The Influence of Geographic Sampling Methods on Vegetation Map Accuracy Evaluation in a Swampy Environment

C.P. Lo and Lee J. Watson

Abstract

The effectiveness of five sampling methods was evaluated empirically to determine the thematic accuracy of six-class vegetation maps of two study areas in the Okefenokee Swamp produced by computer-assisted supervised classification of Landsat TM data. An exhaustive survey of the two study areas permitted estimates of overall, user's, and producer's accuracies to be made by each sampling method. The spatial pattern of classification errors was also examined with the aid of spatial autocorrelation analysis. By comparing Kappa coefficients of agreement among error matrices generated by the different sampling methods and by applying Chi-square tests, the stratified random sampling method was found to perform best for thematic accuracy evaluation in a swampy environment where vegetation exhibited a mixture of simple and complex spatial patterns. The higher spatial resolution of the image data employed to produce the vegetation map tended to cause stronger positive spatial autocorrelation in the resultant map error pattern.

Introduction

In remote sensing applications, the thematic accuracy of the land-use/land-cover maps generated from aerial photographs or satellite images is usually evaluated with the aid of a geographic sampling procedure. A number of points, traverses (lines), or quadrats (areas) are selected by means of a geographic sampling method from the land-use/land-cover maps. Ground truthing of these points, traverses, or quadrats is then conducted to produce an error matrix, from which the overall accuracy, producer's accuracy, and user's accuracy can be calculated (Story and Congalton, 1986). There are different types of geographic sampling methods or designs for land-use/land-cover map accuracy evaluation. The commonly used ones for thematic map accuracy evaluation include simple random, stratified random, systematic random, stratified systematic unaligned, and random clusters (Berry and Baker, 1968; Clark and Hosking, 1986; Griffith and Amrhein, 1991). Previous research by Congalton (1988a) using three types of land covers with varying degrees of spatial complexity (agriculture, range, and forest) generated by Landsat MSS images suggested that spatial complexity of a given environment would affect the choice of an appropriate sampling method. He found that simple random sampling was adequate in all situations, while stratified random sampling was also appropriate, especially when small areas were represented in the sample. However, the results of a stratified systematic unaligned sampling and systematic sampling varied depending on the spatial complexity of the image, and should be used with extreme caution. Because of the "systematic" nature of these two sampling methods, periodicity in remotely sensed data would also adversely affect their performance. It was further suggested that this type of analysis should be extended to higher spatial resolution Thematic Mapper (TM) data or SPOT data to see whether the same conclusions applied. In a subsequent research conducted by Stehman (1992) using simulated periodic patterns and actual land-use/land-cover derived from TM images, it was confirmed that systematic sampling methods were generally more precise and efficient than simple random sampling. When strong periodicity in the spatial pattern of misclassifications was suspected, systematic sampling and stratified systematic unaligned sampling should be employed only if the unfavorable sampling interval could be avoided.

This research is an attempt to extend the empirical evaluation of the influence of sampling methods on thematic accuracy to higher spatial resolution TM data based on the vegetation pattern of the Okefenokee Swamp, which consists of forested upland, forested wetland, scrub-shrub wetland, emergent wetland, and aquatic bed (Hamilton, 1982). The swampy environment provides varying degrees of spatial complexity, which exhibits some of the characteristics of the forested, rangeland, and agricultural environments employed in Congalton's (1988a) study, for the evaluation of the influence of geographic sampling methods on thematic mapping accuracy.

Characteristics of Geographic Sampling Methods

The geographic sampling methods employed in this research are simple random, stratified random, systematic random, stratified systematic unaligned, and random cluster, each of which possesses distinctive properties.

In simple random sampling, each point is chosen independently and randomly. Point locations are defined by coordinate positions. The x and y coordinates of these points are derived using either a random numbers table or a random numbers generator. Simple random sampling is usually conducted "without replacement," meaning that each sample point can only be chosen once, and every point has an equal chance of being chosen. In remotely sensed data, pixels are so numerous that simple random sampling may be justifiable (Griffith and Amrhein, 1991).

C.P. Lo is with the Department of Geography, University of Georgia, Athens, GA 30602 (chpanglo@uga.cc.uga.edu).

L.J. Watson is with 3D Imaging, LLC, 29111 Newnam Road, Easton, MD 21601.

Photogrammetric Engineering & Remote Sensing,
Vol. 64, No. 12, December 1998, pp. 1189–1200.

0099-1112/98/6412–1189$3.00/0

Stratified random sampling involves subdividing the study area into strata. Sampling points within each stratum are then chosen randomly. Two approaches are possible. The first is to subdivide the study area into strata of roughly equal size, within which sample points are randomly chosen. This will ensure a more even spread of sample points throughout the study area. The second approach is to use strata of different sizes so that a larger proportion of sample points can be chosen within a particular stratum of interest. In evaluating a land-use/land-cover map, it is obviously possible to also stratify according to the land-use/land-cover categories, and within each category a specific number of sample points are randomly chosen. Stratified sampling ensures that a more representative set of sample points is chosen than in the case of simple random sampling. Stratification helps reduce internal variability and, hence, sampling error. This means that fewer sample points are required than those in simple random sampling for the same level of accuracy. Based on studies by Rudd (1971), Zonneveld (1972), and Van Genderen and Lock (1977), stratified random sampling has been accepted as the most appropriate method of sampling in land-use studies using remotely sensed images so that smaller areas can be satisfactorily represented. There are, however, some disagreements on the most statistically sound method of allocating sample points to each land-use/land-cover category. Proponents of Hay (1979) feel that all thematic categories should receive an equal number of sample points regardless of the actual size of the category. Others argue that allocation should be proportional to categorical size, because simple random and stratified random techniques have a tendency to oversample larger categories and undersample smaller categories (Card, 1982; Fitzpatrick-Lins, 1981; Quirk and Scarpace, 1980).

Systematic random sampling is generally regarded as the simplest of the geographic sampling methods in terms of administration. Initially, only the first sample point is randomly chosen. All other sample points are located at fixed intervals from this initial point in both the x and y directions. The interval is determined as a function of sample size. Although systematic sampling does ensure an even distribution of sample points among strata, biased and unrepresentative samples can result should any spatial periodicities within the data become coincident with the systematic pattern. Because only the first sample point is randomly selected, all points are not afforded an equal probability of inclusion in the sample (Berry and Baker, 1968).

Stratified systematic unaligned sampling has generally been recommended by some researchers as the preferred sampling procedure for assessing thematic map accuracy (Berry and Baker, 1968; Fitzpatrick-Lins, 1981; Rosenfield and Melley, 1980). In essence, the method involves establishing a grid of a suitable size over the map to be evaluated. In the uppermost left cell, establish a random x and y coordinate. In each cell along the first row, fix a point by combining the random x coordinate value of the first cell with a new random y coordinate value. Similarly, in each cell along the first column, fix a point by combining the random y coordinate value of the first cell with a new random x coordinate value. For all the succeeding rows, a point in each cell is fixed by combining the x coordinate value from the leftmost cell with the y coordinate value from the topmost cell (Clark and Hosking, 1986). The unaligned aspect of this method circumvents any bias resulting from spatial periodicities occurring within the data set, unlike a simple systematic approach. At the same time, uniform coverage of the study area is achieved by areal stratification of the study area. Berry and Baker (1968) contend that the hybrid nature of the method makes it suitable under most circumstances, especially where the spatial autocorrelation function of the mapped phenomena is unknown.

Finally, random cluster sampling uses a cluster of points as the sampling unit. The study area is divided into a number of clusters, out of which a random subset of clusters is selected. A high degree of heterogeneity within a cluster is desirable for maximum information content. Congalton (1988a) recommended that no more than ten pixels be used for a cluster, with 25 pixels as the maximum. Cluster sampling reduces the number of possible samples upon which a sampling distribution is based, but it also gives rise to greater sampling error than simple random sampling of the same size based on points. Cluster sampling is used because it facilitates field checking, thus saving resources and time (Todd *et al.*, 1980).

Vegetation Mapping of the Study Area

Field Mapping of Vegetation Communities

The entire Okefenokee Swamp covers about 167,000 ha in Ware, Charlton, and Clinch counties in Georgia, and Baker County in Florida (Hamilton, 1982). It is a deep fresh water swamp where a woody community occurs. Cohen (1973) described the Okefenokee Swamp as a "swamp-marsh complex" because approximately 25 percent of the area is void of woody vegetation. Such areas are known locally as "prairies."

The vegetation communities in the swamp were mapped by McCaffrey and Hamilton in 1980 at a scale of 1:63,360, based on an interpretation of 1:30,000-scale color infrared aerial photographs obtained in November, 1977, and on field observations. The mapping units employed by McCaffrey and Hamilton (1984) corresponded to classes, subclasses, or dominance vegetation types, using a minimum mapping unit of 10 ha. Where vegetation patches were smaller than 10 ha, several patches were grouped into a larger map unit comprising two or more classes, subclasses, or dominance types. It is important to emphasize that vegetation pattern in the swamp exhibits a "patchwork quilt" characteristic because of the various stages of succession of different vegetation communities juxtaposed together.

Because of the difficulty of collecting vegetation data in the Swamp (which can be accessed only by boats following the boat/canoe trails), two small study areas within the swamp were selected for the production of vegetation maps by digital image classification from Landsat TM data acquired on 03 October 1987 (Figure 1). These two study areas were carefully selected to represent the full diversity and spatial complexity of the vegetation communities in the Swamp. Study Area #1, to the northwest of Chase Prairie, is 2.28 km wide by 6.84 km long, or 15.60 km^2 in area. The Suwannee Canal, which is open to boat traffic, stretches along the southern edge of this study area (Figure 1). Study Area #2, to the east of Chase Prairie, is 2.28 km wide by 8.85 km long, or 20.18 km^2 in area. Certain sections of Study Area #2 are accessible by small roads leading down to the edge of the swamp from Swamp Perimeter Road, which runs along Trail Ridge. Approximately 17 percent of Study Area #2 is an upland ridge bordering the swamp, and therefore not considered part of the swamp proper (Figure 1).

The small study areas also made possible highly intense exhaustive survey of vegetation communities in the field (carried out in the 1988-89 period) to produce reference maps in UTM coordinates with the aid of color infrared aerial photography and 1:24,000-scale USGS topographic maps. The McCaffrey and Hamilton vegetation classification scheme was followed with only slight modifications, producing a total of 21 classes (Table 1). The minimum mapping unit was determined with reference to the spatial resolution of TM images, using the general rule of 4 times the ground area of a single pixel of the image (Star and Estes, 1990). Because a

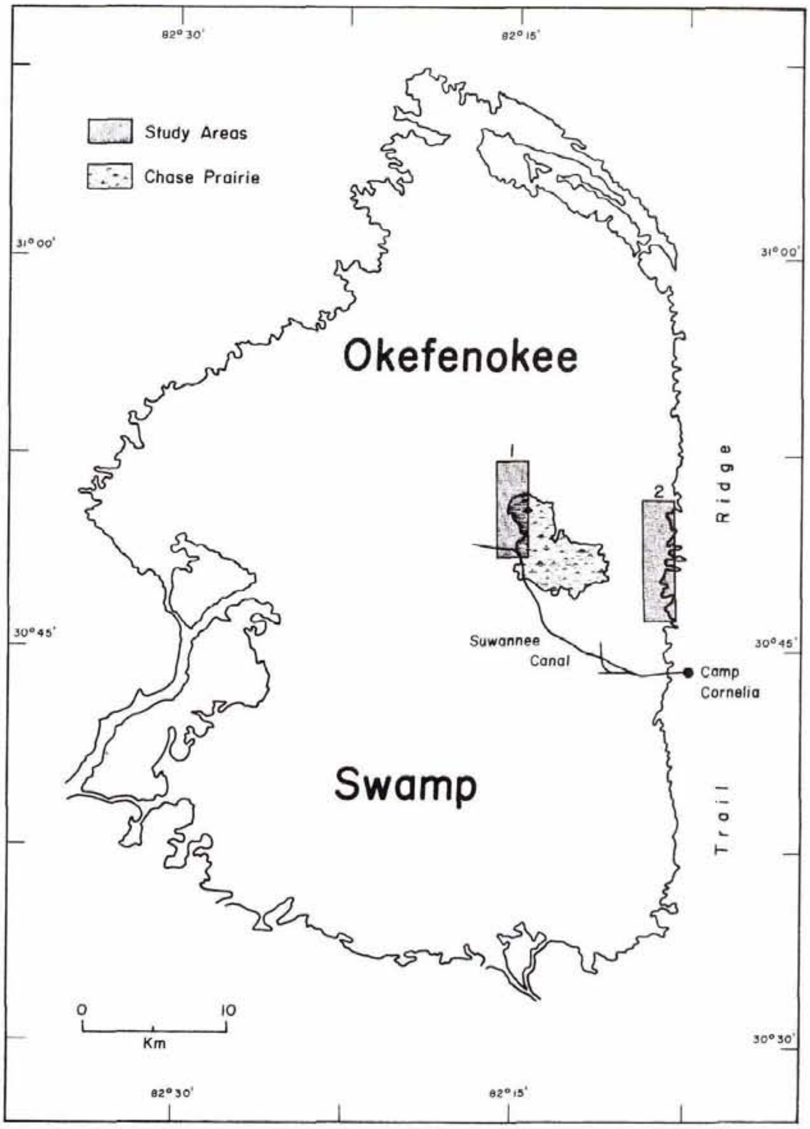


Figure 1. The location of the two study areas in the Okefenokee Swamp.

single TM pixel is 812.25 m² (28.5 m by 28.5 m), the minimum mapping unit used for the reference map was 3,249 m² or about 0.33 ha. The resultant maps for the two study areas exhibit a highly complex "patchwork quilt" pattern (Figure 2). The reference maps of the two study areas were digitized and converted to raster format with a spatial resolution corresponding to that of the TM images, i.e., 28.5 metres. They are the ground truth maps in digital form with which the computer-classified maps will be compared.

Computer-Assisted Mapping of Vegetation Communities from Landsat TM Data

There are 80 columns by 240 rows for Study Area #1 and 80 columns by 300 rows for Study Area #2 as extracted from the Landsat TM data. These data have also been rectified to the same UTM coordinate system of the reference maps. Based on 11 ground control points, the final root-mean-square error (RMSE) of planimetry was 0.384 pixel or 10.95 metres ground distance, thus ensuring accurate registration between the reference maps and the TM images. A supervised approach using a Baysian maximum-likelihood classifier was employed to classify the vegetation types in the two study areas. The full detail of the classification was reported elsewhere (Lo and Watson, 1994). Because the spatial resolution of the TM data is still not high enough to map the small patchy communities of vegetation in the Swamp, the accuracy of the 21-class vegetation map produced from the computer-assisted processing of Landsat TM data is very low (with an overall accuracy of 43 percent, and a Kappa coefficient of 0.3881). The 21-class vegetation classification scheme was aggregated, by means of species association, into six distinct broad vegetation groups (Table 2). Only the black gum swamp (N) and floating aquatic macrophyte prairie (Paq) remain as stand-alone vegetation classes. The remaining four classes are mixed vegetation swamp patches, discriminated on the basis of broad vegetation types found in them: broad-leaved evergreen and shrubs, cypress, grass, and mixed pines. As a result of the aggregation, the classification scheme is more compatible to the spatial resolution of the TM data. The map pattern becomes more simplified. The overall accuracy and the Kappa coefficient of the map increased to 63.6

TABLE 1. THE 21-CLASS VEGETATION SCHEME FOR USE IN MAPPING VEGETATION IN THE OKEFENOKEE SWAMP

B	— broad-leaved evergreen swamp
BC	— broad-leaved evergreen swamp with scattered cypress
C	— cypress swamp (*Taxodium ascendens*) [cypress dominated swamp]
Cm	— mixed cypress swamp [cypress canopy with some blackgum, pine, and/or broad-leaved evergreens]
CS	— mixed cypress and shrub swamp
CSP	— mixed cypress and shrub swamp with prairie
N	— black gum swamp (*Nyssa sylvatica* var. *biflora*) [blackgum dominated canopy, other hardwood and broad-leaved evergreen species present]
Paq	— floating aquatic macrophyte prairie
Ph	— herbaceous prairie
S	— shrub swamp
SP	— mixed shrub swamp and prairie
SB	— mixed shrub and broad-leaved evergreen swamp
ST	— mixed shrub swamp and pine forest
T1	— upland pine forest (*Pinus elliottii*)
T2	— pine swamp (*Pinus elliottii*)
Tm	— mixed pine swamp [pine canopy with blackgum, cypress, or shrub]
Y	— young cypress or blackgum swamp [canopy appears more even-aged and shorter (younger) than categories C or N]
YS	— mixed young and shrub swamp
YP	— mixed young swamp and prairie
YT	— mixed young and pine swamp
T3	— young upland pine

percent and 0.5345, respectively. The six class vegetation maps of the two study areas provide the basis for the evaluation of the sampling methods in this paper (Figure 3).

Comparison of Sampling Methods

Selection of Samples

The accuracy of the classified maps was investigated using the five sampling methods mentioned before. The minimum sampling size was determined using the following formula from Snedecor and Cochran (1967):

$$n = Z^2 \cdot p\,(1 - p)/E^2$$

where n is the minimum sample size; p is the expected population proportion (e.g., normally, 0.5); Z is the standard normal deviate, normally set at 1.65 to correspond to a 90 percent two-sided confidence level; and E is the tolerable error, set at 0.04 (implying that the sample estimate should not deviate by more than 4 percentage points). The sampling unit used is the pixel of the computer classified vegetation map, which has an area of 812.25 m^2. Given these terms, a minimum of 426 pixels was calculated, or slightly over 1 percent of the total number of pixels (39,191) of the two study areas combined. Although it is realized that this formula is based on the binomial distribution (Congalton, 1991), it produces a sufficient number of samples for a six-class vegetation map.

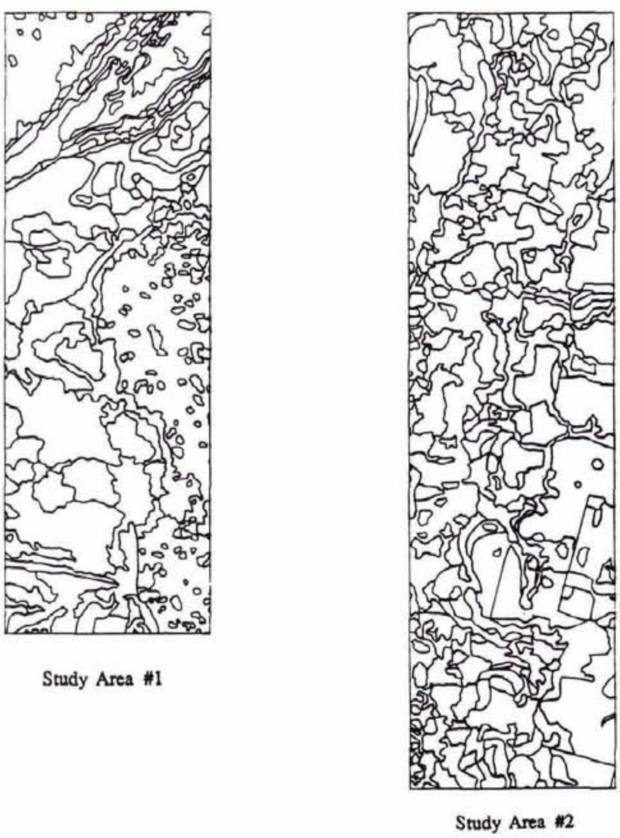


Figure 2. The "patchwork quilt" pattern of polygons of the 21-class reference vegetation maps of the two study areas.

TABLE 2. THE SIX-CLASS VEGETATION SCHEME FORMED BY THE AGGREGATION OF THE 21 ORIGINAL WETLAND VEGETATION CLASSES FROM TABLE 1

B, BC, S, SB, ST, Tm	BS -- mixed broad-leaved evergreen and shrub swamp (either dominant)
N	N -- black gum swamp
C, Cm, CS, Y, YT, YS	Cd -- mixed cypress swamp (cypress dominant)
CSP, Ph, SP, YP	Pm -- mixed herbaceous prairie (no dominance)
Paq	Paq -- floating aquatic macrophyte prairie
T1, T2, T3	Td -- mixed pine forest (pine dominant)

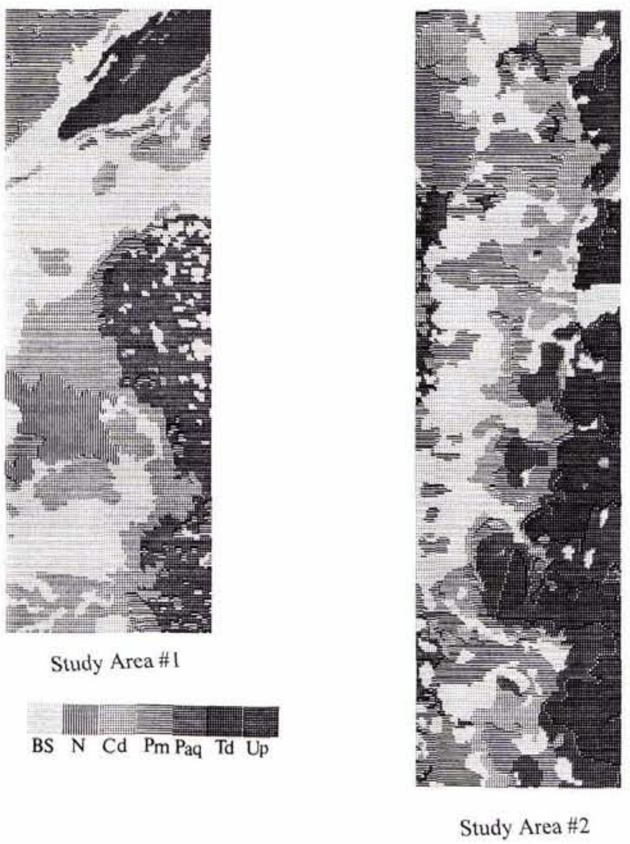


Figure 3. The six-class reference vegetation maps of the two study areas used in this research. For an explanation of the Key, see Table 2. Up = Upland Ridge.

Another consideration of this study is to keep the number of samples as small as possible in view of the difficulty and cost of collecting ground truth information in a swampy environment in the real-life situation. As pixels are used as sampling units, a total of 426 pixels distributed over six classes is considered to be adequate.

The same sample size was maintained for all the sampling methods as far as possible. For simple random sampling, 432 point coordinates were randomly selected. For stratified random sampling, the study area was partitioned geometrically, not categorically, so that sample points could be more evenly distributed. The study area was divided into six equal strata, each containing 80 columns and 90 rows. Seventy-two sample points were selected per stratum for a total of 432 sample points. For random systematic sampling and stratified systematic unaligned sampling, a grid of ten pixels each was used, thus giving rise to 432 grids, each of which contained a sample point. In the case of random systematic sampling, only the upper-left most sample point was selected randomly. All remaining sample points were spaced at ten-pixel intervals in both the x and y directions. For cluster sampling, in accordance with the recommendation by Congalton (1988), clusters consisting of ten pixels, each measuring two by five pixels, were selected. This resulted in 44 clusters. Cluster locations were determined by randomly selecting the pixel coordinates of the northwest pixel of each cluster. Every cluster had to fall entirely within the study area, and all clusters had to be non-overlapping. After adjustment for the 17 percent of upland ridge area excluded from classification in Study Area #2, the effective sample size for each sampling method is still maintained at 1 percent of the total population of pixels (Table 3).

The classified vegetation maps and the reference vegetation maps of the two study areas were overlaid pixel by pixel to produce an error matrix by cross-tabulation (Table 4). For each of the five sampling methods, the sample points on the classified maps were overlaid on the reference maps to produce an error matrix (Tables 5 through 9).

Kappa Coefficients of Agreement and Goodness-of-Fit Chi-Square Tests

To compare each sampling method's result in a thematic accuracy assessment, two approaches were adopted. The first approach involved computation of the Kappa coefficient of agreement and its variance, a discrete multivariate measure developed by Cohen (1960) and widely used by remote sensing scientists (e.g., Congalton and Mead, 1983; Rosenfield and Fitzpatrick-Lins, 1986). Kappa was calculated based on the formulae provided by Hudson and Ramm (1987), using a Microsoft FORTRAN program written by one of the authors for personal computers. Unlike the overall accuracy statistic, which is based solely on the main diagonal of the error matrix, the Kappa coefficient indirectly incorporates the off-diagonal elements as a product of the row and column marginals and thus provides a more realistic measure of the characteristics of the error matrix. The Kappa coefficient computed from the error matrix of each sampling method was compared in turn with the Kappa coefficient of the error matrix for the whole computer-classified vegetation map as determined by the reference map. The following tests of significance between two independent Kappa coefficients as developed by Cohen (1960) were used:

$$Z = (\kappa_1 - \kappa_2)/\sqrt{[V(\kappa_1) + V(\kappa_2)]}$$

where Z is the standard normal deviate, κ is the Kappa coefficient, and $V(\kappa)$ is the variance of Kappa coefficient. If Z exceeds 1.96, then the difference is significant at the 95 percent confidence level.

The second approach made use of the Chi-square test to accept or reject the null hypothesis of no difference between two error matrices. As observed by Cohen (1960), the Chi-square test tests the null hypothesis with regard to association, and can be applied here because it meets all the statistical conditions required for the use of the test. The Chi-square test is a method for comparing nominal data, in which individual observations are assigned to categories (i.e, differentiated on a nominal scale). Requirements for the use of the Chi-square test include (1) the data must be in the form of frequencies counted; (2) the total numbers observed must exceed 20; (3) the expected frequency in any one fraction must not normally be less than 5, but if more than two categories are involved (i.e., with the degree of freedom greater than 2), no more than 20 percent of the expected frequencies may be less than 5, and in no case must any expected frequency be

TABLE 3. EFFECTIVE NUMBER OF SAMPLE POINTS AND PERCENTAGE OF CLASSIFIED AREA SAMPLED PER SAMPLING METHOD

Sampling Method	Total Number of Points	Effective Points	Percentage of Classified Area Sampled
Simple Random Sample	432	394	1.01
Stratified Random Sample	432	383	0.98
Random Systematic Sample	432	382	0.98
Stratified Systematic Unaligned Sample	432	389	0.99
Random Cluster Sample	440	427	1.09
Average	434	395	1.01

Table 4. Error Matrix of the Overall Computer-Classified Vegetation Maps for the Two Study Areas (Combined) as Determined by the Reference Maps (Compiled from Ground Surveys)

Reference Data

Classified Data: Class	BS	N	Cd	Pm	Paq	Td	Row Total	% Error of Commission
BS	8625	42	1666	1789	91	459	12,672	32
N	1059	821	255	2	0	1	2,138	64
Cd	1708	194	5290	889	72	497	8,650	39
Pm	924	5	891	3632	505	317	6,274	42
Paq	425	1	257	473	3658	166	4,980	26
Td	583	0	495	411	97	2891	4,477	35
Column Total	13,324	1,063	8,854	7,196	4,423	4,331	39,191	
% Error of Omission	35	23	40	49	17	33		

Main Diagonal = 24,917

Overall Accuracy = 24,917 / 39,191 = 63.6%   KAPPA = 0.5344705

Table 5. Error Matrix Derived from Simple Random Sampling

Reference Data

Classified Data: Class	BS	N	Cd	Pm	Paq	Td	Row Total	% Error of Commission
BS	105	0	18	18	1	4	146	28
N	13	4	4	0	0	0	21	81
Cd	18	0	38	7	0	9	72	47
Pm	14	0	7	33	3	5	62	47
Paq	4	0	6	2	37	2	51	27
Td	2	0	8	0	0	32	42	24
Column Total	156	4	81	60	41	52	394	
% Error of Omission	33	0	53	45	10	38		

Main Diagonal = 249

Overall Accuracy = 249 / 394 = 63.2%   KAPPA = 0.5180896

TABLE 6. ERROR MATRIX DERIVED FROM STRATIFIED RANDOM SAMPLING

Classified Data \ Reference Data	BS	N	Cd	Pm	Paq	Td	Row Total	% Error of Commission
BS	78	1	18	20	1	5	123	37
N	12	13	3	0	0	0	28	54
Cd	18	1	55	5	0	4	83	34
Pm	8	0	7	30	2	3	50	40
Paq	2	0	2	5	36	3	48	25
Td	7	0	7	3	2	32	51	37
Column Total	125	15	92	63	41	47	383	
% Error of Omission	38	13	40	52	12	32		

Main Diagonal = 244

Overall Accuracy = 244 / 383 = 63.7% KAPPA = 0.5400413

TABLE 7. ERROR MATRIX DERIVED FROM RANDOM SYSTEMATIC SAMPLING

Classified Data \ Reference Data	BS	N	Cd	Pm	Paq	Td	Row Total	% Error of Commission
BS	80	0	18	19	0	3	120	33
N	7	8	1	0	0	0	16	50
Cd	18	3	53	12	1	8	95	44
Pm	9	0	14	30	5	3	61	51
Paq	4	0	2	8	38	2	54	30
Td	7	0	4	3	0	22	36	39
Column Total	125	11	92	72	44	38	382	
% Error of Omission	36	27	42	58	14	42		

Main Diagonal = 231

Overall Accuracy = 231 / 382 = 60.5% KAPPA=0.4934488

TABLE 8. ERROR MATRIX DERIVED FROM STRATIFIED SYSTEMATIC UNALIGNED SAMPLING

Reference Data

Classified Data	Class	BS	N	Cd	Pm	Paq	Td	Row Total	% Error of Commission
	BS	88	1	13	21	0	8	131	33
	N	12	8	1	0	0	0	21	62
	Cd	15	2	58	5	1	5	86	33
	Pm	7	0	10	39	4	4	64	39
	Paq	6	0	3	1	34	2	46	26
	Td	4	0	8	1	1	27	41	34
	Column Total	132	11	93	67	40	46	389	
	% Error of Omission	33	27	38	42	15	41		

Main Diagonal = 254

Overall Accuracy = 254 / 389 = 65.3%

KAPPA = 0.554149

TABLE 9. ERROR MATRIX DERIVED FROM RANDOM CLUSTER

Reference Data

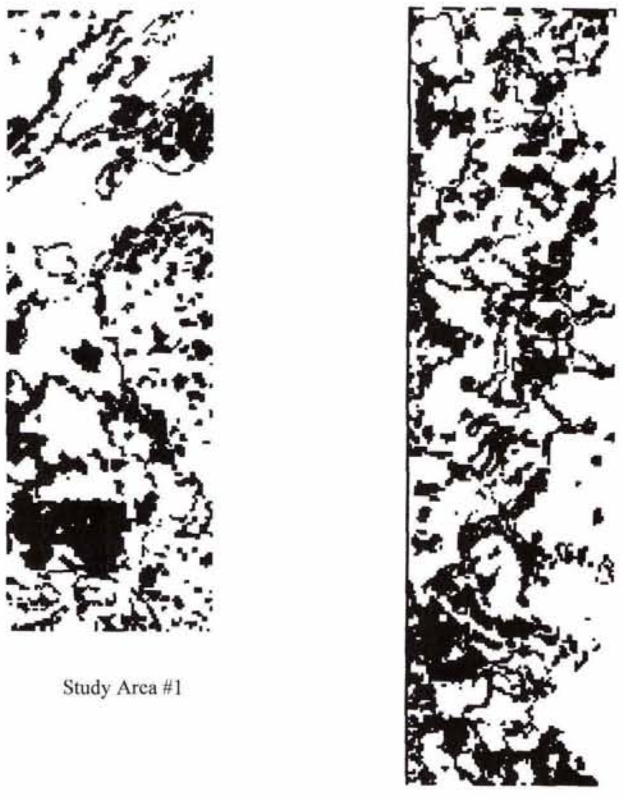
Classified Data	Class	BS	N	Cd	Pm	Paq	Td	Row Total	% Error of Commission
	BS	98	0	10	44	1	1	154	36
	N	0	10	0	0	0	0	10	0
	Cd	17	2	35	17	3	3	77	54
	Pm	3	0	7	58	0	7	75	23
	Paq	2	0	6	6	45	0	59	24
	Td	16	0	5	7	0	24	52	54
	Column Total	136	12	63	132	49	35	427	
	% Error of Omission	28	17	44	56	8	33		

Main Dia onal = 270

Overall Accuracy = 270/427 = 63.2%

KAPPA = 0.5272385

* Unsampled Class



Study Area #1

Study Area #2

Figure 4. Difference images for Study Areas #1 and #2 (black = error; white = correct).

less than 1; and (4) the observations must be independent (Hammond and McCullagh, 1974). The following formula was applied:

$$\chi^2 = \sum (O - E)^2/E$$

where O is the observed frequency defined as the number of sample points correctly classified per vegetation category (i.e., the main diagonal in the error matrix), and E is the expected frequency calculated by multiplying the total number of sample points falling within a category by the actual percentage of that category, as determined by the reference map. The actual categorical percentage used was producer's accuracy extracted from the error matrix for the whole computer-classified map as determined by the reference map. The calculated Chi-square value of each sampling method was then compared to the Chi-square value tabulated at the 95 percent confidence level for $(n - 1)$ degrees of freedom.

Compactness Ratio, Difference Image, and Spatial Autocorrelation Analysis

In order to determine the spatial complexity of the vegetation map patterns, a shape index known as Compactness Ratio which relates the area of an polygon with its perimeter was used (Haggett *et al.*, 1977). As implemented in the IDRISI program employed, the Compactness Ratio (CR) is computed as

$$\mathrm{CR} = \sqrt{(A_p/A_c)}$$

where A_p is the area of the polygon being calculated, and A_c is the area of a circle having the same perimeter as that of the polygon being calculated. The ratio varies from 0 to 1, and the closer the ratio is to 1, the more the polygon is like a circle and hence more compact.

A difference image between the reference maps and the computer-classified maps of the two study areas was also generated to help visualize the spatial pattern of errors of classification, following the approach used by Congalton (1988b). Because the two sets of maps have been accurately georeferenced, the method of overlay with subtraction in the IDRISI program was employed to produce the difference map, which was then re-coded into a binary error map with 1 (in black) to represent errors and 0 (in white) to show no errors (Figure 4).

To analyze quantitatively the spatial patterns of errors revealed for the two study areas, a method of spatial autocorrelation was applied. This sheds light on the degree to which values in a pixel are similar to those of the pixels immediately surrounding it. The measure used was Moran's I as implemented in the IDRISI program (Cliff and Ord, 1973; Griffith and Amrhein, 1991). The range of Moran's I varies from -1, when adjacent pixels are very different, to $+1$, when they are very much alike. The King's Case of spatial autocorrelation, which examines the pixels diagonally connected to each pixel as well as those to the left, right, above, and below each pixel, was selected. Spatial autocorrelation was computed for ten lags. Because of the small size of the two study areas, ten lags provide a long enough distance to evaluate the effects of spatial autocorrelation. To compute spatial autocorrelation of the second and higher lags, the difference images of the two study areas had to be first contracted by the method of pixel thinning using the desired lag number as the contraction factor in both the rows and columns. This resulted in a generalized difference image with reduced numbers of rows and columns (i.e., a lower spatial resolution).

Results of Analysis

Based on a comparison of the Kappa coefficients and the corresponding Z-scores as shown in Table 10, the error matrices of all the sampling methods are not significantly different from the error matrix of the total computer-classified vegetation map, because their Z-scores are all smaller than 1.96. However, the different magnitudes of the Z-scores indicate that the sampling method which produced a Kappa coefficient most similar to that of the total computer-classified vegetation map is stratified random sampling (with a Z-score of only 0.17) followed by stratified systematic unaligned sampling (with a Z-score of 0.60).

A similar but more specific conclusion is given by the Chi-square goodness-of-fit test in Table 11. At the 95 percent level of confidence, the null hypothesis of no difference was accepted only for the stratified random sampling method. The null hypothesis of no difference, however, was rejected for the stratified systematic unaligned sampling method at the 95 percent level of confidence, but accepted at the 90 percent level of confidence. A hypothesis to test the proportional allocation of sample points by vegetation classes for each sampling method revealed that it was accepted only for stratified systematic unaligned sampling method at the 95 percent level of confidence in a Chi-square test (Table 12). In other words, stratified systematic unaligned sampling is the only method capable of generating sample points distribution more representative of the actual distribution of vegetation by class in the swamp.

The average Compactness Ratio for the six-class reference map of Study Area #1 was 0.2870 while that for Study Area #2 was 0.3701. These values suggest that the mapped vegetation patterns are not very compact, and Study Area #1, with many perforations of BS (mixed broad-leaved evergreen and shrub) within a sea of Paq (floating aquatic macrophyte prairie), is less compact than Study Area #2. Taken as a

TABLE 10. KAPPA COEFFICIENTS, KAPPA VARIANCES, AND Z-SCORES BY SAMPLING METHODS

	WHOLE MAP	SIMRAND	STRARAN	RANSYS	STRSYSU	RANCLUS
KAPPA	0.5344705	0.5180896	0.5400413	0.4934488	0.554149	0.5272385
V(KAPPA)	1.083E-05	0.00115719	0.00109255	0.00116974	0.00107489	0.00103315
Z-SCORE		-0.479305728	0.167708476	-1.19389953	0.597218252	-0.223826925

KEY:
SIMRAND = SIMPLE RANDOM
STRARAN = STRATIFIED RANDOM
RANSYS = RANDOM SYSTEMATIC
RANCLUS = RANDOM CLUSTER
STRSYSU = STRATIFIED SYSTEMATIC UNALIGNED

whole, the vegetation pattern of the six-class map is more fragmented than compact.

Visually, the difference images of the two study areas reveal elements of patchiness and linearity of the agricultural and forested environments identified in Congalton's (1988b) study (Figures 4 and 5). The patchy pattern was regarded to be spatially simple and the linear pattern complex by Congalton (1988b). The swampy environment therefore shows a mixture of both simple and complex spatial elements. The linear elements are the result of boundary errors between two vegetation types. The large patchy elements represent gross errors stemming from spectral class confusion or omission of a spectrally distinct class in the classification scheme. The smaller patchy elements reveal errors related to unsuitable sensor spatial resolution or an incompatible minimum mapping unit, as observed in the Chase Prairie region of Study Area #1. The Moran's Coefficients computed for the ten lags for the two study areas revealed strong positive spatial autocorrelation for the first lag, followed by a rapid drop in the second lag and then gradual and very gradual drop in the third and subsequent lags (Figure 5). The implication is that the errors tend to be clustered, but, as the spatial resolution of the difference image decreases, even by a small factor, the clustering effect im-

TABLE 11. CHI-SQUARE TEST RESULTS PER SAMPLING METHOD, COMPARING OBSERVED AND EXPECTED CATEGORICAL ACCURACY FREQUENCIES

Class	Simple Random		Stratified Random		Random Systematic		Strat Sys Unaligned		Random Cluster	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
BS	105	100.9	78	80.9	80	80.9	88	85.4	98	88.0
N	4	3.1	13	11.6	8	80.5	8	8.5	10	9.3
Cd	38	48.4	55	55.0	53	55.0	58	55.6	35	37.7
Pm	33	30.3	30	31.8	30	36.4	39	33.8	58	66.7
Paq	37	33.9	36	33.9	38	36.4	34	33.1	45	40.5
Td	32	34.7	32	31.4	22	25.4	27	30.7	24	23.4
Calculated Chi-Square Statistic	3.41		0.52		1.27		1.26		2.91	
Tabulated Chi-Square Statistic	1.15		1.15		1.15		1.15		1.15	
Decision	Not Similar		Similar		Not Similar		Not Similar		Not Simil	

* At 95 % level of confidence

TABLE 12. CHI-SQUARE TEST RESULTS PER SAMPLING METHOD, COMPARING OBSERVED AND EXPECTED SAMPLE POINT FREQUENCIES PER CLASS

Class	Simple Random		Stratified Random		Random Systematic		Strat Sys Unaligned		Random Cluster	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
BS	156	133.1	125	133	125	133.1	132	133.1	136	133.1
N	4	10.6	15	10.6	11	10.6	11	10.6	12	10.6
Cd	81	88.5	92	88.5	92	88.5	93	88.5	63	88.5
Pm	60	71.9	63	71.9	72	71.9	67	71.9	132	71.9
Paq	41	44.2	41	44.2	44	44.2	40	44.2	49	44.2
Td	52	43.4	47	43.4	38	43.4	46	43.4	35	43.4
Calculated Chi-Square Statistic	12.59		4.09		1.32		1.14		57.42	
Tabulated Chi-Square Statistic *	1.15		1.15		1.15		1.15		1.15	
Decision	Not Similar		Not Similar		Not Similar		Similar		Not Similar	

* At 95% Level of Confidence.

mediately becomes lessened. Study Area #1, which is less compact in its mapped vegetation pattern, has higher positive spatial autocorrelation than Study Area #2 for all ten lags. Finally, there are some slight fluctuations noticed in the graph beginning from lag 5 to lag 9 for both study areas — an indication of weak periodicities in the two difference images (Figure 5).

The finding that stratification in the sampling design was found to be best applied over the Okefenokee Swamp is related to the fragmented nature of its vegetation pattern. It allows a more even spread of sample points to be allocated to the small vegetation patches scattered all over the swamp.

Conclusions

The findings from this research suggest that any sampling method will be capable of generating an accurate overall thematic accuracy value. However, if categorical accuracy is required, the Kappa coefficient of agreement should be computed, which will reveal differences in thematic accuracy evaluation by different sampling methods. This paper also recommends the use of the Chi-square test, a much simpler hypothesis testing technique, to complement the Kappa coefficient of agreement in order to also show the degree of association between two error matrices. In this study, the findings from the Chi-square tests are very similar to those using tests of significance between two independent Kappa coefficients of agreement.

Based on the six-class vegetation maps of the two study areas in the Okefenokee Swamp, we find that the stratified random sampling method is probably the best to use in thematic accuracy evaluation. Another choice is the stratified systematic unaligned sampling method, which is somewhat more complicated to use than the stratified random sampling method in the selection of sample points. The stratified systematic unaligned sampling method has one advantage over the stratified random sampling method in that it can generate a pattern of sample point distribution by vegetation class more representative of the actual distribution of vegetation by class in the Swamp.

The spatial patterns of classification errors of the two study areas in the Okefenokee Swamp were investigated by using spatial autocorrelation analysis. A mixture of simple patchy and complex linear patterns of errors was evident. The linear pattern was the result of boundary placement errors and transitional zone errors. The patchy pattern was associated with the low sensor spatial resolution not capable of resolving small islands of vegetation in the swamp and with spectral confusion between classes. The errors for both study areas were found to be clustered in the first lag, but, as the spatial resolution decreased, positive spatial autocorrelation declined. The spatially more complex vegetation pattern of Study Area #1 produced higher positive spatial autocorrelation. The implication is that high spatial resolution data and spatially complex mapped patterns will produce strong spatial autocorrelation, or, highly clustered error patterns.

In view of the physical difficulty of carrying out ground checking of vegetation in a swampy environment, it is doubly important to select an appropriate sampling method to evaluate the thematic accuracy of the vegetation map produced from high-resolution Thematic Mapper data. The mixed spatial complexity of the mapped pattern and the strong positive spatial autocorrelation associated with high spatial resolution of the image data are the two main problems to deal with. Our research suggests that stratified random sampling is the best sampling method to use in such a situation, and, in view of the weak periodicity of the vegetation data in the swampy environment, stratified systematic unaligned sampling can also be applied. The use of stratification in the sampling scheme ensures an even allocation of sample points to the small, juxtaposing vegetation patches in the swamp.

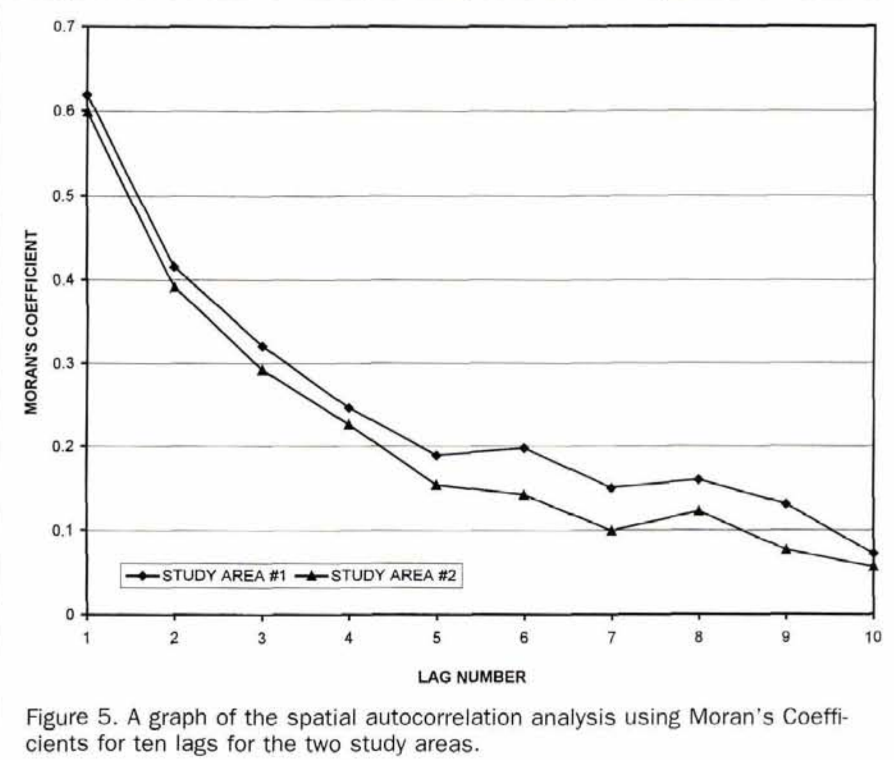


Figure 5. A graph of the spatial autocorrelation analysis using Moran's Coefficients for ten lags for the two study areas.

Acknowledgments

The authors wish to acknowledge the help provided by Professor Bernard C. Patten of the Institute of Ecology in the University of Georgia, which made possible our field work in the Okefenokee Swamp. The authors are also grateful to the constructive criticisms given by three reviewers of this paper. In particular, the comments given by Professor Russ Congalton of the University of New Hampshire have provided excellent ideas for the authors to rewrite the paper in the present form. Of course, any shortcomings that remain in the revised paper are the responsibility of the authors.

References

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

Card, D.H., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy, *Photogrammetric Engineering & Remote Sensing*, 48:431–439.

Clark, W.A., and P.L. Hosking, 1986. *Statistical Methods for Geographers*, John Wiley and Sons, New York, 518 p.

Cliff, A.D., and J.K. Ord, 1973. *Spatial Autocorrelation*, Pion, London, 178 p.

Cohen, A.D., 1973. Possible influences of subpeat topography and sediment type upon the development of the Okefenokee Swamp-Marsh complex of Georgia, *Southeastern Geology*, 15:141–151.

Cohen, J., 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20:37–46.

Congalton, R.G., 1988a. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54:593–600.

———, 1988b. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54:587–592.

———, 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R.G., and R.A. Mead, 1983. A quantitative method to test for consistency and correctness in photointerpretation, *Photogrammetric Engineering & Remote Sensing*, 49:69–74.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map, *Photogrammetric Engineering & Remote Sensing*, 47:343–351.

Griffith, D.A., and C.G. Amrhein, 1991. *Statistical Analysis for Geographers*, Prentice Hall, Englewood Cliffs, New Jersey, 478 p.

Haggett, P., A.D. Cliff, and A. Frey, 1977. *Locational Methods*, Edward Arnold, London, 605 p.

Hamilton, D.B., 1982. *Plant Succession and the Influence of Disturbance in the Okefenokee Swamp, Georgia*, Ph.D. Dissertation, University of Georgia (reproduced as *Okefenokee Ecosystem Investigation*, Technical Report Number 9, Institute of Ecology, Athens, Georgia).

Hammond, R., and P.S. McCullagh, 1974. *Quantitative Techniques in Geography: An Introduction*, Clarendon Press, Oxford, 318 p.

Hay, A.M., 1979. Sampling designs to test land-use map accuracy, *Photogrammetric Engineering & Remote Sensing*, 45:529–533.

Hudson, W.D., and C.W. Ramm, 1987. Correct formulation of the Kappa coefficient of agreement, *Photogrammetric Engineering & Remote Sensing*, 53:421–422.

Lo, C.P., and L.J. Watson, 1994. Okefenokee swamp vegetation mapping with Landsat Thematic Mapper data: an evaluation, *ASPRS Technical Papers, ASPRS/ACSM Annual Convention and Exposition*, Reno, Nevada, 25–28 April 1994, ACSM and ASPRS, Bethesda, Maryland, 1:365–374.

McCaffrey, C.A., and D.B. Hamilton, 1984. Vegetation mapping of the Okefenokee Swamp ecosystem, *The Okefenokee Swamp: Its Natural History, Geology and Geochemistry* (A.D. Cohen, D.J. Casagrande, M.J. Andrejko, and G.R. Best, editors), Wetlands Survey, Los Alamos, New Mexico, pp. 201–211.

Quirk, B.K., and F.L. Scarpace, 1980. A method of assessing accuracy of a digital classification, *Photogrammetric Engineering & Remote Sensing*, 46:1427–1431.

Rosenfield, G.H., and K. Fitzpatrick–Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52:223–227.

Rosenfield, G.H., and M.L. Melley, 1980. Applications of statistics to thematic mapping, *Photogrammetric Engineering & Remote Sensing*, 46:1287–1294.

Rudd, R.D., 1971. Macro land-use mapping with simulated space photographs, *Photogrammetric Engineering*, 37:365–372.

Snedecor, G.W., and W.F. Cochran, 1967. *Statistical Methods*, State University Press, Ames, Iowa, 517 p.

Star, J., and J. Estes, 1990. *Geographic Information Systems: An Introduction*, Prentice Hall, Englewood Cliffs, New Jersey, 295 p.

Stehman, S.V., 1992. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 58:1343–1350.

Story, M., and R.G. Congalton, 1986. Accuracy assessment: a user's perspective, *Photogrammetric Engineering & Remote Sensing*, 52:397–399.

Todd, W.J., D.G. Gehring, and J.F. Haman, 1980. Landsat wildland mapping accuracy, *Photogrammetric Engineering & Remote Sensing*, 46:509–520.

Van Genderen, J.L., and B.F. Lock, 1977. Testing land-use map accuracy, *Photogrammetric Engineering & Remote Sensing*, 51:1135–1137.

Zonneveld, I.S., 1972. *Lectures on Vegetation Science (Ecology) and Vegetation Survey*, I.T.C., Enschede.

Received 24 April 1995, revised 02 October 1995; revised 23 March 1998)