STUDY OF SAMPLING METHODS FOR ACCURACY ASSESSMENT OF CLASSIFIED REMOTELY SENSED DATA

M.S.Hashemian [a], A.A.Abkar [b], S.B.Fatemi [b]

[a] Department of image processing, National Cartographic Centre (NCC) of Iran, Meraj Av., Azadi Sq., Tehran, Iran, hashemian@ncc.neda.net.ir
[b] Faculty of Geodesy and Geomatics Eng., KN Toosi University of Technology, Vali_Asr Street, Mirdamad Cross, Tehran, Iran, abkar@itc.nl, sbfatemi@yahoo.com

KEY WORDS: Remote Sensing, Land cover, Classification, Sampling, Accuracy

ABSTRACT:

Accuracy assessment is an important step in the process of analyzing remote sensing data. It determines the value of the resulting data to a particular user, i.e. the information value. Remote sensing products can serve as the basis for political as well as economical decisions. Users with a variety of applications should be able to evaluate whether the accuracy of the map suits their objectives or not. In the conventional accuracy assessment an error matrix and some accuracy measures derived from it are used. An error matrix is established using some known reference data and corresponding classified data. There are various factors that affect the performance of the accuracy assessment by influencing the error matrix through out the ground truth data collection. In practice, the techniques are of little value if these effective factors are not considered. In this paper the necessity considerations for accuracy assessment including the sampling schemas and the sample size for these sampling methods are studied. Also the factors that affect selecting and applying appropriate sampling schemas and sample size are investigated. For this study numbers of synthetic images and one real image and some reference data are used. Sensitivity of the various sampling schemas has been investigated using the synthetic images and using the real image the obtained results have been confirmed. The results represent that depend on specific conditions such as type and size of the study region and object characteristics, different sampling methods and sample sizes are preferred.

1. INTRODUCTION

The value of the classified map is clearly a function of the accuracy of the classification. A responsible use of the stored geodata is only possible if the quality of these data is known. Then accuracy assessment is an important step in the process of analyzing remote sensing data. In the conventional accuracy assessments an error matrix is established using some known reference data and corresponding classified data and some accuracy measures derived from it are used for accuracy assessment. Assessing the accuracy of land cover map 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 improper sampling approach can be costly and yield poor results. To wit, poor choice in sampling scheme can result in significant biases being introduced into the error matrix, which may over or under estimate the true accuracy. In this paper the suitability of five sampling methods including simple random sampling, stratified random sampling, systematic sampling, stratified systematic unaligned sampling and cluster sampling are investigated with some synthetic images and one real image.

2. DATA USED

In this paper, in order to have a controllable process and reliable investigation, some experiments are performed using generated synthetic images. Also, for comparison of the different approaches that have been hindered by the scarcity of satellite imagery for which full reference data are available, use of simulated datasets is one means of overcoming this problem. One real image is used for confirmation of the obtained results from synthetic images. The synthetic images are:

- Image number 1 is a quadrangle area with three generated bands and 200 pixels in rows and 200 pixels in columns and has 10 classes and large fields. There are 10 fields in the image totally and the average area of the fields is 4000 pixels.
- Image number 2 is a quadrangle area with three generated bands and 200 pixels in rows and 200 pixels in columns and has 10 classes and small fields with good distribution in total of image. The average area of the fields is 99 pixels.
- Image number 3 is a quadrangle area with three generated bands and 512 pixels in rows and 512 pixels in columns and has 10 classes and small fields with good distribution in total of image. The average area of the fields is 119 pixels
- Image number 4 is an image with three bands (R, G, B) and 768 pixels in rows and 576 pixels in columns. This image is a RGB-CCD image of a model of agricultural fields that means a color CCD video camera has produced three bands (R,G,B) of it, this image has 9 classes consist of two types of roads, five different crop fields, farmhouses and forest (Abkar, 1999). The average area of the fields is 29491 pixels.

For generation of first three synthetic images firstly it has been assumed that our data has a known normal distribution. Then by having this assumption, some values for mean vectors and covariance matrices have been considered and using them the pixel values, on the basis of normal distribution have been generated. The values of means and standard deviations of each class for each band have been derived from a real case study (Fatemi, 2002). In the all cases, for generation of the pixel values, covariances between all of the bands were assumed to be zero. Images have three bands that were generated by the above consideration. For considering the noise that exists in all real images the random values of noise were added to the pixels in the synthetic images. Again these values of noise are according to normal distribution with certain means and variances.

The area of interest in real word is Moghan and is located in Ardebil province of Iran. One TM (Landsat 5) image of the study area acquired on 1998-06-08 is used for this study. 6 bands of this image are used for producing classified map.

For classification of images, based on produced reference data, training samples for each class were collected. Then Maximum Likelihood Classification based on equal prior probability of the classes was implemented.

3. SAMPLING SCHEMAS

Five sampling schemes typically are used in the accuracy assessment. These five sampling schemes are Simple Random Sampling (SRS), Cluster Sampling (CS), Stratified Random Sampling (STRAT), Systematic Sampling (SYSTEM), and Stratified Systematic Unaligned Sampling (SSUS) (Congalton, 1988). 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. Cluster sampling is a method of sampling in which the sample units are not single pixels but, rather, groups (clusters) of pixels. In this study cluster sampling is based on random selection of clusters. Stratified random sampling is a sampling method that divides or stratifies the population into nonoverlapping subpopulations (strata).

In this paper, stratification of images was done geometrically and by dividing images into four equal parts. A systematic sample is one in which the sample units (pixels) are selected at some equal interval over space. In stratified systematic unaligned sampling with random selecting of samples in each strata produced from stratification of image in specified intervals, a combination of systematic and random sampling is used.

4. INVESTIGATIONS ON SAMPLING SCHEMAS

For investigating the sampling schemas used in the accuracy assessment, each of the five sampling schemas, were simulated. In each simulation overall accuracy of produced error matrix were calculated. Each simulation was repeated 30 times for each image with 50, 100, 200… 1000 samples per images for each sampling schema and the results were averaged together. This experiment was done in three cases and results of averaging overall accuracies have been graphed for each image. Overall accuracy was computed using the all pixels in reference data for each image without sampling. Thus, it is possible to compare the results of each sampling schema simulation with these true and actual values. This comparison allowed determining the best sampling schemas to use for each data set. The results of the sampling simulations were graphically displayed with the overall accuracy on the y-axis and the number of samples on the x-axis. In such graphs, the actual values were also plotted on each graph as a horizontal line originating from the appropriate place on the y-axis. The plotted values are the averaged overall accuracy for a particular sample size. Repeating the sampling several times and taking the average eliminates the problem of the odd chance of obtaining an unrepresentative sample, which is always possible in any sampling schema. In some below experiments the results of investigations are discussed.

4.1 Experiment #1: Investigation of Simple Random Sampling (SRS) Schema

The results of averaging overall accuracies with SRS method in four images have been shown in graphs of Figure 1.

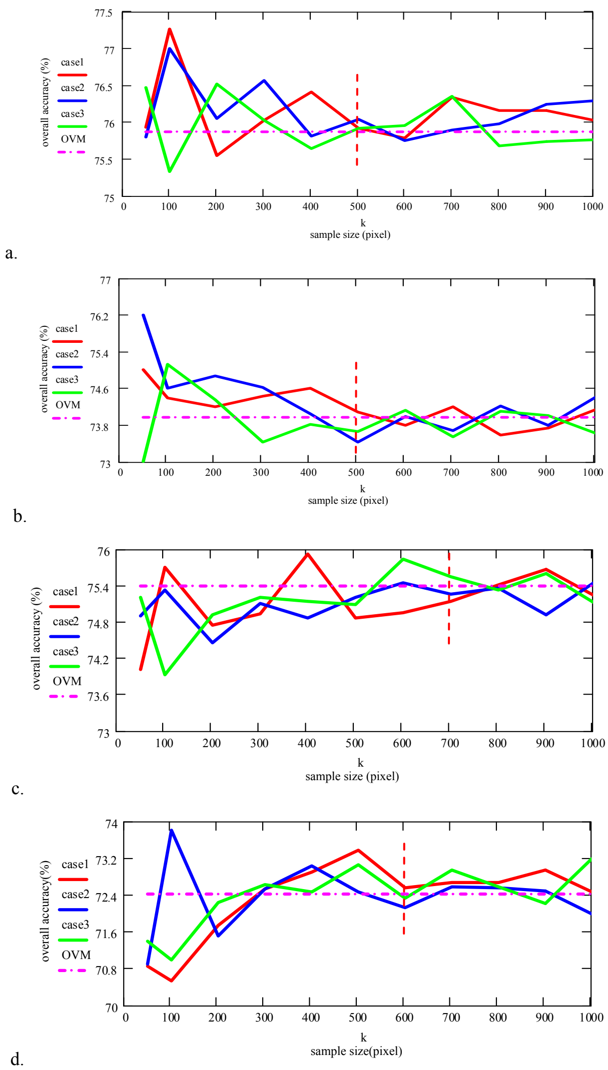

Figure 1. Overall accuracies resulted from using SRS method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged) for 3 cases in image#1 (a), image#2 (b), image#3 (c) and image#4 (d)

Figure 1(a) and Figure 1(b) that are related to images with smaller image size show that approximately with sample size larger than 50 samples for each class, the results go towards stability. In images with larger image size i.e. image#3 and image#4 (Figure 1(c) and Figure 1(d)), the results go towards stability after approximately 70 samples for each class. This result is accordance with a rule of thumb recommended by

Congalton that stats at least 50 samples and in large area at least 75-100 samples should be taken per class (Congalton, 1991).

In images with large fields, i.e. image#1 and image#4 (Figure 1(a) and Figure 1(d)), the graphs show that these results have tendency to overestimating actual overall accuracy. In fact in images with large fields optimistic estimation of errors is occurred. The reason for this matter is that in images with small fields, distribution and dispersion of classes and consequently errors in image is better and sampling with simple random method is more suitable.

For comparing results of SRS method in the all synthetic images, average and standard deviation of overall accuracies in each sampling cases after stability were computed and results of each three cases related to each image are averaged. The results were graphically displayed in Figure 2. The y-axis in Figure 2(a) shows differences of means after stability with actual overall accuracies and in Figure 2(b) shows standard deviations from means and in Figure 2(c) shows standard deviations from real values of overall accuracies. The x-axis shows image numbers that image number 5 is the real TM image.

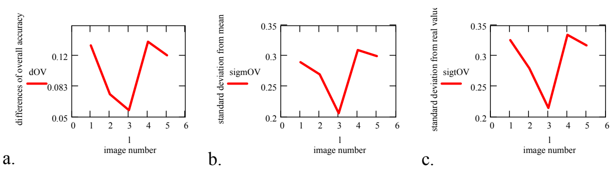

Figure 2. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b), and standard deviations from real values (c), using SRS method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged)

With respect to these graphs, the best results are related to image #2 and image#3 with smaller difference of means (Figure 2(a)) and standard deviations (Figure 2(b) and Figure 2(c)). Also with due attention to Figure 2(a) the overestimating in image#1 and image#4 is distinctive. In addition to largeness of difference values in these images in Figure 2(a), the value of standard deviations from real overall accuracies in these two images are clearly bigger than standard deviations from means (Figure 2(b) and Figure 2(c)).

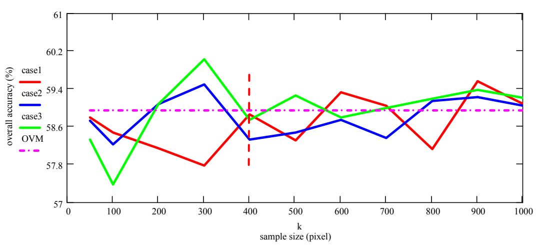

Figure 3. Overall accuracies resulted from using SRS method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged) for 3 cases in the real TM image

The results of SRS schema in real image has been graphed in Figure 3 that it shows, the results go towards stability almost after 50 samples for each class.

4.2 Experiment #2: Investigation of Stratified Random Sampling (STRAT) Schema

With due attention to graphs of overall accuracies using STRAT sampling schema in 3 cases, it was seen that with nearly 50 samples for each class, i.e. 500 samples in first three images with 10 classes and almost 400 samples in image #4 and image #5 with 9 classes, the results went towards stability, either for larger images or smaller images. So, produced samples with this sampling schema have better distribution in image relative to SRS method, therefore, with fewer samples, good results are achieved.

In the graph of image #4, continuously, the overestimating of results was seen. Because this image is an image with large fields and large size, that this sampling method can not sample this image in a good way. But in image #1 in spite of having large fields because of smallness of image size the results are better. From this, it is concluded that the size of image is an effective factor for STRAT method.

Also with comparing image #2 with image # 3 in Figure 4, it is distinguishable that STRAT method has better results in images with smaller image size used in this paper. The nearness of means of overall accuracies to real amounts in image #2 (Figure 4(a)) that is smaller image in comparing to image #3, is better, in spite of same distribution of fields and classes in image.

On the other hand with due attention to results of image #1 and image #2 in Figure 4, the later has better results in mean of overall accuracy (Figure 4.4(a)), but the former has better standard deviation (Figure 4(b) and Figure 4(c)). Then results of these two images with the same image size have not advantage upon each other.

Totally, Stratified Random Sampling has better results in image with smaller image size, and with considering results of real image in Figure 4, this matter is well confirmed.

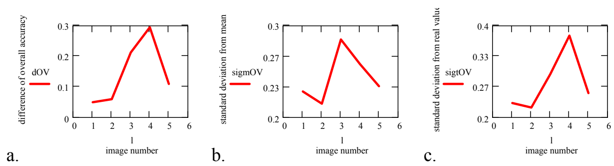

Figure 4. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b), and standard deviations from real values (c), using STRAT method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged)

The results of computing of overall accuracies with various sample sizes using stratified sampling schema in real TM image showed that the results go towards stability almost after 50 samples per class.

4.3 Experiment #3: Investigation of Systematic Sampling (SYSTEM) Schema

Studying of graphs of STRAT method showed that in this method (similar to STRAT method) the size of images is not an important factor for stability of results. These graphs showed that almost with more than 30 or 40 sample for each class, stable results are acquired, i.e. some more quickly than two previous methods. This case demonstrate that in this method samples have better distribution in image, then with less samples can achieve appropriate results. Also there was not overestimating tendency problem in these four study areas. This subject also is a confirmation about efficiency of this sampling schema in all of these images. In view of graphs of means and standard deviations in Figure 5 this subject is delineated that in images with large image size i.e. image #3 and image #4, means are closer to real values (Figure 5(a)) whereas standard deviations are appropriate (Figure 5(b) and Figure 5(c)), then it seems that this method is more efficient in large study areas.

If results of image #3 and image #4 in Figure 5 are compared together, it is seen that image #4 with larger field sizes has better results i.e. smaller difference of means (Figure 5(a)) and smaller standard deviations (Figure 5(b) and Figure 5(c)), then in addition to large image size, large object size is a factor for achievement of good results. In the case of real image, this matter is confirmed; this image with smaller image size than image #3 and because of larger field size has results better than image #3.

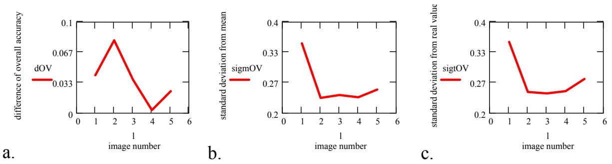

Figure 5. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b), and standard deviations from real values (c), using SYSTEM method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged)

With comparing Figure 5(a) with same graphs in previous methods (Figure 2(a) and Figure 4(a)), it is obvious that the range of values in this graph is lower than the others. Then this method is more appropriate from other methods. The results of computing overall accuracies in real image show that the results go towards stability after almost 50 samples for each class.

4.4 Experiment #4: Investigation of Stratified Systematic Unaligned Sampling (SSUS) schema

In this sampling method the same result as STRAT sampling about the stability of the means of overall accuracies was achieved. In other words with 50 sample for each class and without considering the size of images, the stability of results was acquired. The reason for this matter is that SSUS method has either random or systematic characteristics. Then results of SSUS method not as SYSTEM method with small sample size for all images and not as SRS method with large sample size for large images, but with almost 50 samples for each class go towards stability.

There is not overestimating problem in graphs of this method. This subject is also visible in comparing of graphs in Figure 6, because standard deviations from means (Figure 6(b)) are close to standard deviations from real overall accuracies (Figure 6(c)) in all of images. This subject is a reason for efficiency of this sampling schema in all of these images.

The best results have been produced in image #3 and image #4 that are images with large image size. These results are the best because of smallness of differences from real overall accuracies (Figure 6(a)) and standard deviations (Figure 6(b) and Figure 6(c)).

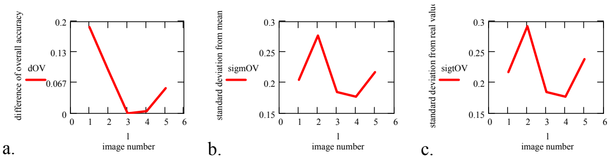

Figure 6. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b), and standard deviations from real values (c), using SSUS method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged)

The results of computing of overall accuracies with various sample sizes with stratified systematic unaligned sampling schema for the real TM image showed that the results go towards stability after 50 samples for each class.

4.5 Experiment#5: Investigation of Cluster Sampling (CS) Schema

In this experiment CS schema with cluster shape 3 by 3 is investigated. In this sampling method in all of images the means of overall accuracies go to stability after 60 or 70 samples for each class. These values for sample size are the largest values in comparing with the other methods. This case shows that in this method samples have not suitable distribution in image then with further samples can achieve the some better results.

In graphs of image #1 the overestimating of results are observed. In the case of image #4 although the overestimating tendency don't exist, this image has the maximum value of standard deviation (Figure 7(b) and Figure 7(c)) even with respect to graphs of the other methods. Then the results of this image are not appropriate too. Totally neither of these images has preferable and suitable results because if averages of calculations (Figure 7(a)) are small value, standard deviation of that (Figure 7(b) and Figure 7(c)) is large and vice versa. Then this method is not a suitable sampling schema in neither of images in this paper. The reason for this is that with cluster sampling, the distribution of samples in images is not suitable and the samples don't represent the population properly.

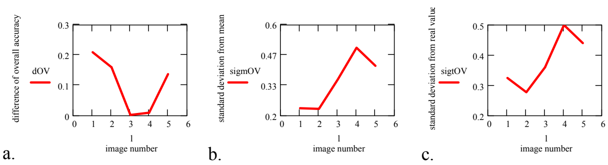

Figure 7. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b), and standard deviations from real values (c), using CS method (each sampling schema for each sample size has been repeated 30 times and the results have been averaged)

The results of the real TM image with this sampling schema showed that the results go towards stability after 70 samples for each class.

After previous experiments, it is possible to draw some new graphs using the same values. In Figure 8 these graphs has been shown. This figure is consisting of five graphs for each type of images used in this paper, in one diagram. In all of these graphs, x-axis introduces sampling methods with numbers of 1 to 5 that these are accordance to: SRS, STRAT, SYSTEM, SSUS and CS, and y-axis in graphs 8(a) shows the values of difference of means of overall accuracies with real values and in graphs 8(b) is the values of standard deviations from means.It is seen that in all of images systematic sampling is clearly better than the other methods because of minimum values of mean differences (Figure 8(a)) and almost most uniform and small standard deviations (Figure 8(b)). Then with assurance this sampling schema produces suitable results in all of these images.

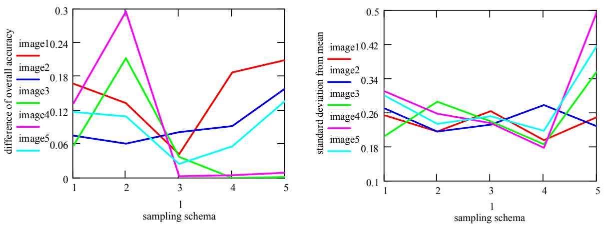

Figure 8. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b) using all of sampling methods for all of images with 30 times repeating, each color shows results of using all of methods in one image

On the other hand with graphically displaying the graphs of mean differences and standard deviations related to all of the sampling methods (graphs of Experiment#1 to 5) according to Figure 9, it is obvious that in image #2 that has small image size and small field size, in all of methods appropriate results are achieved. This image has the least differences in resulted mean differences (Figure 9(a)) and resulted standard deviations (Figure 9(b)), using all of the sampling schemas.

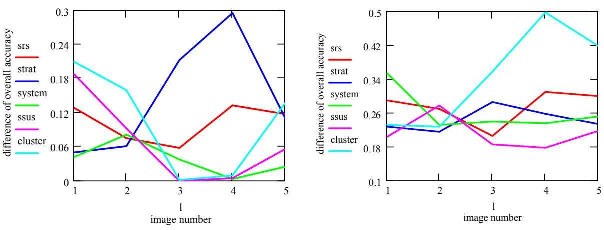

Figure 9. The difference of average overall accuracies after stability with actual overall accuracies (a) and standard deviations from means (b) in all of images using 5 sampling methods with 30 times repeating, each color shows results of one sampling methods in all of images

In addition to previous experiments and for confirmation of results, with sample size equal to 1000 as a confident sample size in each of sampling schemas, with one , 10 and 30 times calculating of overall accuracies several graphs has been produced that because of similarity of results with previous graphs, it has been restrained from displaying of them. However it is important to be realized that with one time sampling and calculating of overall accuracy, the previous results are not achieved, because the problem of the chance of obtaining an unrepresentative sample which is always possible doesn't eliminate. Then for assurance, these computations have to be done in repetitive manner and average of values must be considered as the final result (Hashemian, 2004).

5. CONCLUSIONS

With due attention to experiments in this paper, several results are achieved. Here these results are summarized:

1. With 50 to 70 samples for each class in all of images used in this paper, the results of overall accuracies go towards stability and the best results. Then 100 samples for each class in image can be a reliable sample size.
2. In images with large fields, SRS and STRAT methods overestimate the overall accuracy, but with SYSTEM and SSUS methods this problem doesn't exist.
3. SRS method has the best results for images with small fields used in this paper.
4. STRAT method produces better results in smaller study areas.
5. SYSTEM schema is the more efficient method in large images with large fields.
6. SSUS schema is a more suitable method for accuracy assessment in images with large size.
7. CS schema is not a suitable method in neither of the images used in this paper, and don't have the good results on comparing with the other methods.
8. Totally, the sampling schemas with systematic basis achieve rather more suitable results in all of images used in this paper.
9. In image with small image size and small field size (images with good distribution of classes), it is expectable that the good results are produced from each of these sampling schemas.
10. For achieving the best results, computations of accuracy assessment have to be done in repetitive manner and average of values be considered as final result.

REFERENCES

Abkar, A.A., 1999. *Likelihood-Based Segmentation and Classification of Remotely Sensed Images, A Bayesian Optimization Approach for Combining RS and GIS*, PhD-thesis, University of Twente, Enschede / ITC, Enschede, 132 p.

Congalton, R.G., 1988. A Comparison of Sampling Schemes Used in Generation Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data, *Photogrammetric Engineering & Remote Sensing*, 54(5), pp. 593 – 600.

Congalton, R.G., 1991. A Review of Assessing the Accuracy of Classification of Remotely Sensed Data, *Remote Sensing Environ*, 37, pp. 35 - 46.

Fatemi, S.B., 2002. *Model-based Image Analysis of Remotely Sensed Images: An Agricultural Case Study*, MSc Thesis, KN Toosi University of Technology.

Hashemian, M.H., 2004. *Study of accuracy assessment techniques for classification of remote sensing data*. MSc Thesis, KN Toosi University of Technology.