would also like to acknowledge funding for this project by the Nationwide Forestry Applications Program and McIntire-Stennis project 4633-MS.

REFERENCES

Barrett, James P., and Mary E. Nutt, 1979. *Survey Sampling in the Environmental Sciences: A Computer Approach.* Compress, Inc., P.O. Box 102, Wentworth, N.H. 319 p.

Berry, Brian J. L., 1962. *Sampling, Coding, and Storing Flood Plain Data.* USDA Agriculture Handbook No. 237, 27 p.

Berry, B. J. L., and A. M. Baker, 1968. Geographic sampling. Spatial Analysis: A Reader. *Statistical Geography,* (B. J. L. Berry and D. F. Marble, eds.) Prentice Hall, Englewood Cliffs, N. J. pp. 91–100.

Card, D. H., 1980. *Testing the Accuracy of Land Use and Land Cover Maps: A Short Review of the Literature.* NASA Ames Research Center. 18 p.

Cochran, William G., 1977. *Sampling Techniques.* John Wiley & Sons. New York. 428 p.

Congalton, Russell G., 1984. *A Comparison of Five Sampling Schemes Used in Assessing the Accuracy of Land Cover/Land Use Maps Derived from Remotely Sensed Data.* Ph. D. Dissertation, Virginia Polytechnic Institute and State University. 147 p.

———, 1988. Using Spatial Autocorrelation Analysis to Explore the Errors in Maps Generated from Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing,* Vol. 54, No. 5, pp. 587–592.

Congalton, Russell, Richard Oderwald, and Roy Mead, 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. *Photogrammetric Engineering and Remote Sensing,* Vol. 49, No. 12, pp. 1671–1678.

Fitzpatrick-Lins, K., 1978a. Accuracy of Selected Land Use and Land Cover Maps in the Greater Atlanta Region, Georgia. *Jour. Research USGS,* Vol. 6, No. 2, pp. 169–173.

———, 1978b. Accuracy and Consistency Comparisons of Land Use and Land Cover Maps Made from High-Altitude Photographs and Landsat MSS Imagery. *Jour. Research USGS,* Vol. 6, No. 1, pp. 23–40.

Ginnevan, M. E., 1979. Testing Land-Use Map Accuracy: Another Look. *Phototgrammetric 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. 40, 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.

Hudson, W. D., and C. W. Ramm, 1987. Correct Formulation of the Kappa Coefficient of Agreement. *Photogrammetric Engineering and Remote Sensing,* Vol. 53, No. 4, pp. 421–422.

Kish, Leslie, 1965. *Survey Sampling.* John Wiley & Sons. New York. 643 p.

Ling, H. S., G. H. Rosenfield, and K. Fitzpatrick-Lins, 1979. *Sample Selection for Thematic Map Accuracy Testing.* USGS circular.

Quenouille, M. H., 1949. Problems in Plane Sampling. *Annals of Mathematical Statistics,* Vol. 20, pp. 355–375.

Rhode, W.G., 1978. Digital Image Analysis Techniques for Natural Resource Inventories. *National Computer Conference Proceedings,* pp. 43–106.

Rosenfield, G. H., 1982. Sample Design for Estimating Change in Land Use and Land Cover. *Photogrammetric Engineering and Remote Sensing,* Vol. 48, No. 5, pp. 793–801.

Rosenfield, G. H., and K. Fitzpatrick-Lins, 1986. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering and Remote Sensing,* Vol. 52, No. 2, pp. 223–227.

Rosenfield, G. H., and M. L. Melley, 1980. Applications of Statistics to Thematic Mapping. *Photogrammetric Engineering and Remote Sensing,* Vol. 48, No. 5, pp. 1287–1294.

Rudd, R. D., 1971. Macro Land Use Mapping with Simulated Space Photographs. *Photogrammetric Engineering,* Vol. 37, pp. 365–372.

Story, M., and R. Congalton, 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing,* Vol. 53, No. 3, pp. 397–399.

Todd, W. J., D. G. Gehring, and J. F. Haman, 1980. Landsat Wildland Mapping Accuracy. *Photogrammetric Engineering and Remote Sensing,* Vol. 46, No. 4, pp. 509–520.

van Genderen, J. L., B. F. Lock, and P. A. Vass, 1978. Remote Sensing: Statistical Testing of Thematic Map Accuracy. *Proc. Twelfth International Symposium on Remote Sensing of Environment.* ERIM. pp. 3—14.

Yates, Frank, 1981. *Sampling Methods for Censuses and Surveys.* Charles Griffin & Company Ltd. London. 458 p.

Zonneveld. I. S., 1974. Aerial Photography, Remote Sensing and Ecology. *ITC Journal,* Part 4. pp. 553–560.

(Received 29 August 1987; revised and accepted 21 January 1988)

Forthcoming Articles

F. Becker, F. Nerry, and *P. Ramanantisizehena,* Towards a Quantitative Use of Aerial Color Infrared Films.
Y. Doytsher and *B. Shmutter,* A New Approach to Monitoring Data Collection in Photogrammetric Models.
R. W. Gaskell, Digital Identification of Cartographic Control Points.
U. V. Helava, Object-Space Least-Squares Correlation.
Michael A. Karteris, Manual Interpretation of Small Forestlands on Landsat MSS Data.
C. P. Lo, Comparative Evaluation of the Large Format Camera, Metric Camera, and Shuttle Imaging Radar-A Data Content.