

Nationwide Forestry Applications Program  
Analysis of Forest Classification Accuracy

Virginia Polytechnic Inst. and State Univ.  
Blacksburg

Prepared for

Forest Service  
Houston, TX

Feb 81

U.S. Department of Commerce  
National Technical Information Service

**NTIS.**

<b>REPORT DOCUMENTATION PAGE</b>	<b>1. REPORT NO.</b> NFAP-239	<b>2.</b>	<b>3. Recipient's Accession No.</b> PB82 164500
<b>4. Title and Subtitle</b>  Analysis and Forest Classification Accuracy			<b>5. Report Date</b> January 1981
<b>7. Author(s)</b> R. G. Congalton, R. A. Mead, R. G. Oderwald, and J. Heinen			<b>6.</b> 049098
<b>9. Performing Organization Name and Address</b> School of Forestry and Wildlife Resources Virginia Polytechnic Institute & State University Blacksburg, VA 24061			<b>8. Performing Organization Rept. No.</b>
<b>12. Sponsoring Organization Name and Address</b> U.S. Department of Agriculture, Forest Service Nationwide Forestry Applications Program 1050 Bay Area Blvd. Houston, Texas 77058 Tech. Monitor: J.R. Bell			<b>10. Project/Task/Work Unit No.</b>
			<b>11. Contract(C) or Grant(G) No.</b> (C) 13-1134 (G)
			<b>13. Type of Report &amp; Period Covered</b> Annual Progress
<b>15. Supplementary Notes</b>  An AgRISTARS, Renewable Resources Inventory Project Report			<b>14.</b>
<b>16. Abstract (Limit: 200 words)</b> <p>Many studies have been conducted to determine the usefulness of Landsat data for mapping land cover. However, very little research has been done to determine the degree of success (i.e., accuracy) in doing this. The apparent absence of quantitative methods for comparing classification accuracy is certainly a stumbling block that must be overcome. The effects of imaging date, spectral band combination, classification algorithm, training set selection procedure, and the image analyst on final classification accuracy must be studied. Therefore, the following study was proposed with these objectives:</p> <ol style="list-style-type: none"> <li>1. To develop a computer system that implements an iterative proportional fitting technique to "normalize" the coefficients within classification error matrices.</li> <li>2. To develop hierarchical models for testing the significance of several factors (e.g., image date, classification algorithm, the analyst, etc.) on the resulting classification accuracy.</li> <li>3. To test the above techniques and determine their usefulness with actual data for classification accuracy.</li> </ol> <p>Results of the development and test of procedures for standardizing forest classification accuracy assessment are presented. Appendices I through VIII include specific error matrices, Fortran listings and various technical contributions related to forest classification accuracy.</p>			
<b>17. Document Analysis a. Descriptors</b> Forest Classification Landsat Accuracy Assessment Spatial Analysis System  <b>b. Identifiers/Open-Ended Terms</b>     <b>c. COSATI Field/Group</b> Forestry (Remote Sensing)			
<b>18. Availability Statement:</b>  Release Unlimited		<b>19. Security Class (This Report)</b> Unclassified	<b>21. No. of Pages</b> 85
		<b>20. Security Class (This Page)</b> Unclassified	<b>22. Price</b> 



# AgRISTARS

RR-111-04066  
JSC-17123

A Joint Program for  
Agriculture and  
Resources Inventory  
Surveys Through  
Aerospace  
Remote Sensing

## Renewable Resources Inventory

January 20, 1981



USDA Forest Service

## NATIONWIDE FORESTRY APPLICATIONS PROGRAM

Cooperative Research Report

Analysis of Forest Classification Accuracy

Cooperative Agreement No. 13-1134

Virginia Polytechnic Institute  
and State University  
Blacksburg, Virginia 24061

REPRODUCED BY  
NATIONAL TECHNICAL  
INFORMATION SERVICE  
U.S. DEPARTMENT OF COMMERCE  
SPRINGFIELD, VA 22161



**NASA**



Lyndon B. Johnson Space Center  
Houston, Texas 77058

Project Supported by  
Nationwide Forestry Applications Program  
Renewable Resources Inventory Project  
Houston, Texas

Cooperative Agreement Number 13-1134

February 1981

Remote Sensing Research Report 81-1  
Analysis of Forest Classification Accuracy

by

Russell G. Congalton<sup>2/</sup>

Roy A. Mead<sup>1/</sup>

Richard G. Oderwald<sup>1/</sup>

Joel Heinen<sup>2/</sup>

1/ Assistant Professors

2/ Graduate Research Assistants, School of Forestry and Wildlife  
Resources, Virginia Polytechnic Institute and State University,  
Blacksburg, VA 24061.

RECEIVED  
FEB 15 1981  
FORESTRY  
APPLICATIONS  
PROGRAM

## TABLE OF CONTENTS

List of Tables	
Preface	
Abstract	
1.0	Introduction
1.1	Objectives
1.2	Justification
1.3	State of the Art of Landsat Classification Accuracy Assessment
1.31	Accuracy Assessment Techniques
1.32	Sampling Techniques
1.33	National Data Base for Error Matrices
2.0	Statistical Methods
2.1	Methods of Comparing Similarity Matrices
2.2	Categorical Data Analysis
3.0	Sample Data Analyses
3.1	MARGFIT Analysis
3.2	KAPPA Analysis
3.3	CONTABLE Analysis
4.0	Accuracy Conference
5.0	Wildlife Habitat Assessment Methods
6.0	Effect of Classification Accuracy on Interspersion Maps
7.0	Summary and Future Work
8.0	Literature Cited



Appendix I	List of Sources of Error Matrices
Appendix II	Listing of FORTRAN Computer Program MARGFIT
Appendix III	Listing of FORTRAN Computer Program KAPPA
Appendix IV	Listing of APL Computer Program CONTABLE
Appendix V	A Quantitative Method to Test for Similarity between Photo Interpreters
Appendix VI	Landsat Classification Accuracy Assessment Procedures: An Account of a National Working Conference
Appendix VII	Listing of FORTRAN Computer Program WILDINT2
Appendix VIII	A Computerized Spatial Analysis System for Assessing Wildlife Habitat from Vegetation Maps

# LIST OF TABLES

<u>Table</u>	<u>Title</u>	<u>Page</u>
1	Josesigs error matrix before normalization .....	15
2	Josesigs error matrix after normalization .....	15
3	Overall and normalized overall performance results for five classification error matrices .....	16
4	Overall and normalized overall performance results for four cover type error matrices .....	17
5	$\hat{K}$ statistic with upper and lower limits at 95% confidence interval for four cover type error matrices .....	18
6	$\hat{K}$ statistic with upper and lower limits at 95% confidence interval for four classification algorithms .....	19
7	List of factors and effects for 5-way contingency table .....	20
8	List of hypothesis for CONTABLE example .....	21
9	List of results and conclusions for CONTABLE example .	21
10	Similarity matrix for five fictional cover types on the base map and on map II .....	25
11	Similarity matrix for five fictional cover type maps on the base map and on map III .....	25
12	Similarity matrix for three categories of inter- spersion, high (3), medium (2), and low (1) produced from the base map and map II .....	26



<u>Table</u>	<u>Title</u>	<u>Page</u>
13	Similarity matrix for three categories of interspersion, high (3), medium (2), and low (1) produced from the base map and map III .....	26

## PREFACE

The authors would like to thank the Department of Statistics at Virginia Polytechnic Institute and State University for their support with the computer programs used in this study. Special thanks go to Dr. S. K. Lee who wrote the CONTABLE program. Thanks also go to Daniel Yaussy for translating MARGFIT and KAPPA into FORTRAN.

Acknowledgment should also be given to all those who contributed error matrices for use in this study. Without this contribution of data the testing of these techniques could never have been done.

## ABSTRACT

This report documents and summarizes the accomplishments over the past year in two areas: (1) development of Landsat classification accuracy assessment techniques, and (2) development of a computerized system for assessing wildlife habitat from land cover maps. This report includes a literature review on accuracy assessment techniques, a complete explanation for the techniques developed under both projects, including example analyses and listings of the computer programs.

A summary of the presentations and discussions at the National Working Conference on Landsat Classification Accuracy is included. Also, two symposium papers which have been published on the results of this project are included as appendices.



## 1.0 Introduction

Many studies have been conducted to determine the usefulness of LANDSAT data for mapping land cover. However, very little research has been done to determine the degree of success (i.e., accuracy) in doing this. A recent literature review by Mead (1977) indicated that:

...more work is needed to develop reliable techniques for estimating classification accuracies. A means of comparing the accuracies (i.e., to compare classification matrices) obtained in different areas on different dates, or estimated by different techniques is needed. Such techniques should permit the investigator to test hypotheses that at specified level of confidence the accuracies from several areas, dates, etc. are not different.

(p. 59)

Mead (1977) continues by suggesting "Future studies might consider iterative proportional fitting of the classification matrices as a means of doing this." (Bishop et al. 1975).

The apparent absence of quantitative methods for comparing classification accuracy is certainly a stumbling block that must be overcome. The effects of imaging date, spectral band combination, classification algorithm, training set selection procedure, and the image analyst on final classification accuracy must be studied. Therefore, the following study was proposed with these objectives:

### 1.1 Objectives

1. To develop a computer system that implements an iterative proportional fitting technique to "normalize" the coefficients within classification error matrices.

2. To develop hierarchal models for testing the significance of several factors (e.g., image date, classification algorithm, the analyst, etc.) on the resulting classification accuracy.
3. To test the above techniques and determine their usefulness with actual data for classification accuracy.

### 1.2 Justification

Research will undoubtedly continue toward development of a system for classification of land cover from digitally recorded Landsat imagery. Such research efforts will in part be measured by improvements in the classification accuracies achieved. Therefore scientists will need ways of assessing the accuracy. Also the accuracy of the final maps produced must be verified before they are distributed to users. Once standards are established, rigorous statistical procedures will be needed to maintain the quality of the maps. Therefore, it can be seen that accuracy assessment techniques will be needed in both the research and operational environments.

### 1.3 State of the Art of Landsat Classification Accuracy Assessment

Landsat, like any other remote sensing system, is only as good as our ability to evaluate it. The need for techniques to assess the accuracy of the Landsat sensor systems cannot be understated. As Freese (1960) states, "testing the accuracy of some measurement against



an accepted standard requires a statement of the accuracy required, a measure of the accuracy attained, and an objective method of deciding whether the accuracy attained is equal to the accuracy required". If there are no methods for measuring the accuracy attained with a certain sensor system, then there will be no way to make comparisons between systems to determine which is better.

If Landsat is ever to become an operational system, then evaluation and accuracy assessment techniques must be developed to show where such sensor systems give more adequate results than conventional methods. These assessment techniques must then be applied to specific applications. For example, "the usefulness of satellite imagery for forestry depends on the extent to which forest data can be recorded by a remote sensing system from satellite altitudes, processed by an image interpretation system, and used in forest mapping and inventories" (Kalensky and Scherk, 1975).

#### 1.31 Accuracy Assessment Techniques

There have been very few studies done on accuracy of Landsat classification. Most of the early assessments were done as an "after thought" without much consideration given to the statistical methods used. These studies, such as the one done by Kalensky and Scherk (1975), usually dealt only with training set accuracy. The use of training sets as well as other possible areas to be assessed will be discussed later.

A review of the current assessment techniques are necessary before any of the applications of these techniques can be understood.

The most common way to describe the accuracy of a Landsat image is in the form of an error matrix (e.g., Todd et al., 1980; Mead and Meyer, 1977; Hoffer, 1975). An error matrix is a square array of numbers set out in rows and columns which express the number of pixels assigned as a particular land cover type relative to the actual land cover as verified in the field or from photos. The columns usually represent the ground truth and the rows indicate the computer assigned land cover category. This form of expressing accuracy as an error matrix allows for an effective way to evaluate both errors of inclusion (commission errors) and errors of exclusion (omission errors) present in the classification. Also, the error matrix allows the analyst to determine the performance for individual categories as well as for the overall classification (Hoffer and Fleming, 1978). In the ideal situation, all the non-major diagonal elements of the error matrix would be zero, indicating that no pixel had been misclassified (Lillesand and Kiefer, 1979).

There are two basic types of accuracy assessments. They are site specific accuracy and non-site specific accuracy. All the methods described to assess accuracy can be applied to either type. Non-site specific accuracy is less useful than site specific accuracy. Meyer et al. (1975) used a non-site specific accuracy assessment to evaluate

classification of Landsat imagery in Southeastern Montana. Total area acreages were calculated for each informational class. There were no tests made for positional accuracy (site specific), just relative total acreages. Meyer found the estimate of the relative proportion of each cover type compared favorably with the ground truth (i.e., actual acres of each land cover category). However, he also noticed that omission and commission errors were very obvious and that the overall positional accuracy of the cover types within the areas studied was poor.

This example points out the major disadvantages of a non-site specific accuracy assessment. If only total acreage estimates are needed, then this method may apply. However, the natural resource manager is usually interested in the location as well as the acreage of a certain land cover category. If this is the case, it is obvious that non-site specific accuracy assessment is not adequate.

Site specific accuracy, on the other hand, is a measure of how well the computer (classification algorithm) classifies each pixel with respect to the ground truth. It is a more meaningful representation of the accuracy of the classification. The analyst can see which categories are easily identifiable and which are being confused. Although Lyon (1979) used site specific accuracy assessment, he includes no error matrices in his paper. Instead, he gives just one number as a measure of the accuracy. This is a common problem throughout the literature. Without error matrices,



the reader has little chance of understanding how an accuracy figure was determined. The reader also loses the knowledge of which categories were easily identified and which were difficult.

Once the error matrix has been generated, a very simple procedure can be used to determine the overall accuracy. Since all the values on the major diagonal represent those pixels that have been correctly classified, if one adds up the major diagonal and divides this number by the total number of pixels classified, one will obtain the overall accuracy of that error matrix. This is the most common use of the error matrix in accuracy assessment.

In recent years, some new techniques have been developed to assess classification accuracy. Among these new methods are analysis of variance techniques, regression analysis techniques, and discrete multivariate analysis techniques. Each of these methods has certain assumptions that must be met before the technique can be used for assessing classification accuracy. If these assumptions are not met, the technique loses its power.

The data used in classification accuracy assessment is of the discrete type. Discrete data, as opposed to continuous data, may take on only a limited number of distinct values (Snedecor and Cochran, 1976). In analysis of variance, the data must be normally distributed in order to meet the assumptions of the technique. Since discrete data is not

normally distributed, it would seem that ANOVA is not a good technique for accuracy assessment. However, Rosenfield (1980) has proposed the use of the logit transformation or the arcsine transformation as described by Snedecor and Cochran (1976) to transform the data into an approximately normal distribution. Rosenfield states, "the statistically interpreted results of the weighted adjustment agree fairly well with what might be technologically expected, and are therefore judged technically acceptable". After the transformation is applied to the data, the analysis of variance can be run. From the resulting ANOVA table, multiple range tests are applied to population means found to be significantly different (Rosenfield, 1980). Analysis of variance is a powerful statistical tool. However, other techniques that do not require so much data manipulation should also be tested. Rosenfield (1978) agrees, "this does not mean that they (ANOVA) are the best; however, the tools available should be used until something better comes along".

Regression analysis is another way of visually representing accuracy. In this case the ground truth (i.e., actual land cover) is the independent variable, X, and the computer classification is the dependent variable, Y. If the computer is completely correct in its classification, then all the points will lie on a forty five degree line. More likely, the points will be spread out from this line. The value of the correlation coefficient can then be used to get an idea of the



relative agreement between the ground truth and the computer classification. Regression analysis has not been widely used in the literature and therefore no more will be said about it.

### 1.32 Sampling Techniques

The need to use more than just training areas for accuracy assessment has already been discussed. However, one could not afford nor desire to assess the entire scene. Instead, a representative sample should be chosen and assessed as the accuracy for the entire scene. Sampling allows not only the calculation of a number that represents the accuracy of the classification, but also allows for a confidence interval to be placed around that number.

Ginevan (1979) states three criteria that should be satisfied in any sampling scheme. These criteria are: (1) the sampling scheme should have a low probability of accepting a map of low accuracy, (2) the sampling scheme should have a high probability of accepting a map of high accuracy, and (3) the sampling scheme should have a minimum number,  $N$ , of ground truth samples. Many researchers (Hay, 1979; Ginevan, 1979; and Genderen and Lock, 1977) agree that stratified random sampling is the best sampling scheme to use. Rhode (1978) proposes other schemes including cluster-stratified sampling and two phase sampling. No matter which sampling scheme is used, it should be chosen so to obtain the maximum information with the minimum amount of work. This involves

considering many variables such as terrain, image identifiable locations, and variability of land cover categories.

It should also be noted that errors arise in classification from other sources besides the sampling scheme chosen. Problems arise in radiometric correction and geometric rectification. Also, the time interval between when the imagery is attained and when the field checking is done may cause differences in land cover category. It must also be realized that just because the classification of a category seems perfect, this does not always mean that the method is error free. The result may occur purely by chance because of the sampling design. "This fact is seldom appreciated by many image interpreters when checking the accuracy results of their remote sensing land use survey (Genderen and Lock, 1977).

Finally, no matter which sampling scheme is chosen, a sample size must be determined. This situation is described by Ginevan (1979), "The sampling problem as defined here is the determination of the optimal number,  $N$ , of ground truth samples and an allowable number,  $X$ , of misclassifications of these samples." Once these have been determined, the results of image interpretation are checked against the  $N$  ground truth samples and the map is accepted as accurate if  $X$  or fewer of the ground truth samples are misclassified. The optimum number of samples,  $N$ , to be taken has met with widespread disagreement throughout the literature (Todd et al., 1980; Hay, 1979; Genderen et al., 1978; Genderen and Lock, 1977; and Hord and Brooner, 1976). Each researcher seems to have his

own ideas about sample size determination and it is obvious that a great deal more research is needed in this area.

### 1.33 National Data Base for Error Matrices

Letters were sent out to potential sources of error matrices asking that any matrices they had be sent to us for inclusion in a National Data Base for Error Matrices. An information questionnaire was sent along with each request for data. This questionnaire contained questions about the location of the area analyzed, the analyst, the algorithm, and the date the data were taken.

All error matrices that we have received have been compiled along with their corresponding pertinent information and placed on a computer tape. This data are available for distribution to other users upon request. A listing of the sources of error matrices can be found in Appendix I.



## 2.0 Statistical Methods

### 2.1 Methods of Comparing Similarity Matrices

Two methods were used to compare two-dimensional matrices representing ground classification versus machine classification from different methods. In the first method of comparison the cell entries in each matrix are successively balanced until the sum of each of the matrix margins is one. The entries in the matrix then represent a normalized percentage of the total observations occurring in each matrix cell. Within an individual matrix these percentages can be used to examine omission and commission errors. Classification errors between two or more machine classification methods can be evaluated by comparing the percentages in corresponding cells in each matrix. Matrices with differing numbers of observations can be compared since the entries in each matrix are transformed to percentages.

The second method of comparison was a measure of agreement for two-dimensional square matrices presented by Bishop et al. (1975). This measure,  $\hat{K}$ , is calculated as the difference between the actual agreement and chance agreement between two classification methods. In this application the two methods are ground classification and machine classification. The measure is calculated as

$$\hat{K} = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r X_{i+} X_{+i}}{N^2 - \sum_{i=1}^r X_{i+} X_{+i}}$$

where  $r$  is the number of rows in the matrix,  $X_{ii}$  is the number of observations in row  $i$  and column  $i$ ,  $X_{i+}$  and  $X_{+i}$  are the marginal totals of row  $i$  and column  $i$ , respectively, and  $N$  is the total number of observations. An approximate large sample variance, based on the asymptotic normality of  $\hat{K}$ , is available, and can be used to derive a confidence interval for  $\hat{K}$  from a single matrix and to perform tests for equality of  $\hat{K}$  between two matrices.

The two methods described above can be used together. Method two,  $\hat{K}$ , will indicate whether two matrices exhibit the same degree of classification success (or error). If a difference exists, method one can be used to determine in which particular category or categories the difference lies.

## 2.2 Categorical Data Analysis

The influence of factors such as season of imaging, film type, and interpreter bias on classification accuracy was examined using categorical data analysis (Bishop *et al.*, 1975). Using this analysis technique the dependence of classification accuracy on a single factor or combination of factors can be assessed.



Categorical data analysis requires only that each factor being examined for influence on accuracy can be assigned to an unambiguous category within each factor. These categories may be normative, ordinal, or interval. The result of data collection is a multidimensional matrix with each factor, including ground and machine classification, serving as a dimension of the matrix.

This method of analysis avoids the more restrictive assumptions inherent in alternative analysis methods such as multivariate regression or analysis of variance. No normality assumption is necessary, no factors need be considered as continuous, and interpretation of many dummy variables is avoided.

### 3.0 Sample Data Analyses

#### 3.1 MARGFIT Analysis

As previously discussed, the FORTRAN computer program MARGFIT (see Appendix II) implements a normalization procedure which standardizes each error matrix for purposes of comparison. The accuracy of the classification can then be represented as a normalized overall performance. This value is calculated the same way as in overall performance (i.e., summing the major diagonal and dividing by the total) except that the matrix is normalized first.

Smith and Itkowsky (1978) compiled five error matrices for a study in north central Colorado. Two of the matrices were for training sets; Original was compiled using a supervised classification while Josesigs was compiled using a modified unsupervised classification. The other three matrices (Scrambl1, Scrambl2, Scrambl3) were attempts to reclassify incorrect pixels using a computer program called SCRAMBL. Table 1 shows the Josesigs error matrix before normalization and Table 2 shows the matrix after normalization.



Table 1. Josesigs error matrix before normalization.

		Reference Data							
		Decid.	Conif.	Grass	Meadow	Shrub	Water	Sage	
Computer Classification	Decid.	17	2	0	0	0	0	0	Overall Performance  $\frac{398}{463} = .8596$
	Conif.	28	127	0	0	0	0	0	
	Grass	0	0	2	2	1	0	0	
	Meadow	16	0	0	122	6	0	0	
	Shrub	6	0	0	4	3	0	0	
	Water	0	0	0	0	0	127	0	
	Sage	0	0	0	0	0	0	0	
									$\Sigma = 398$

Table 2. Josesigs error matrix after normalization.

		Reference Data							
Computer Classification		Decid.	Conif.	Grass	Meadow	Shrub	Water	Sage	
	Decid.	.5363	.0996	.1001	.0168	.0645	.0155	.1674	Normalized Overall Performance  $\frac{4.2958}{7} = .6136$
	Conif.	.1382	.8044	.0158	.0027	.0102	.0025	.0265	
	Grass	.0154	.0200	.5025	.0842	.1943	.0156	.1681	
	Meadow	.0879	.0035	.0174	.7139	.1457	.0027	.0291	
	Shrub	.1804	.0180	.0906	.1366	.4089	.0141	.1515	
	Water	.0035	.0046	.0230	.0039	.0148	.9108	.0385	
	Sage	.0384	.0499	.2505	.0420	.1615	.0389	.4190	
									$\Sigma = 4.2958$



Table 3 shows the results of overall performance and normalized overall performance for all five error matrices. Note that the relative accuracies are similar for the two performance values except for the Josesigs matrix. Careful study of Table 1 shows why this is so. Only three pixels in the shrub category were correctly classified. This forced the normalization procedure to inflate the values in the shrub row and column decreasing the normalized performance accuracy. Also, no sage category pixels were classified at all resulting in the same type of normalization problem.

Table 3. Overall and normalized overall performance results for five classification error matrices.

Matrix	Overall Performance	Normalized Overall Performance
Original	90.37%	86.03%
Josesigs	85.96%	61.36%
Scramb11	85.43%	79.97%
Scramb12	78.94%	70.49%
Scramb13	80.18%	74.17%

Similar results were achieved for matrices compiled by Hoffer (1975a). Here four error matrices were compiled at two different dates comparing a classification of major land cover types versus forest cover types. The results of normalization shown in Table 4 agree with the overall performance values calculated by Hoffer.

Table 4. Overall and normalized overall performance results for four cover type error matrices.

Matrix	Overall Performance	Normalized Overall Performance
Major Land Cover Types 6-5-73	85.96%	89.51%
Major Land Cover Types 8-8-73	69.35%	72.53%
Forest Cover Types 6-5-73	71.79%	76.87%
Forest Cover Types 8-8-73	48.83%	57.88%

### 3.2 KAPPA Analysis

The FORTRAN computer program KAPPA (see Appendix III) calculates a  $\hat{K}$  statistic for a given error matrix which allows one to compare

error matrices to see if they are significantly different. This type of comparison has many uses. In an example cited above, Hoffer (1975a) compiled two classifications at two different dates. The  $\hat{K}$  statistic and corresponding confidence interval (i.e., upper and lower bounds) are presented for each error matrix in Table 5.

Table 5.  $\hat{K}$  statistic with upper and lower limits at 95% confidence interval for four cover type error matrices.

Matrix	Lower Limit	$\hat{K}$	Upper Limit
Major Land Cover Types 6-5-73	.69396	.69458	.69521
Major Land Cover Types 8-8-73	.62880	.62929	.62978
Forest Cover Types 6-5-73	.38961	.39055	.39150
Forest Cover Types 8-5-73	.33004	.33074	.33144



As can be seen from Table 5, none of the confidence intervals overlap; therefore, all these matrices are significantly different. This means that the imagery taken at two different dates is significantly different which implies that one date must then be better than the other. A quick look at the data indicates that 6-5-73 was the significantly better date.

Another example of this technique is provided by Hoffer (1975b). In this example, four matrices were generated from four different classification algorithms. The results presented in Table 6 show that all the matrices are significantly different.

Table 6.  $\hat{K}$  statistic with upper and lower limits at 95% confidence interval for four classification algorithms.

Matrix	Lower Limit	$\hat{K}$	Upper Limit
Nonsupervised (10 cl.)	.60271	.60479	.60686
Nonsupervised (20 cl.)	.58348	.58573	.58799
Modified Supervised	.47326	.47581	.47837
Modified Cluster	.71631	.71846	.72001

A final example of the  $\hat{K}$  statistic is found in Appendix V. This example deals with comparing photo interpreters to see if they are significantly different.

### 3.3 CONTABLE Analysis

The APL computer program CONTABLE (see Appendix IV) allows one to analyze multi-way contingency tables.\* In the example here a 5-way table is analyzed. This table (Carneggie, 1972) deals with 5 factors or effects listed in Table 7. The data consists of 18 5 x 5 error matrices with various films, dates, and interpreters.

Table 7. List of factors and effects for 5-way contingency table.

<u>FACTOR</u>	<u>EFFECT</u>	
1	Date	(6/10, 7/25, 10/25)
2	Film	(Color, CIR)
3	Interpreter	(#1, #2, #3)
4	Row	(1, 2, 3, 4, 5)
5	Column	(1, 2, 3, 4, 5)

The hypotheses to be tested in this example are listed in Table 8 while the results and conclusions are listed in Table 9.

\*Without the use of this program and its Iterative Proportional Fitting Procedure, analysis of tables larger than 3 dimensions would be impossible.

Table 8. List of hypothesis for CONTABLE example.

1.  $H_0: \mu_2 = 0$  No film effect
2.  $H_0: \mu_3 = 0$  No interpreter effect
3.  $H_0: \mu_1 = 0$  No date effect
4.  $H_0: \mu_4 = \mu_5 = 0$  No row-column effect
5.  $H_0: \mu_{12} = 0$  No date-film interaction
6.  $H_0: \mu_{13} = 0$  No date-interpreter interaction
7.  $H_0: \mu_{23} = 0$  No film-interpreter interaction

Table 9. List of results and conclusions for CONTABLE example.

<u>HYPOTHESIS</u>	<u>CHI SQUARE VALUE</u>	<u>CONCLUSION</u>
$H_0: \mu_2 = 0$	623.487	reject $H_0$
$H_0: \mu_3 = 0$	613.142	reject $H_0$
$H_0: \mu_1 = 0$	591.543	reject $H_0$
$H_0: \mu_4 = \mu_5 = 0$	134.485	fail to reject $H_0$
$H_0: \mu_{12} = 0$	145.961	fail to reject $H_0$
$H_0: \mu_{13} = 0$	162.393	fail to reject $H_0$
$H_0: \mu_{23} = 0$	144.707	fail to reject $H_0$



Table 9 shows that although no single factor significantly affects the classification, the combination of two or more factors does. This means that none of the three factors (film, date, interpreter) is more important than the others. Instead all three factors interact together to give the best classification. From the analysis so far there is no significant one factor on which most of the accuracy depends.

#### 4.0 Accuracy Conference

A National Working Conference on Landsat Classification Accuracy Assessment Procedures was held in Sioux Falls, South Dakota. A summary of this conference is given in Appendix VI as a draft manuscript which will be revised and submitted for publication in journal.

#### 5.0 Wildlife Habitat Assessment Methods

A secondary task in this year's plan of work was to develop digital spatial analysis techniques for assessing wildlife habitat. Appendix VII includes a FORTRAN computer program for doing this, and the techniques are described in a manuscript which is Appendix VIII.

#### 6.0 Effects of Classification Accuracy on Interspersion Maps

Artificial land cover type maps were made in order to test the effects of classification accuracy on computer generated interspersion maps. Three cellular maps were made, each containing 10 rows and 10 columns with each cell assigned to one of 5 classes. The first map was used as a reference base map for comparison with the other two maps. The second map had 90% of its cells classified similar to the first (i.e., 90% accurate), and the third map was 70% similar to the reference base map. Five cover types, designated 1 through 5, were used on each map. Similarity matrices were generated between the accurate

(90% similar) "Map II" and the base map (Table 10), and between the less accurate, "Map III" (70% similar) and the base map (Table 11). KAPPA was used to compare the two resulting similarity matrices.

The interspersion index described by Mead et al. in Appendix VIII was used to create interspersion maps from each of the three fictional cover type maps. The maps delineate areas of high (designated 3), medium (2), and low (1) interspersion. Similarity matrices were created by comparing each of the interspersion maps (from the cover type maps II and III) with the interspersion map made from the base map (Tables 12 and 13).

The implementation of the KAPPA program (see Section 3.2) was then used to test for a significant difference between the interspersion maps. The resulting KHAT values indicate that cover type maps II and III were significantly different. A significant difference was also found between the two matrices for the interspersion maps. However, further work is needed to understand the effect of map accuracy on computer generated interspersion maps, juxtaposition maps, and spatial diversity maps. Also, the effect of increasing the number of cover types or the number of interspersion classes (high, medium and low) is unknown.



Table 10. Similarity matrix for five fictional cover types on the base map and on Map II.

		Base Map Classification				
		1	2	3	4	5
Map II Classification	1	19			1	
	2		18	1		
	3		1	16		
	4	1		2	27	1
	5				3	10

$$\text{Overall Accuracy} = \frac{90}{100} = 90\%$$

Table 11. Similarity matrix for five fictional cover type maps on the base map and on map III.

		Base Map Classification				
		1	2	3	4	5
Map III Classification	1	14	1		2	
	2	2	16	4	3	
	3	1	2	13	2	
	4	3		2	17	2
	5				6	10

$$\text{Overall Accuracy} = \frac{70}{100} = 70\%$$

Table 12. Similarity matrix for three categories of interspersions, high (3), medium (2), and low (1) produced from the base map and map II.

		Base Map Interspersion		
		1	2	3
Map II Interspersion	1	60		
	2		15	1
	3		3	21

Table 13. Similarity matrix for three categories of interspersions, high (3), medium (2), and low (1) produced from the base map and map III.

		Base Map Interspersion		
		1	2	3
Map III Interspersion	1	60		
	2		13	
	3		5	22

## 7.0 Summary and Future Work

The literature review and preliminary investigations show that: (1) the statistical techniques initially proposed are sound and are useful for analysis of Landsat classification accuracy data, (2) substantial amounts of data from accuracy assessments exist but few sets are comparable prohibiting hypotheses from being tested, (3) preliminary results show that the method used in sampling a classification can significantly affect the estimated accuracy. An "automatic" computerized system needs to be developed for compiling error matrices for any classification given the necessary ground truth and a specified sampling strategy. Experiments need to be designed in the future so that fundamental questions can be answered about factors which affect classification accuracy.

The wildlife habitat assessment system has greatest potential when animals with requirements related to the spatial characteristics of the landscape are considered. Juxtaposition can be of great importance or of very little importance depending upon the specific geographic area and the wildlife species of interest. When this data on the spatial characteristics of the landscape are coupled with basic land cover information and ancillary data (e.g., elevation, slope, soil type, political or ownership boundaries), an over-all system for habitat assessment may be realized. Such a system could be implemented on a computer and merged with data on other resource



attributes (e.g., timber producing capability). Further work should include pilot testing the system and an evaluation by field level resource managers.

### 8.0 Literature Cited

- Bishop, Y., S. Fienberg, and P. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press: Cambridge, Massachusetts. 575 pp.
- Carnegie, D. M. 1972. Large scale 70 mm aerial photographs for evaluating ecological conditions, vegetational changes, and range site potential. Ph.D. Dissertation, University of California, Berkeley. 180 pp.
- Coggeshall, M. G. and R. M. Hoffer. 1973. Basic forest cover mapping using digitized remote sensor data and ADP techniques. Laboratory for Applications in Remote Sensing, Purdue University. LARS Information Note 030573.
- Fleming, M. D., J. S. Berkebile, and R. M. Hoffer. 1975. Computer aided analysis of Landsat-1 MSS data: a comparison of three approaches, including a modified clustering approach. Laboratory for Applications in Remote Sensing, Purdue University. LARS Information Note 072475.
- Freese, Frank. 1960. Testing accuracy. Forest Science. Vol. 6, No. 2. pp. 139-145.
- Genderen, J. L. van and B. F. Lock. 1977. Testing land use map accuracy. Photogrammetric Engineering and Remote Sensing. Vol. 43, No. 9. pp. 1135-1137.

- Genderen, J. L. van, B. F. Lock, and P. A. Vass. 1978. Remote sensing: statistical testing of thematic accuracy. In: International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan. pp. 3-14.
- Ginevan, Michael E. 1979. Testing land-use map accuracy: another look. Photogrammetric Engineering and Remote Sensing. Vol. 45, No. 10. pp. 1371-1377.
- Hay, Alan M. 1979. Sampling designs to test land-use map accuracy. Photogrammetric Engineering and Remote Sensing. Vol. 45, No. 4. pp. 529-533.
- Hoffer, R. M. 1975. Computer aided analysis of skylab multi-spectral scanner data in mountainous terrain for land use, forestry, water resources, and geologic applications. Laboratory for Applications in Remote Sensing, Purdue University. LARS Information Note 121275. 381 pp.
- Hoffer, R. M. 1975. Natural resource mapping in mountainous terrain by computer analysis of ERTS-1 satellite data. Laboratory for Applications in Remote Sensing, Purdue University. LARS Research Bulletin 919. LARS Information Note 061575. 124 pp.
- Hoffer, R. M. and M. D. Fleming. 1978. Mapping vegetative cover by computer aided analysis of satellite data. Laboratory for Applications in Remote Sensing, Purdue University. LARS Technical Report 011178. 10 pp.



- Hord, R. Michael and William Brooner. 1976. Land-use map accuracy criteria. Photogrammetric Engineering and Remote Sensing. Vol. 42, No. 5. pp. 671-677.
- Itkowsky, Frank A. and James A. Smith. 1978. Sensitivity of variable probability sampling estimates to initial Landsat classification. Final Report, Volume I, Rocky Mountain Forest and Range Experiment Station. USFS Coop. Agreement 16-741-CA. Colorado State University. 199 pp.
- Kalensky, Z. and L. Scherk. 1975. Accuracy of forest mapping from Landsat computer compatible tapes. In: International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan. Tenth Proceedings. pp. 1159-1167.
- Landgrebe, D. A. 1973. Machine processing for remotely acquired data. Laboratory for Applications in Remote Sensing, Purdue University. LARS Information Note 031573. 30 pp.
- Lillesand, T. M. and R. W. Kiefer. 1979. Remote Sensing and Image Interpretation. John Wiley and Sons: New York. 612 pp.
- Lyon, John G. 1978. An analysis of vegetation communities in the lower Columbia river basin. In: Pecora IV Symposium. Sioux Falls, South Dakota. pp. 321-327.
- Lyon, John G. 1979. Remote sensing analyses of coastal wetlands characteristics: the St. Clair Flats, Michigan. In: International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan. Thirteenth Proceeding. pp. 1117-1129.

- Mead, R. A. and M. P. Meyer. 1977. Landsat digital data application to forest vegetation and land use classification in Minnesota. In: Machine Processing of Remotely Sensed Data, Proceedings, Purdue University. pp. 270-280.
- Meyer, M., J. Brass, B. Gerbig, and F. Batson. 1975. ERTS data applications to surface resource surveys of potential coal production lands in southeast Montana. IARSL Research Report 75-1. Final Report, University of Minnesota. 24 pp.
- Rohde, Wayne. 1977. Digital image analysis techniques required for natural resource inventories. National Computer Conference Proceedings. Vol. 47. pp. 93-106.
- Rosenfield, G. H. 1978. Personal Correspondence, Comments on Roy A. Mead's Ph.D. dissertation, March 16, 1978.
- Rosenfield, G. H. 1980. Analysis of variance of thematic mapping experiment data. Presented at the American Society of Photogrammetry Annual Conference, St. Louis, Missouri.
- Snedecor, George W. and William G. Cochran. 1976. Statistical Methods. Sixth Edition. The Iowa State Press, Ames, Iowa. 593 pp.

Todd, W., D. Gehring, and J. Haman. 1980. Landsat wildland mapping accuracy. Photogrammetric Engineering and Remote Sensing. Vol. 43, No. 9. pp. 1135-1137.



## Appendix I.

## List of Sources of Error Matrices

1. Mead, Roy A., Landsat Digital Data Application to Forest Vegetation and Land-Use Classification in Minnesota. Ph.D. Dissertation, University of Minnesota, 1977.  
4 matrices (training set, test set, 2 evaluation areas)  
Minnesota.
2. Fleming, Michael, Computer Aided Analysis Techniques for an Operational System to Map Forest Lands Utilizing Landsat MSS Data, LARS Technical Report 112277.  
2 matrices      Colorado.
3. Smith, James and Frank Itkowsky, Sensitivity of Variable Probability Sampling Estimates to Initial Landsat Classification, Final Report R.M.F. & R.E.S. USFS Coop-Agree. 16-741-CA, September 1978, CSU, Fort Collins, Colorado.  
5 matrices (training set, test set, 3 evaluation areas)  
Colorado.
4. Madding, Robert and Harland Hogan, Detection and Mapping of Spruce Budworm Defoliation in Northern Wisconsin Using Digital Analysis of Landsat Data. Proceedings of ASP Convention. Feb. 26 - Mar. 4, 1978. pp 285-300.  
2 matrices (normal and collapsed)  
Wisconsin.

5. Voss, A. W., J. E. Baker, G. E. Hauser, and D. W. Newton, The Use of Landsat Derived Land Cover Data in a Flood Peak Correlation Study, Proceedings ASP, Feb. 26-Mar. 4, 1978, pp. 135-146.  
2 matrices (normal and collapsed)  
North Carolina - Tennessee.
6. Hoffer, Roger, Natural Resource Mapping in Mountainous Terrain by Computer Analysis of ERTS-1 Satellite Data, LARS Research Bulletin 919. Info. Note 061575.  
10 matrices (different classification systems)  
Colorado.
7. Hoffer, Roger, Computer-Aided Analysis of Skylab MSS Data in Mountainous Terrain for Land Use, Forestry, Water Resources, and Geologic Applications, LARS Info. Note 121275, 1975.  
4 matrices (varying spectral bands)  
Colorado.
8. Hoffer, Roger, Mapping Vegetative Cover by Computer Aided Analysis of Satellite Data, LARS Technical Report 011178.  
2 matrices (test sites)  
Colorado.
9. Hoffer, Roger, Variables in Automatic Classification over Extended Remote Sensing Test Sites, LARS Information Note 061571.  
1 matrix (test site)  
Indiana - Illinois.

10. Hoffer, Roger, Basic Forest Cover Mapping Using Digitized Remote Sensor Data and ADP Techniques, LARS Information Note 030573.  
13 matrices (tests at different spectral bands)
11. Heller, R. C., R. C. Aldrich, R. S. Driscoll, R. E. Francis, and F. P. Weber, Evaluation of ERTS-1 Data for Inventory of Forest and Rangeland and Detection of Forest Stress. PSW & RM For & Range Exp. Sta. Aug. 9, 1974.  
12 matrices.
12. Ernst, Carola Lisette, Digital Processing of Remotely Sensed Data for Mapping Wetland Communities, Ph.D. Dissertation, Purdue University, Dec. 1979.  
6 matrices (classification)  
Indiana.
13. Nelson, R. and R. Hoffer, Computer Aided Processing of Landsat MSS Data for Classification of Forest Lands, LARS Technical Report 102679, 1979.  
12 matrices  
Colorado.
14. Carneggie, D. M., Large Scale 70 mm Aerial Photographs for Evaluating Ecological Conditions, Vegetational Changes, and Range Site Potential. Ph.D. Dissertation, University of California, Berkeley.  
18 matrices (photo interpretation)



15. Lauer, Donald, Claire Hay and Andrew Benson, Quantitative Evaluation of Multiband Photographic Techniques, Final Report for Earth Observation Division Manned Spacecraft Center, NASA Contract NAS 9-9577, 1970.  
79 matrices (photo interpretation)
16. Bryant, Emily and Gibb Dodge  
1 matrix. Maine.
17. Roberts, Edwin  
1 matrix. Colorado (test set for Grand County).
18. Roller, Norman and Larry Visser, Accuracy of Landsat Forest Cover Type Mapping in the Lake States Region of the U.S., Fourth International Symposium on Remote Sensing of Environment, April 23-30, 1980.  
1 matrix (Forest Cover Types).  
Michigan.
19. Newcomer, Jeffrey  
3 matrices. Pennsylvania.
20. Harrington, John A. and Charles W. Dunn, Jr.  
3 matrices (forest - other) Oklahoma.

## Appendix II. Listing of FORTRAN Computer Program MARGFIT

```

//WATFIV          ,PAGES=30
C                *****
C                *
C                *   MARGFIT WAS REWRITTEN AND DOCUMENTED BY
C                *   RUSSELL G. CONGALTON
C                *   DEPT. OF FORESTRY, VPI&SL
C                *   JULY 1979
C                *****
C
C *****
C * THIS PROGRAM WAS DESIGNED TO CHANGE A MATRIX OF MAXIMUM DIMENSIONS
C * OF 50X50 INTO A MATRIX WITH PREDETERMINED ROW AND COLUMN MARGINALS
C *
C * NUMMAT   = THE NUMBER OF MATRICES TO BE CHANGED
C * TAB(I,J) = THE VALUE IN ROW I AND COLUMN J OF THE GIVEN MATRIX
C * RMAR(I)  = THE MARGINAL VALUE FOR ROW I
C * CMAR(I)  = THE MARGINAL VALUE FOR COLUMN J
C * MAXIT    = THE MAXIMUM NUMBER OF ITERATIONS
C * MAXDEV   = THE MAXIMUM ALLOWABLE DEVIATION
C * IR       = THE NUMBER OF ROWS IN THE MATRIX
C * IC       = THE NUMBER OF COLUMNS IN THE MATRIX
C *****
C
1  DIMENSION TAB(50,50),RMAR(50),CMAR(50),FIT(50)
2  REAL MAXDEV
C
3  NCOUNT=0
4  READ(5,15) NUMMAT
5  FORMAT(12)
C
6  50 READ(5,10) IR,IC
7  10 FORMAT(2I5)
8  DO 200 I=1,IR
9  READ(5,20) (TAB(I,J),J=1,IC)
10 20 FORMAT(12(F6.1))
11 200 CONTINUE
12 READ(5,30) (RMAR(I),I=1,IR)
13 30 FORMAT(12(F6.0))
14 READ(5,40) (CMAR(I),I=1,IC)
15 40 FORMAT(12(F6.0))
16 READ(5,51) MAXIT,MAXDEV
17 51 FORMAT(15,F10.3)
18 READ(5,52)
19 52 FORMAT('AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA')
C
20 WRITE(6,999)
21 999 FORMAT('1')
22 WRITE(6,52)
23 WRITE(6,53)
24 53 FORMAT(1X,'*****'//)
25 WRITE(6,54)
26 54 FORMAT(1X,'THE ORIGINAL MATRIX IS:')
27 WRITE(6,55)
28 55 FORMAT(1X,'-----'//)
29 DO 250 I=1,IR

```

```

30      WRITE(6,56) (TAB(I,J),J=1,IC)
31      56 FORMAT(12(1X,F6.1))
32      250 CONTINUE
C
33      NIT=0
34      DO 300 I=1,IR
35      FIT(I)=0.000001
36      DO 400 J=1,IC
37      400 FIT(I)=FIT(I)+TAB(I,J)
38      300 CONTINUE
39      DO 150 I=1,IR
40      DO 500 J=1,IC
41      TAB(I,J)=TAB(I,J)*RMAR(I)/FIT(I)
42      500 CONTINUE
43      150 CONTINUE
44      DO 600 J=1,IC
45      B=0.000001
46      DO 700 I=1,IR
47      700 B=B+TAB(I,J)
48      DO 800 I=1,IR
49      IF (TAB(I,J).LT.1.0E-10) TAB(I,J)=0.0
50      TAB(I,J)=TAB(I,J)*CMAR(J)/B
51      800 CONTINUE
52      600 CONTINUE
53      NIT=NIT+1
54      D=0.0
55      DO 900 I=1,IR
56      FIT(I)=0.000001
57      DO 1000 J=1,IC
58      1000 FIT(I)=FIT(I)+TAB(I,J)
59      H=ABS(FIT(I)-RMAR(I))
60      IF (D.GE.H) GO TO 900
61      D=H
62      900 CONTINUE
63      IF (D.LE.MAXDEV) GO TO 260
64      IF (NIT.LE.MAXIT) GO TO 100
65      WRITE(6,60) MAXIT,D
66      60 FORMAT(' NO CONVERGENCE AFTER ',I5,' ITERATIONS.'// ' THE CURRENT MAX
        AIMUM DEVIATION IS: ',F10.3)
67      GO TO 1200
68      260 WRITE(6,70) NIT,D
69      70 FORMAT(/// ' CONVERGENCE AFTER ',I5,' ITERATIONS WITH A MAXIMUM DEV
        AIATION OF: ',F10.3/)
70      DO 1100 I=1,IR
71      WRITE(6,80) (TAB(I,J),J=1,IC)
72      80 FORMAT(20(1X,F7.4))
73      1100 CONTINUE
C
74      1200 NCOUNT=NCOUNT+1
75      IF (NCOUNT.LT.NUMMAT) GO TO 50
C
76      STOP
77      END

```

//DATA



## Appendix III. Listing of FORTRAN Computer Program KAPPA

```

C /WATFIV          ,PAGES=35
C *****
C * KAPPA WAS REWRITTEN AND DOCUMENTED BY
C * RUSSELL G. CONGALTON
C * DEPT. OF FORESTRY, VPI&SU
C * JULY 1979
C *****
C *****
C * THIS PROGRAM WAS DESIGNED TO TEST FOR SIMILAR DEGREES OF AGREEMENT
C * BETWEEN TWO OR MORE SQUARE ERROR MATRICES
C *
C * ME      = THE NUMBER OF TABLES OR MATRICES TO BE COMPARED
C * NR      = NUMBER OF ROWS; ALSO THE NUMBER OF COLUMNS SINCE THE
C *          MATRIX IS SQUARE
C * X(I,J) = THE VALUE IN THE MATRIX FOR ROW I AND COLUMN J
C *****
C
1  REAL KHAT,LCL
2  DIMENSION X(20,20),SXR(20),SXC(20)
3  DIMENSION UCL(20),LCL(20),KHAT(20)
4  L=20
5  M=0
6  K=1
C
7  READ(5,10) ME
8  10  FORMAT(12)
C
9  100 DO 200 I=1,L
10     SXR(I)=0.0
11     SXC(I)=0.0
12     DO 300 J=1,L
13     300 X(I,J)=0.0
14     200 CONTINUE
C
15  READ(5,20) NR
16  20  FORMAT(12)
17     DO 400 I=1,NR
18     READ(5,30) (X(I,J),J=1,NR)
19     30  FORMAT(12(F6.0))
20     400 CONTINUE
21  READ(5,31)
22  31  FORMAT('AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA')
C
23  WRITE(6,999)
24  999  FORMAT('1')
25  WRITE(6,31)
26  WRITE(6,32)
27  32  FORMAT(1X,'*****'/'//')
28  WRITE(6,34)
29  34  FORMAT('///',1X,'THE ORIGINAL ERROR MATRIX IS:')
30  WRITE(6,35)
31  35  FORMAT(1X,'_____')

```

```

32      DO 450 I=1,NR
33      WRITE(6,36) (X(I,J),J=1,NR)
34      36  FORMAT(20(1X,F6.0))
35      450 CONTINUE
C
36      XN=0.0
37      DO 500 I=1,NR
38      DO 600 J=1,NR
39      SXR(I)=SXR(I)+X(I,J)
40      SXC(J)=SXC(J)+X(I,J)
41      600 CONTINUE
42      XN=XN+SXR(I)
43      500 CONTINUE
44      TH1=0.0
45      TH2=0.0
46      TH3=0.0
47      TH4=0.0
48      DO 700 I=1,NR
49      TH1=TH1+X(I,I)
50      TH2=TH2+SXR(I)*SXC(I)
51      TH3=TH3+X(I,I)*(SXR(I)+SXC(I))
52      DO 800 J=1,NR
53      TH4=TH4+X(I,J)*(SXR(I)+SXC(J))*2
54      800 CONTINUE
55      700 CONTINUE
56      TH1=TH1/XN
57      TH2=TH2/(XN**2)
58      TH3=TH3/(XN**2)
59      TH4=TH4/(XN**3)
60      KHAT(K)=(TH1-TH2)/(1.-TH2)
61      SD=SQRT(TH1*(1.-TH1)/((1.-TH2)**2)+(2.*((1.-TH1)*(2.*TH1*TH2-TH3))
      A)/(1.-TH2)**3)+(1.-TH1)**2*(TH4-4.*TH2**2)/(1.-TH2)**4)/XN
C
C C THE STEPS THAT FOLLOW CALCULATE THE 95% CONFIDENCE INTERVAL FOR KHAT
62      UCL(K)=KHAT(K)+1.96*SD
63      LCL(K)=KHAT(K)-1.96*SD
64      WRITE(6,40)
65      40  FORMAT(//,1X,'LOWER LIMIT',4X,'KHAT',4X,'UPPER LIMIT')
66      WRITE(6,45)
67      45  FORMAT(1X,'_____',4X,'_____',4X,'_____'//)
68      WRITE(6,50) LCL(K),KHAT(K),UCL(K)
69      50  FORMAT(3X,F8.5,3X,F8.5,3X,F8.5)
70      K=K+1
71      M=M+1
72      IF(M.LT.ME) GO TO 100
73      WRITE(6,900)
74      900  FORMAT('1','SUMMARY TABLE AND COMPARISONS')
75      WRITE(6,910)
76      910  FORMAT(1X,'*****'//)
77      WRITE(6,920)
78      920  FORMAT(1X,'MATRIX',2X,'LOWER LIMIT',4X,'KHAT',4X,'UPPER LIMIT')
79      WRITE(6,930)
80      930  FORMAT(1X,'_____',2X,'_____',4X,'_____',4X,'_____'//)
81      DO 940 K=1,ME
82      WRITE(6,950) K,LCL(K),KHAT(K),UCL(K)
83      950  FORMAT(4X,12,5X,F8.5,3X,F8.5,3X,F8.5)
84      940 CONTINUE
85      WRITE(6,960)
86      960  FORMAT(//////////)

```

```

87      N=ME-1
88      DO 1300 I=1,N
89      DO 1400 J=2,M
90      IF (J.LE.I) GO TO 1400
91      IF (LCL(J).GT.LCL(I).AND.LCL(J).LT.UCL(I)) GO TO 1100
92      IF (UCL(J).LT.UCL(I).AND.UCL(J).GT.LCL(I)) GO TO 1100
93      WRITE(6,1000) I,J
94      1000 FORMAT(1X,'MATRIX ',I2,' IS SIGNIFICANTLY DIFFERENT FROM MATRIX ',
95      AX I2/)
95      GO TO 1400
96      1100 WRITE(6,1200) I,J
97      1200 FORMAT(1X,'MATRIX ',I2,' IS NOT SIGNIFICANTLY DIFFERENT FROM MATRI
98      AX ',I2/)
98      1400 CONTINUE
99      1300 CONTINUE
100     WRITE(6,1500)
101     1500 FORMAT('1')
102     STOP
103     END

```

//DATA



## Appendix IV. Listing of APL Computer Program CONTABLE

```

CONTABLEHOW [0]
CONTABLEHOW
[1] 'CONTABLE'
[2] 'ANALYSES OF MULTIDIMENSIONAL CONTINGENCY TABLES'
[3] 'S. K. LEE - - - DEPARTMENT OF STATISTICS, VPI SU'
[4] 'ENTERED 7/5/1976'
[5] ' '
[6] ' THIS PROGRAM WILL PERFORM ANALYSES OF COMPLETE OR INCOMPLETE'
[7] 'MULTIDIMENSIONAL CONTINGENCY TABLES USING A LOGLINEAR MODEL'
[8] 'APPROACH,'
[9] ' '
[10] ' DATA SHOULD BE ARRANGED INTO A CONTINGENCY TABLE, AND MAY BE'
[11] 'STORED IN AN ARRAY SMF PRIOR TO PROGRAM EXECUTION OR MAY BE '
[12] 'ENTERED UPON REQUEST, WHEN ANALYZING INCOMPLETE TABLES THE INITIAL'
[13] 'FITTING TABLE MAY BE STORED SIMILARLY IN AN ARRAY NAMED ONES PRIOR'
[14] 'TO PROGRAM EXECUTION, UPON REQUEST, THE USER SHOULD ENTER A LOGLINEAR'
[15] 'MODEL WITH WHICH HE INTENDS TO FIT THE DATA, THE LOGLINEAR MODEL'
[16] 'SHOULD BE ENTERED BY THE CONFIGURATIONS AND THE DIGIT 0 IS USED TO'
[17] 'SEPARATE CONFIGURATIONS, FOR EXAMPLE, THE NO-THREE-FACTOR-INTERACTION'
[18] 'MODEL OF A 3-DIMENSIONAL TABLE, (C12, C13, C23), IS ENTERED AS:'
[19] ' '
[20] ' 1 2 0 1 3 0 2 3'
[21] ' '
[22] 'THE PROGRAM FOLLOWS AN ITERATIVE PROPORTIONAL FITTING SCHEME TO '
[23] 'COMPUTE THE MAXIMUM LIKELIHOOD ESTIMATES OF THE EXPECTED CELL VALUES,'
[24] 'THEN THE PROGRAM COMPUTES THREE GOODNESS-OF-FIT CHI-SQUARE STATISTICS,'
[25] 'PEARSON CHI-SQUARE, LIKELIHOOD RATIO, AND FREEMAN-TUKEY CHI-SQUARE;'
[26] 'HOWEVER, THE DEGREE OF FREEDOM ASSOCIATED WITH THESE STATISTICS SHOULD'
[27] 'BE DETERMINED BY THE USER, AFTER ALL THE COMPUTATIONS RELATING TO THE'
[28] 'CURRENT MODEL ARE PERFORMED, THE PROGRAM ASKS WHETHER MORE HYPOTHESIS'
[29] '(I.E., ANOTHER LOGLINEAR MODEL) IS TO BE FITTED, IF NOT, PROGRAM'
[30] 'EXECUTION IS TERMINATED,'
[31] ' '
[32] ' ALL VARIABLES ARE LOCALIZED EXCEPT SMF AND ONES, PROGRAM'
[33] 'EXECUTION REQUIRES THE FUNCTIONS NAMED SUM AND YESNO, ALL OF THESE'
[34] 'PROGRAMS PLUS THIS DOCUMENTATION ARE GROUPED UNDER THE NAME CONTABLEGRP,'
[35] ' '
[36] ' FOR MORE DETAILED DISCUSSION OF THE LOGLINEAR MODEL APPROACH SEE'
[37] 'THE FOLLOWING REFERENCE:'
[38] ' '
[39] ' BISHOP, Y.M.M., FIENBERG, S.E. AND HOLLAND, P.W., DISCRETE'
[40] ' MULTIVARIATE ANALYSIS; THEORY AND PRACTICE, CAMBRIDGE,'
[41] ' MASS.; THE MIT PRESS, 1975.'
[42] ' '
[43] 'A SAMPLE RUN OF THE PROGRAM CAN BE OBTAINED FROM S. K. LEE,'
[44] ' '

```

```

      VCONTABLE [0]V
      V CONTABLE;DIM;FOLD;A;W;T;I;J;V;VV;STEPS;IND;M;U;D;C;MNI;MD
[1]  ' '
[2]  'THE DEFAULT PARAMETER VALUES FOR THE ITERATIVE PROPORTIONAL'
[3]  'FITTING SCHEME ARE: MAXIMUM DEVIATION = 0.01'
[4]  '          MAXIMUM NUMBER OF ITERATIONS = 15.'
[5]  ' '
[6]  BEGIN;'ENTER THE DIMENSIONS OF THE TABLE'
[7]  DIM=0
[8]  'IS THE DATA ALREADY IN ARRAY, SMF?(YES, NO, OR STOP)'
[9]  W=YESNO(A+0)
[10]  +((W=0),(W=1))/0,GO
[11]  RESTART;'ENTER THE DATA WITH LAST SUBSCRIPT CHANGING FASTEST'
[12]  SMF=0
[13]  GET;'ANY MORE DATA?'
[14]  W=YESNO(A+0)
[15]  +((W=0),(W=2))/0,GO
[16]  'ENTER MORE DATA'
[17]  SMF+SMF,T+0
[18]  +GET
[19]  GO:+((X/PSMF)=X/DIM)/GO1
[20]  'ERROR, NUMBER OF ELEMENTS INCORRECT, HERE IS THE INPUT'
[21]  SMF
[22]  'DID YOU ENTER ZERO COUNTS? LETS TRY AGAIN'
[23]  +RESTART
[24]  GO1:SMF+DIM*SMF
[25]  'DO YOU WANT TO SEE THE TABLE?'
[26]  W=YESNO(A+0)
[27]  +((W=0),(W=2))/0,GOON
[28]  SMF
[29]  GOON;'ANY FIXED ZEROS?'
[30]  W=YESNO(A+0)
[31]  +((W=0),(W=2))/0,GO11
[32]  'IS THE MATRIX OF ONES AND ZEROS ALREADY IN ARRAY ONES?'
[33]  W=YESNO(A+0)
[34]  +((W=0),(W=1))/0,GO7
[35]  'ENTER A MATRIX OF ONES WITH ZEROS IN THE APPROPRIATE PLACES'
[36]  GO22:ONES=0
[37]  GO6;'ANY MORE?'
[38]  W=YESNO(A+0)
[39]  +((W=0),(W=2))/0,GO7
[40]  'ENTER MORE ONES AND ZEROS'
[41]  ONES+ONES,0
[42]  +GO6
[43]  GO7:A+P,ONES
[44]  +(((X/DIM)=A),((X/DIM)(A)))/GO19,GO20
[45]  'YOU HAVE NOT ENTERED ENOUGH NUMBERS YET! ALL THE REMAINING'
[46]  'NUMBERS ARE TAKEN TO BE ONE'
[47]  ONES+ONES,((X/DIM)-A)*1
[48]  +GO21
[49]  GO20;'YOU HAVE ENTERED TOO MANY NUMBERS;'
[50]  GO21:ONES+DIM*ONES
[51]  'HERE IS THE MATRIX OF ONES AND ZEROS AS IT STANDS NOW'
[52]  ONES

```

```

[53]  'IS THIS THE MATRIX YOU WANT?'
[54]  W←YESNO(A+[])
[55]  →((W=0),(W=1))/0,G02
[56]  'RE-ENTER YOUR MATRIX OF ONES AND ZEROS'
[57]  →G022
[58]  G019:ONES←DIM×ONES
[59]  'DO YOU WISH TO SEE THE TABLE OF ZEROS AND ONES?'
[60]  W←YESNO(A+[])
[61]  →((W=0),(W=2))/0,G02
[62]  ONES
[63]  →G02
[64]  G011:ONES←DIM×1
[65]  G02:'ENTER MARGINALS TO BE FIT'
[66]  G018:V←[]
[67]  CYCLE:'ANY MORE MARGINALS?'
[68]  W←YESNO(A+[])
[69]  →((W=0),(W=2))/0,G03
[70]  'ENTER MORE MARGINALS'
[71]  V←V,[]
[72]  →CYCLE
[73]  G03:V←V,0
[74]  I←0
[75]  J←0
[76]  G012:J←J+1
[77]  →(J)P V)/G013
[78]  →(V[J]=0)/G012
[79]  VV←V,V[J]
[80]  G014:J←J+1
[81]  →(J)P V)/G016
[82]  →(V[J]=0)/G017
[83]  VV←VV,V[J]
[84]  →G014
[85]  G017:→(V[J-1]=0)/G014
[86]  I←I+1
[87]  A←P VV
[88]  VV←VV,((I×P DIM)-A)P 0
[89]  →G014
[90]  G013:'YOU HAVE NOT ENTERED ANY MARGINALS YET, TRY AGAIN'
[91]  →G018
[92]  G016:V←(I,P DIM)P VV
[93]  (←'YOU HAVE ENTERED '), (←I), ←' MARGINALS,'
[94]  MD←0.01
[95]  MNI←15
[96]  FOLD←ONES
[97]  STEPS←I
[98]  M←1
[99]  'DO YOU WISH TO SET THE PARAMETER VALUES FOR THE ITERATIVE SCHEME?'
[100]  W←YESNO(A+[])
[101]  →((W=0),(W=2))/0,BIGLOOP
[102]  G05:'WHAT IS YOUR VALUE OF MAXIMUM DEVIATION?'
[103]  MD←[]
[104]  'WHAT IS YOUR VALUE OF MAXIMUM NUMBER OF ITERATIONS?'
[105]  MNI←[]
[106]  BIGLOOP:→(M>MNI)/NOCON

```



```

[107] D←0
[108] I←1
[109] LITLOOP:→(I>STEPS)/CHECK
[110] IND←(√(V[I;]ε0))/V[I;]
[111] FOLD←FOLDx(IND SUM SMF)÷((IND SUM FOLD)+(IND SUM FOLD)ε0)
[112] I←I+1
[113] →LITLOOP
[114] CHECK:I←1
[115] CHECKER:→(I>STEPS)/OUT
[116] IND←(√(V[I;]ε0))/V[I;]
[117] C←(1/I,(IND SUM SMF)-(IND SUM FOLD))
[118] →(C<D)/GO10
[119] D←C
[120] GO10:→(C>MD)/STEP
[121] I←I+1
[122] →CHECKER
[123] STEP:M←M+1
[124] →BIGLOOP
[125] OUT:U←((,FOLDε0)^(,SMFε0))
[126] ' '
[127] ' '
[128] (↑'MAXIMUM DEVIATION = '), (↑D), (↑' NO. OF ITERATION = '), ↑M
[129] ' '
[130] ' '
[131] (↑'PEARSON CHI-SQUARE '), ↑+/(((√U)/(,SMF-FOLD))*2)÷((√U)/,FOLD)
[132] T←(SMF×0.5)+((SMF+1)×0.5)-((1+4×FOLD)×0.5)
[133] (↑'FREEMAN-TUKEY CHI-SQUARE '), ↑+/,(T×2)
[134] U←,SMFε0
[135] (↑'LIKELIHOOD RATIO '), ↑2x+/(√U)/,SMF)×((√U)/,SMF)÷((√U)/,FOLD)
[136] (↑'TOTAL NUMBER OF FIXED ZEROS '), ↑(x/DIM)-(+/ONES)
[137] ' '
[138] ' '
[139] 'DO YOU WANT TO SEE THE EXPECTED TABLE?'
[140] W←YESNO(A+□)
[141] →((W=0),(W=2))/0,CONT
[142] ' '
[143] ' '
[144] (L0.5+FOLD×100)÷100
[145] ' '
[146] ' '
[147] 'DO YOU WANT TO SEE THE FREEMAN-TUKEY DEVIATIONS?'
[148] W←YESNO(A+□)
[149] →((W=0),(W=2))/0,GO9
[150] ' '
[151] ' '
[152] (L0.5+T×100)÷100
[153] ' '
[154] ' '
[155] GO9:'DO YOU WANT TO SEE THE RESIDUALS (OBSERVED - EXPECTED)?'
[156] W←YESNO(A+□)
[157] →((W=0),(W=2))/0,GO4
[158] ' '
[159] ' '
[160] (L0.5+(SMF-FOLD)×100)÷100
[161] ' '

```

```

[162] ' '
[163] GO4: 'DO YOU WANT TO SEE THE STANDARDIZED RESIDUALS?'
[164] W←YESNO(A+[])
[165] →((W=0),(W=2))/0,GO8
[166] ' '
[167] ' '
[168] (L0.5+((SMF-FOLD)/(FOLD+FOLD*0)*0.5))*100)+100
[169] ' '
[170] ' '
[171] GO8: 'DO YOU WANT TO SEE THE LOG EXPECTATIONS?'
[172] W←YESNO(A+[])
[173] →((W=0),(W=2))/0,CONT
[174] ' '
[175] ' '
[176] (L0.5+(FOLD+FOLD*0)*100)+100
[177] ' '
[178] ' '
[179] CONT: 'WANT TO TRY SOME OTHER HYPOTHESES?'
[180] W←YESNO(A+[])
[181] →((W=0),(W=1))/0,GO2
[182] 'DO YOU WANT TO ANALYZE ANOTHER TABLE?'
[183] W←YESNO(A+[])
[184] →((W=0),(W=1))/0,BEGIN
[185] →END
[186] NOCON: (↑'NO CONVERGENCE AFTER '),(↑MNI),↑' ITERATIONS.'
[187] (↑'THE DEVIATION AT THIS STAGE IS '),(↑D
[188] 'DO YOU WANT TO DO MORE ITERATIONS OR CHANGE YOUR VALUE'
[189] 'OF MAXIMUM DEVIATION?'
[190] W←YESNO(A+[])
[191] →((W=0),(W=1))/0,GO5
[192] →CONT
[193] END: ' '
[194] ' '
[195] ' '

```

▽  
 \

▽CONTINGENCY [ ]▽  
 ▽ CONTINGENCY

```

[1] 'ENTER OBSERVED CELL FREQUENCIES BY ROWS'
[2] O←10
[3] L1:O←O,[]
[4] →L1X1'Y'=([],[]←'MORE OBSERVED FREQUENCIES TO ENTER?')[1]
[5] 'ENTER NUMBER OF ROWS AND COLUMNS, RESPECTIVELY'
[6] DF←X/(SHAPE+[])-1
[7] ROWTOT←+/MAT←SHAPE,0
[8] N←+/COLTOT←+/MAT
[9] EXP←ROWTOT*,XCOLTOT÷N
[10] CHISQUARE←+/,((MAT-EXP)*2)÷EXP
[11] 'OBSERVED FREQUENCIES ARE: '
[12] MAT
[13] 'EXPECTED FREQUENCIES ARE: '
[14] EXP
[15] (↑'CONTINGENCY CHI-SQUARE WITH '),(↑DF),(↑' DF IS: '),(↑CHISQUARE

```

▽  
 \

```

      ▽TWOWAY [ ] ▽
      ▽ D TWOWAY Y;A;C;K;M;R;S;SIDE;T;TAB;TOP;TV;V
[1] TOP←0,0↑←'ENTER TOP HEADING'
[2] SIDE←0,0↑←'ENTER SIDE HEADING'
[3] T←D[2]+0XS+D[1]+0XR+D[3]+0XC+D[4]
[4] TAB←(S,T)F0XM+1
[5] T1;K←1
[6] V←(Y[R]=M)/1590
[7] T2;TAB[M;K]←+/Y[V;C]=K
[8] →T2X1T1K←K+1
[9] →T1X1S1M←M+1
[10] ''
[11] ' ,TOP
[12] ''
[13] A←F TV←' , (9XT)F'110
[14] TV FMT1T
[15] (A)F'-'
[16] SIDE
[17] TV←8↓TV
[18] K←1
[19] T3; ' |'
[20] (↑' '), (↑K), (↑' |'), ↑TV FMT TAB[K;]
[21] →T3X1S1K←K+1
[22] (A)F'-'
[23] □←□←□←''
[24] (↑'FOR THIS TABLE, TOTAL RESPONSES ARE: '), ↑+/ ,TAB
      ▽
      \

```

```

      ▽YESNO [ ] ▽
      ▽ W←YESNO A
[1] A←4FA, '
[2] →((^/A='STOP'), (^/A='YES '), (^/A='NO '))/G1,G2,G3
[3] 'TYPE YES, NO, OR STOP PLEASE'
[4] A←4F(A←□), '
[5] →2
[6] G1;W←0
[7] →0
[8] G2;W←1
[9] →0
[10] G3;W←2
      ▽
      \

```



```

      7THREEWAY [0]7
      7 W THREEWAY X
[1] DEPTH←8↑0,0↑0←'ENTER DEPTH HEADING'
[2] ROW←8↑0,0↑0←'ENTER ROW HEADING'
[3] COLUMN←0,0↑0←'ENTER COLUMN HEADING'
[4] D←W[1]+0xR←W[2]+0xC←W[3]+0xS←W[4]+0xT←W[5]+0xU←W[6]
[5] TAB←(D,R,C)P0xI←1+0xA←1↑PX
[6] T1:B←1↑PY←X[(X[;S]=I)/\A;]+0xJ←1
[7] T2:Z←Y[(Y[;T]=J)/\B;]+0xK←1
[8] T3:TAB[I;J;K]←+/Z[;U]=K
[9] →T3x\C]K←K+1
[10] →T2x\R]J←J+1
[11] →T1x\D]I←I+1
[12] 0←0←''
[13] ' , COLUMN
[14] ''
[15] E←P TV←' , (9xC)P'110
[16] TV FMT\C+0xI←1
[17] (E)P'_'
[18] DEPTH,ROW
[19] (E)P'_'
[20] TV←18↓TV
[21] T4:J←1
[22] T5:' |'
[23] (↑' 111 111 |' FMT(I,J)),↑TV FMT TAB[I;J;]
[24] →T5x\R]J←J+1
[25] →T4x\D]I←I+1
[26] (E)P'_'
[27] 0←0←0←''
[28] (↑'FOR THIS TABLE, TOTAL RESPONSES ARE: '),↑+/ , TAB
      7

```

## Appendix V.

A Quantitative Method to Test for Similarity  
between Photo Interpreters

Russell G. Congalton, Research Assistant  
and  
Roy A. Mead, Assistant Professor  
School of Forestry and Wildlife Resources  
Virginia Polytechnic Institute and State University  
Blacksburg, Virginia 24061

## Biographical Sketch

Mr. Congalton is a native of New Jersey and received his B.S. in Forestry at Rutgers University. He is presently a graduate research assistant at VPI and SU. Dr. Mead received his B.S. in Botany at Northern Arizona State University, his M.S. in Remote Sensing at Colorado State University, and his Ph.D. in Remote Sensing at the University of Minnesota.

## Abstract

A method has been developed to quantitatively test the degree of similarity between photo interpreters. This method involves giving each photo interpreter the same set of photos to interpret. An error matrix is then generated for each interpreter by comparing his interpretation to the actual ground cover. This error matrix is then analyzed using a computer program called KAPPA. This program uses discrete multivariate analysis techniques to determine if one error matrix (i.e., photo

interpreter) is significantly different from another. The program can be altered to test for similarity at different confidence levels. Not only does this technique allow one to compare two separate interpreters, but it also allows one to test whether an individual photo interpreter is consistent through time.

### Introduction

Photo interpretation is the art and science of identifying objects and deducing their significance on aerial photos. Good, consistent photo interpretation depends upon the experience and skill of the individual who delineates the boundaries between vegetation/land cover types over the landscape. The judgment involved is generally qualitative in nature, and therefore difficult to evaluate or compare with interpretations made by others. Usually the interpreter has intuitive feelings about how well he is doing, but is unable to support these feelings with any specific tests. This paper suggests a way of quantifying photo interpretation results and gives a statistical method for comparing these results.

The procedure proposed in this paper can test for the degree of similarity between interpreters, or test the consistency of the same interpreter over time. Testing to see if interpreters are similar is useful when more than one interpreter is to work on a project. If it can be determined that the delineations made by all interpreters are not



significantly different, then the project will yield uniform results for all interpreters. Also, it would be useful to test the same interpreter over a period of time to check for changes in his interpretation. It may also be important to determine if varying types of photography (film/filter combinations), or seasons of photography result in significantly different delineations. By placing a grid over each delineation, the individual cells are assigned to the land cover/vegetation type which represents the majority of the cell. Each cell is then compared one-by-one with the corresponding cell (i.e., in the same location) from another delineation. If one of the delineations is assumed to be correct (reference data), then comparison of the two sets of spatially defined cells yields a measure of "photointerpretation accuracy". This is usually expressed in the form of an error matrix.

#### Procedure

An error matrix is a square array of numbers set out in rows and columns which expresses the number of cells assigned by the photo interpreter to a particular land cover type relative to the actual land cover (reference data). The columns represent the reference data and the rows indicate the photo interpreter assigned land cover type (Figure 1).

		Reference Data		
		A	B	C
Photo Interpretation	A			
	B			
	C			

Figure 1. Error matrix format for three land cover types.

The numbers in the error matrix are tallies compiled by comparing the photo interpretation with the actual cover type (reference data) on a cell by cell basis. All correct classifications are located on the major diagonal of the error matrix.

The specific method used to generate an error matrix is dependent on what information is needed. If the degree of similarity between two or more photo interpreters is to be determined, each interpreter is given the same aerial photographs to interpret. An error matrix is then tabulated for each interpreter by comparing his interpretation with a reference data set (correct delineation). If the test involves determining the consistency over time for a single interpreter, then a representative part of a selected stereo pair is interpreted at the beginning of a project. At some later date the remainder of the photos are interpreted and then the two error matrices (Time A and Time B) are compared. Finally, if it is desired to measure the accuracy of delineations made on different types of photography, a separate interpretation is performed on the same area for each set of photos by each interpreter and an error matrix is generated.

Once the error matrices are generated, a discrete multivariate analysis procedure (Bishop et al., 1975) is used to test the degree of similarity between the error matrices. This test is based on a maximum likelihood estimate of the multinomial distribution (Equation 1).

$$K = \frac{\sum_{i=1}^{\Gamma} X_{ii} - \frac{\sum_{i=1}^{\Gamma} (X_{i+} * X_{+i})}{\Gamma}}{N^2 - \sum_{i=1}^{\Gamma} (X_{i+} * X_{+i})} \quad (1)$$

where:

$K$  = # of rows in matrix

$X_{ii}$  = # of obs in row  $i$  and col.  $i$

$X_{i+}$  = marginal total of row  $i$

$X_{+i}$  = marginal total of col.  $i$

$N$  = total # of observations

This equation yields a value KHAT which is a measure of the actual agreement minus the chance agreement. A confidence interval at a given  $\alpha$ -level is then placed around the value of KHAT calculated for each error matrix. If the confidence interval for one error matrix overlaps the confidence interval for another error matrix, the two matrices are said to be not significantly different at that  $\alpha$ -level. However, if no overlap of the KHAT confidence intervals occurs, then the matrices are said to be significantly different at that  $\alpha$ -level.

This entire comparison process can be performed using a FORTRAN computer program called KAPPA. Given the error matrices to be analyzed, the program calculates a KHAT value and a confidence interval for each error matrix. The program then prints out which error matrices are significantly different and which are not.



The data used in this study were taken from Lauer et al. (1970). Five photo interpreters interpreted the same aerial photographs of Yosemite Valley, California, and their individual error matrices were generated. Also, five film and filter combinations were used with a single interpreter, and error matrices were generated.

#### Results and Discussion

All five of the interpreters tested on the photos from Yosemite Valley produced significantly different delineations (Table 1). The confidence interval was calculated at the 95% level.

Table 1. Summary table for five interpreters of Yosemite Valley photos.

<u>Interpreter</u>	<u>Lower Limit</u>	<u>KHAT</u>	<u>Upper Limit</u>
1	0.31167	0.31991	0.32815
2	0.28623	0.29420	0.30216
3	0.36677	0.37485	0.38293
4	0.23115	0.24156	0.25197
5	0.20878	0.21925	0.22972

The results of the five different film and filter combinations are presented in Table 2. These results were also calculated at the 95% confidence level.

Table 2. Summary table for the five film and filter combinations.

<u>Film/Filter</u>	<u>Lower Limit</u>	<u>KHAT</u>	<u>Upper Limit</u>
IR-301/W25	0.31167	0.31991	0.32815
IR/W89B	0.29615	0.30436	0.31258
Ektta Aero IR	0.11318	0.12071	0.12825
Enhancement X	0.25427	0.26163	0.26898
Enhancement Y	0.36704	0.37438	0.38173

As can be seen from Table 2, the interval for IR-301/W25 overlaps with the interval for IR/W89B. Therefore, these two interpretations are not significantly different. All the other interpretations are significantly different.

#### Summary

The examples given in this paper indicate how photo interpretation results can be quantified using error matrices. These error matrices can then be compared using a discrete multivariate analysis procedure and conclusions made.

#### Literature Cited

Bishop, Y., S. Fienberg and P. Holland 1975, Discrete Multivariate Analysis: Theory and Practice. MIT Press: Cambridge, Massachusetts, 575 pp.

Lauer, Donald T., Claire M. Hay and Andrew S. Benson 1970, Quantitative Evaluation of Multiband Photographic Techniques.  
Final Report for Earth Observation Division Manned Spacecraft Center, NASA, Contract No. NAS 9-9577, 110 pp.



**DRAFT**

## Appendix VI.

**Landsat Classification Accuracy Assessment Procedures:****An Account of a National Working Conference**

by

Roy A. Mead  
School of Forestry and Wildlife Resources  
Virginia Polytechnic Institute and State University  
Blacksburg, VA 24061

and

John Szajgin  
Technicolor Graphic Services, Inc.  
EROS Data Center  
Sioux Falls, SD 57198

## ABSTRACT

A working conference was held in Sioux Falls, South Dakota November 12, 13, and 14, 1980 dealing with Landsat Classification Accuracy Assessment Procedures. Thirteen formal presentations were made on three general topics: (1) sampling procedures, (2) statistical analysis techniques, and (3) examples of projects which included accuracy assessment and the associated costs, logistical problems and value of the accuracy data to the remote sensing specialist and the resource manager. Nearly twenty conference attendees participated in two discussion sessions addressing various issues associated with accuracy assessment. This paper presents an account of the accomplishments of the conference.

## INTRODUCTION

In the years since Landsat imagery first became available, an untold number of Landsat scenes have been digitally analyzed to classify land cover. These classifications are not without error, and have been subject to closer scrutiny by critics and potential users than similar products developed by more traditional methods. A few potential users of Landsat data were discouraged by the unfulfilled expectations spirited by the results of early investigations. This has recently led researchers and government agencies to proceed cautiously with technology transfer. Thus, scientists have been keenly aware of the need to assess the accuracy of Landsat classifications before distributing the products to users.

Topographic mapping procedures include routine evaluations for compliance with well defined accuracy standards and the accuracy attainable under specific conditions (terrain characteristics and mapping equipment used) are well known. This capability is the result of many directed research efforts. However, techniques for assessing the accuracy of Landsat classifications have developed in an ad hoc manner. Many such methods are not statistically sound and can yield biased estimates of accuracy.

For example, researchers used the limited available ground information (i.e., maps, photo interpretations or less often actual visits to the field) collected for development of training statistics to estimate classification accuracy. This can result in over optimistic estimates of classification performance, particularly when the training data does not adequately describe the scene variability. Windshield surveys, in which a few easily accessible areas are visited on the ground, are another biased approach to accuracy assessment. In addition, biases can also be introduced by using a different



classification framework for accuracy assessment than that used in developing the digital classification.

The trend, more recently, has been to sample the classifications and measure the degree of agreement with a set of spatially defined reference data (i.e., ground truth). Analysis of the resulting accuracy data can guide researchers in scene selection (season, etc.), and determine the most appropriate methods of classification for particular applications.

The importance of assessing classification accuracy, the lack of any standard procedures, and the limited number of reports in technically reviewed journals, justified the conference discussed in this paper. Only a relatively small number of researchers have worked in the subject area to any great extent. Therefore, attendance at the conference was limited and by invitation only. The specific objectives of the conference were:

1. To determine the state-of-the-art for accuracy assessment procedures.
2. To provide a forum for exchange of ideas concerning accuracy assessment procedures.
3. To identify research needs and recommend the approach that should be taken to improve accuracy assessment procedures.

#### CONFERENCE THEMES

A comprehensive proceedings of the formal conference presentations is planned. However, it is worthwhile to identify and summarize the major themes that developed from the conference in general.

Accuracy is a measure of the amount of agreement between two data sets. Typically this is a thematic map in question and a reference data set often thought of as "ground truth." However, when this procedure is generalized other applications become apparent, including change detection analysis for

the monitoring of particular resources. Furthermore, sequential appraisal of a classification can result in better end results.

There are several types of accuracy and it is important to identify which is being utilized. Two major categories of accuracy are site specific and non site specific accuracy. Non site specific accuracy compares tabular summaries of the proportions of the area mapped into each of the categories. Site specific accuracy utilizes the spatial nature of the data. That is, two spatially defined data sets are registered and compared for the amount of agreement. This can be a polygon, grid cell, or point comparison. In this case, the difference between the two data sets results in a spatially defined binary data set. This represents the population we are sampling for the parameters in question.

An error matrix or contingency analysis approach to accuracy assessment is still another method of comparison of the two data sets. This requires a site specific (spatially defined) approach.

Furthermore, many factors affect the validity of an accuracy assessment. The quantity and quality of ground truth depend upon the methods used for sample size determination and data acquisition. In light of this, it becomes apparent that the term "ground truth" is ill defined. What is "ground truth" with regard to parameters such as percent of ground cover? Can this ever really be measured? For many cover types, this parameter can be estimated more accurately on aerial photographs than by ground procedures.

Finally, one should not lose track of the difference between the usefulness of a specific product and its estimated accuracy. A numerical report of product accuracy may say nothing of how much use the product gets or how well it compares with what was previously available. A quantitative accuracy assessment results in a numerical summary which may or may not represent the

usefulness of the product. In many instances, a classification of low or intermediate accuracy is a welcome and useful product.

The desired information as well as the nature of the scene which was classified, determine which is the most appropriate means of assessing accuracy. Certainly, different landscapes may need to be sampled differently for best results. Therefore, studies should be done to look at the sensitivity of accuracy estimates when different sampling procedures are used. In comparing and assessing sampling procedures for accuracy assessment, not only is statistical variability to be considered, but also the spatial diversity of the data. Furthermore, all of these considerations interact to determine the most appropriate sampling and estimation procedure to use. Much work remains to be done, utilizing designed experiments with specific hypotheses, to identify the relative reliability of various sampling procedures.

Assessing and reporting, by some standard means, the accuracy of a thematic classification will become more vital as these products become a part of geographic information systems. This will be necessary to insure high quality output products and well informed management decisions.

The use of training data for accuracy assessment results in a somewhat biased but possible useful estimate of overall accuracy. The nature of the bias is to overestimate accuracy. The amount of bias depends upon how well the training data represent the variability present in the scene. In some instances, such an approach will be adequate. However, for close scrutiny and for within class estimates of accuracy, and independent accuracy assessment is warranted. An approach to minimizing the cost of an independent accuracy assessment is to collect accuracy assessment data at the time the training data is collected. This data should be earmarked for later use and not used in the training process.



Although this conference did much to establish communication among researchers utilizing accuracy assessment procedures, much work remains to be done in summarizing what procedures are most commonly utilized. In addition, a bibliography of the literature and available computer programs should be compiled and published. A survey of researchers in the field will help to define how well they can map various cover types. This will assist in developing a set of mapping standards. Although accuracy requirements may vary among cover types, acceptable map accuracy standards are needed to match intended uses. Standards such as "second order at level II" can help in minimizing subjective evaluations and finally, perhaps many classifications are more accurate than we think due to geometry problems and edge pixels. It becomes apparent that classification error and mapping error are not one in the same. Much work needs to be done to discriminate between the two sources of error.

## SUMMARY

Many issues were discussed and debated by the participants. Topics for further research were identified and major themes summarized in this paper.

The participants recommended that a working group be established to write a "manual" or "guide book" on accuracy assessment procedures. Possibly this group could be formed as an ad hoc committee within the American Society of Photogrammetry and seek funding to prepare the document described above. Plans are now being made to do this.

The conference succeeded in accomplishing the three objectives stated earlier. A comprehensive proceedings is planned which will represent state-of-the-art accuracy assessment procedures.

## Appendix VII. Listing of FORTRAN Computer Program WILDINT2

```

$JUN  MATFIV, NOLIST
REAL  JX(20), YI(20,20), INDEX(20,50), JOX(20,50), SO(20,50)
REAL  LONCLS(11), HICLAS(11)
CHARACTER*1 CHAR(11)
CHARACTER*4 NAME(20,5)
INTEGER DATA(20,50), A, B, X, A1, B1, CTYPE(20)
READ(5,*) NR, NC
NR2=NR-1
NC2=NC-1

C
C   READ THE DATA MATRIX
C
DO 10 I=1, NR
  READ(5,500) CONTACT(I,J), J=1, NC
10 CONTINUE

C
C   THIS LOOP IS TO SET THE EDGES OF THE OUTPUT
C   MATRICES TO ZERO.
C
DO 5 J=1, NC
  SO(1,J)=SO(NR,J)=0
  JOX(1,J)=JOX(NR,J)=0
  INDEX(1,J)=INDEX(NR,J)=0
5 CONTINUE
DO 6 I=2, NR2
  SO(I,1)=SO(I,NC)=0
  JOX(I,1)=JOX(I,NC)=0
  INDEX(I,1)=INDEX(I,NC)=0
6 CONTINUE
8 PRINT, 'HOW MANY CLASSES DO YOU WANT IN YOUR OUTPUT?'
PRINT, '(I.E. LOW, MEDIUM, HIGH = 3 CLASSES)'
READ(5,*,END=91,ERR=91) HICLAS
91 DO 9 I=1, HICLAS
  WRITE(6,501) I
  READ(5,*,END=92,ERR=92) LONCLS(I), HICLAS(I)
92 PRINT, 'WHAT SYMBOL DO YOU WANT TO REPRESENT THIS CLASS?'
  READ(5,*,END=9,ERR=9) CHAR(I)
9 CONTINUE
PRINT, 'WHAT INDICES DO YOU WANT TO CALCULATE?'
PRINT, 'TYPE 1 FOR INTERSPERSION ONLY,'
PRINT, 'TYPE 2 FOR Juxtaposition ONLY,'
PRINT, 'TYPE 3 FOR ALL INDICES.'
READ(5,*,END=91,ERR=91) NCHECK
91 DO CASE NCHECK
  DO 10 TO 11
    CASE
      DO 10 TO 50
    CASE
      DO 10 TO 11
    END CASE
C
C   CALCULATION OF SPATIAL DIVERSITY INDICES
C
11 DO 25 I=2, NR2
  DO 25 J=2, NC2

```



```

SD1=0
I=J=1
A=1-1
I1=I+2
A1=A+2
DO 20 K=A,A1
DO 20 L=I,I1
IF (DATA(K,L).EQ.0.OR.DATA(K,L).EQ.DATA(1,J)) GO TO 20
SD1=SD1+1
20 CONTINUE
SD(1,J)=SD1/N
25 CONTINUE
CALL UJFUT(SD,1,IR,IC,LORCLS,ICLAS,NCLASS,CHAR)
IF (NCHECK.EQ.1) GO TO 1000

READ MATRIX OF WEIGHTING VALUES

30 READ(S,*) NTYPES
DO 50 K=1,NTYPES
READ(S,SUS) CTYPE(K), (NAME(K,II),II=1,3)
50 CONTINUE
PRINT, 'WHAT ARE THE WEIGHTING FACTORS FOR:'
N12=NTYPES-1
DO 60 K=1,N12
K1=K+1
DO 60 L=K1,NTYPES
WRITE(9,504) (NAME(K,JJ),JJ=1,3), (NAME(L,II),II=1,3)
WRITE(9,504) (NAME(K,SJ),SJ=1,3), (NAME(L,LI),LI=1,3)
READ(S,*) EN1=54, EN2=59) W(K,L)
59 W(K,K)=0.0
W(L,L)=0.0
W(L,K)=W(K,L)
WRITE(9,505) CTYPE(K), CTYPE(L), (NAME(K,K1),K1=1,3), (NAME(L,L2),L2=
*1,3), W(K,L)
60 CONTINUE

CALCULATION OF Juxtaposition Indices

DO 100 I=2,IR2
DO 100 J=2,IC2
X=DATA(I,J)
JUX(I,J)=0
IF (X.EQ.0.OR.DATA(I-1,J-1).EQ.0) GO TO 40
IF (DATA(I-1,J).EQ.0.OR.DATA(I-1,J+1).EQ.0) GO TO 40
IF (DATA(I,J-1).EQ.0.OR.DATA(I,J+1).EQ.0) GO TO 40
IF (DATA(I+1,J-1).EQ.0.OR.DATA(I+1,J).EQ.0) GO TO 40
IF (DATA(I+1,J+1).EQ.0) GO TO 60
JX(1)=1.0*W(X,DATA(I-1,J-1))
JX(2)=2.0*W(X,DATA(I-1,J))
JX(3)=1.0*W(X,DATA(I-1,J+1))
JX(4)=2.0*W(X,DATA(I,J-1))
JX(5)=2.0*W(X,DATA(I,J+1))
JX(6)=1.0*W(X,DATA(I+1,J-1))
JX(7)=2.0*W(X,DATA(I+1,J))
JX(8)=1.0*W(X,DATA(I+1,J+1))

```

```

      DO 70 N=1,8
70  JUX(I,J)=JUX(I,J)+JX(N)/12
      GO TO 100
80  JUX(I,J)=0
100  CONTINUE
      CALL OUTPUT(JUX,2,NR,NC,LONCLS,HICLAS,NCLASS,CHAR)
      IF(NCHECK.EQ.2) GO TO 1000

      CALCULATE THE TOTAL WILDLIFE IMPORTANCE INDEX

      PRINT,'WHAT IS THE WEIGHT FOR THE INTERSPERSION INDEX (0<N<1)?'
      READ(S,*,END=111,ERR=111) ALPHA
111  PRINT,'WHAT IS THE WEIGHT FOR THE JUXTAPOSITION INDEX (0<N<1)?'
      READ(S,*,END=112,ERR=112) BETA
112  ALPHA=ALPHA/(ALPHA+BETA)
      BETA=BETA/(ALPHA+BETA)
      DO 200 I=1,NR2
      DO 200 J=1,NC2
      INDEX(I,J)=ALPHA*SO(I,J)+BETA*JUX(I,J)
200  CONTINUE
      CALL OUTPUT(INDEX,3,NR,NC,LONCLS,HICLAS,NCLASS,CHAR)
1000  PRINT,'DO YOU WANT TO GO AGAIN (1 IF YES,0 IF NO)?'
      READ(S,*)NTRY
      IF(NTRY.EQ.1) THEN GO
      PRINT,'IF YOU WANT TO USE THE SAME INPUTS AS LAST'
      PRINT,'TIME, HIT RETURN AFTER BEING PROMPTED.'
      GO TO 8
      END IF
      STOP
500  FORMAT(30I2)
501  FORMAT(1X,'WHAT ARE THE LOWER AND UPPER BOUNDS FOR CLASS ',I2)
503  FORMAT(12,3X,3A4)
504  FORMAT(2X,3A4,'-',3A4,' EDGE?')
505  FORMAT(2X,12,2X,12,3X,3A4,'-',3A4,' EDGE IS WEIGHTED ',F3.1)
510  FORMAT(4I)
      END

      THIS SUBROUTINE PRINTS A MATRIX OF CHARACTERS
      REPRESENTING THE VALUES OF WHATEVER INDEX IS
      SENT AS AN ARRAY. THE OUTPUT WILL BE IN
      FILE FT09F001.

      SUBROUTINE OUTPUT(ARRAY,N,NR,NC,LONLIM,UPPLIM,NCLASS,CHAR)
      REAL ARRAY(20,30),LONLIM(11),UPPLIM(11)
      CHARACTER CHAR*(11),TYPE(20,30)
      DO CASE N
      WRITE(6,102)
      WRITE(9,102)
      CASE
      WRITE(6,103)
      WRITE(9,103)
      CASE
      WRITE(6,104)

```

```

WRITE(9,104)
END CASE
DO 15 J1=1,NR
DO 15 J2=1,NC
DO 15 J3=1,NCLASS
IF (ARRAY(J1,J2).GE.LOWLIM(J3).AND.ARRAY(J1,J2).LE.UPPLIM(J3)) TYPE
*(J1,J2)=CHAR(J3)
15 CONTINUE
DO 20 KK=1,NR
WRITE(6,110) (TYPE(KK,LL),LL=1,NC)
WRITE(9,110) (TYPE(KK,LL),LL=1,NC)
20 CONTINUE
103 FORMAT(1X,'JUXTAPOSITION MAP')
102 FORMAT(1X,'INTERSPERSION MAP')
104 FORMAT(1X,'TOTAL INDEX MAP')
110 FORMAT(1X,50A1)
RETURN
END
SENTRY

```



## Appendix VIII.

## A COMPUTERIZED SPATIAL ANALYSIS SYSTEM

## FOR ASSESSING WILDLIFE HABITAT

## FROM VEGETATION MAPS\*

Roy A. Mead, Terry L. Sharik,

Stephen P. Prisley, and Joel T. Heinen

Department of Forestry

Virginia Polytechnic Institute

and State University

Blacksburg, VA 24061

## ABSTRACT

Vegetation and land cover patterns affect the quality of habitat available for wildlife. Given the degree of interspersions of cover types and relative value of each edge type and the importance of spatial diversity, an index of habitat spatial diversity can be computed for each parcel of land (of any desired size) relative to each wildlife species or group of species. This is accomplished by defining a grid which is either placed on a land cover map or on an aerial photograph. Each cell is then coded on the basis of (its predominant) cover type. A computer program subsequently analyzes the arrangement of these coded cells and produces maps of (a) interspersions,

---

\* Presented at 47th Annual Meeting American Society Photogrammetry, Washington, D.C., February 23-26, 1981.

(b) juxtaposition, and (c) spatial diversity. Separate multicolor maps can be made for any wildlife habitat of interest using a digital film recorder. These map overlays can be used by the resource manager to compare wildlife habitat quality and potential with maps for forest, range, watershed and recreation potential.

#### INTRODUCTION

There is a tremendous need to develop quantitative methods to assess wildlife habitat. This was specifically mandated by the Resources Planning Act, as well as other legislation. Wildlife habitat must be considered in all management plans together with timber, range, recreation and watershed. While timber inventories have been conducted for many years, techniques for quantifying the wildlife habitat still need to be developed.

The technology of remote sensing has provided the means for mapping land cover/vegetation over very large areas for wildlife habitat management (Pengelly, 1978). However, the maps themselves only partially fulfill the inventory data needed by biologists who must manage for wildlife. The maps must be analyzed and interpreted to enhance the various characteristics of the landscape which have a bearing on management decisions. In short, the standard land cover map is a source of information that may be helpful in making management decisions.

This paper suggests a means to analyze and interpret maps of land cover to produce spatially defined data that will be valuable information for managing wildlife habitat. Emphasis is on the technique and not on the controversial issue of defining habitat QUALITY. It must be understood that the landscape characteristics important in habitat evaluation vary

according to region and the specific wildlife species of interest. The various weighting factors discussed in this paper must be determined by wildlife managers familiar with local conditions or from agency handbooks which give the habitat requirements and preferences for many species.

The specific objective of this study was to develop a computerized system for measuring the spatial diversity component of wildlife habitat from vegetation maps.

#### Study Area and Input Data

The area used as an example for testing the wildlife habitat analysis techniques described in this paper was the Great Dismal Swamp. The area is managed by the U. S. Fish and Wildlife Service as a game refuge and includes approximately 84,900 hectares. This wetland was thoroughly described by Garrett and Carter (1977). The area was ideal for evaluating the proposed habitat analysis techniques for three reasons. First, the Dismal Swamp "contains a remarkable diversity of vegetative communities" (Garrett and Carter, 1977). Second, the area had recently been mapped (Gammon and Carter, 1979). Third, the local resource managers were available for assistance in evaluating the validity/usefulness of the final habitat quality maps that were produced.

The vegetation maps produced by Gammon and Carter (1979) contained "43 separate canopy designations and 243 specific vegetative communities...".



This map was overlaid with a square grid system oriented in a North-South manner. Each cell contained 22 hectares and formed a matrix of 93 rows and 42 columns. Each cell was given a community designation according to the cover type which occupied the most area within that cell. This information was stored on discs for analysis by the computer.

Given the above data the following procedure was used to assess wildlife habitat diversity for the Great Dismal Swamp. Since the primary thrust of this paper is to present a proposed technique, all additional inputs (e.g., juxtaposition weighting factors and restrictive factors) are purely hypothetical, as is the selected wildlife species "A".

#### HABITAT ASSESSMENT PROCEDURE

There are four components that form the package of techniques used for assessing wildlife habitat:

1. Input data
2. Measurement of interspersion
3. Measurement of juxtaposition
4. Recognition of exclusion factors

Basically, the four components interact in the following way. Suitable land cover/vegetation maps are either obtained from existing sources or compiled. The necessary vegetation categories, map scales and minimum mapping unit size may vary from region to region and with the species for which potential habitat is being assessed. The habitat criteria for the species of interest must be known (or estimated). Such criteria include the relative desirability (i.e., the weighting factors) of various

vegetation/land cover edges and the animals' preference for various vegetation distribution patterns. Classification of vegetation groups is sometimes rather arbitrary (Pielou, 1977), and must be made biologically in terms of the requirements of the organisms involved. For example, the timber type classification system used by forest industries may not always be adequate for use in wildlife habitat inventories. Finally, specific restrictive factors or resources (e.g., water) that either MUST or MUST NOT be present for suitable habitat need to be known.

A spatial diversity "SD" index value is computed for each parcel of land (cell) (of any desired size) relative to each wildlife species or groups of species. The index is a function of "IS," interspersion, "JX," juxtaposition, and any number of restrictive factors.

$$SD_A = \left[ \left( \sigma_A \frac{\text{"IS"}}{8} \right) + \left( \alpha_A \frac{\text{"JX"}}{12} \right) \right] \left[ \frac{1}{A} \times \frac{2}{A} \times \frac{3}{A} \right]$$

where:

A → indicates a specific wildlife species or group of species (B, C, D, etc. for others).

$\sigma$  → indicates the relative importance of interspersion to juxtaposition for wildlife species A, B, or C, etc.

$\alpha$  → indicates the relative importance of juxtaposition to interspersion for wildlife species A, B, or C, etc.

Note that  $\sigma$  and  $\alpha$  can range between 0 and +1 but must sum to 1.00. A low value would indicate a very undesirable or unimportant characteristic and +1 as very desirable or important characteristic. Scaling will have to be worked out and a sensitivity analysis performed.

$\boxed{1}_A \rightarrow$  indicates a restrictive factor that is essential for wildlife species group "A".

An example of a restrictive factor might be the presence of water within one mile. If this is present (i.e., satisfying a necessity), then  $\boxed{1}_A$  is given a value of 1 and has no impact on the value of  $IS_A$ .

However, if there is no water (an absolute necessity), then  $\boxed{1}_A$  is assigned a value of "0" and automatically makes  $IS_A = 0$ . In some cases the restrictive factors may be set at intermediate values indicating undesirable conditions but not total exclusion. Values for  $IS_A$  referring to "high," "medium," and "low" have to be determined (categorized).

#### Necessary Land Cover Data

It is assumed that a suitable vegetation map is available which includes the necessary categories of overstory and/or understory communities indicated. This must be determined for each wildlife species for which habitat is to be assessed.

A. Small Area, Manual Analysis. A grid drawn on clear plastic material is placed directly over the vegetation map. The predominant vegetation category in each cell of the grid is determined, and coded directly on the clear plastic using a grease pencil. A key will be needed to relate the



letters or symbols used to the vegetation categories.

B. Large Area, Computerized Analysis. A vegetation map in polygon form is digitized (or manually coded) at any desired cell size. Individual cells are categorized and a file created to store the resulting data.

### Measurement of Interspersion

A. Small Area, Manual Analysis. The vegetation category predominant in each individual cell on the clear plastic grid is compared to each of the immediately adjacent cells. The number of adjacent cells of another vegetation type are counted and that number recorded on the plastic sheet in the lower right hand corner.

Consider the following two examples:

Example I

B	A	A
B	A <sub>3</sub>	A
B	A	A

IS = 3

Example II

B	C	B
A	A <sub>7</sub>	B
B	C	C

IS = 7

The center cell in example 1 has 3 adjacent cells with dissimilar predominant vegetation types. Therefore, the value for interspersion is 3. In the second example, the IS value is 7. It is clear that the land cover patterns are much more intermixed in example II. Those cells with IS values of 7 or 8 could be printed light gray, values of 3-6, intermediate gray, and 0-2 as dark gray. Note that each cell in the entire matrix

becomes the centroid cell for comparison with adjacent cells. Thus, a map of interspersions is produced from all of the "IS" values computed by moving the 3 x 3 matrix throughout the data set.

B. Large Area, Computerized Analysis. A computer could easily be programmed to compare adjacent cells and create a file with the interspersions values for each cell. Any range of IS values could be assigned a specific color or gray tone, and thus an interspersions map could be made.

#### Measure of Juxtaposition

Wildlife biologists know that certain types of vegetation edges are very important for specific wildlife species. Abundance of these species may be considered a consequence of edges where types of food and cover come together (Leopold, 1936). According to Odum (1971), the edge effect may be defined as the tendency for an increase in variety and density of organisms at community junctions. This effect is most marked in animals with relatively low mobility (Leopold, 1936) and high requirements in terms of diversity of vegetative communities (Leopold, 1936). Various edge combinations can be assigned a relative weighting factor for each group of wildlife, e.g.,

A/B	.60
A/C	.30
B/C	.10

In this case the relative value of an A/B edge is twice that of an A/C edge for a particular wildlife species. Therefore, a measure of juxtaposition can be easily computed by summing the various quantity-quality

products for all edges relative to each centroid cell in the data matrix.

Considering example matrices I and II again:

Example I

B	A	A
B	A	A
B	A	A

Edge	Quantity	Quality	Total
Type			
A/B	4*	.60	2.40
A/C	0	.30	0.00
B/C	0	.10	<u>0.00</u>
JX Index =			2.40

Example II

B	C	B
A	A	B
B	C	C

Edge	Quantity	Quality	Total
Type			
A/B	5	.60	3.00
A/C	5*	.30	1.50
B/C	0	.10	<u>0.00</u>
JX Index =			4.50

The JX value for example I is 2.40 and 4.50 in example II which has more edges which are of importance to the wildlife species under consideration.

\* Note that diagonal edges only count 1 while either vertical or horizontal edges count as 2.



## RESULTS AND DISCUSSION

A portion of the original coded vegetation map and the resulting maps for interspersions, juxtaposition and spatial diversity are shown in Figures 1, 2, 3, and 4. The area shown includes 20 rows and 28 columns of the coded input data. The numbers in Figure 1 correspond to coefficients which were arbitrarily assigned to the various vegetation categories mapped by Gammon and Carter (1979).

The dark, intermediate gray and light areas in Figure 2 represent low, medium and high interspersions, respectively. These correspond to the following ranges for the "IS" calculation, respectively:

0 to .3  
>.3 to .6  
>.6 to 1.0

The designations of dark, intermediate gray, and light in Figure 3 show juxtaposition and correspond to these ranges for the "JX" calculation, respectively:

0 to .3  
>.3 to .6  
>.6 to 1.0

Finally, the spatial diversity index "SD" was categorized in an identical way. The resulting map is shown in Figure 4.

## SUMMARY AND CONCLUSIONS

A wildlife habitat diversity map was produced for a hypothetical wildlife species in the Dismal Swamp utilizing a vegetation cover map. This method can be performed very quickly by computer over large areas, given the necessary input data. Maps of interspersions and juxtaposition can be produced as well by assigning printer symbols to arbitrarily designated categories for each of the three parameters (interspersions, juxtaposition and the wildlife habitat diversity index). Such maps are repeatable and would be consistent over large areas. The most crucial part of the operation is the assignment of the weighting factors from "known" ecological information about each wildlife species. The computerized methodology may have tremendous potential when implemented with remotely sensed digital data for land/cover vegetation.

Further work is needed to determine the sensitivity of the output maps to changes in the weighting factors for various species of wildlife. The relation between animal home range and suitable cell size must also be examined. More efficient methods should be used to digitize the land cover/vegetation maps. Finally, the maps must be more thoroughly evaluated by field resource managers and wildlife habitat specialists.

The method proposed here measures only the spatial diversity of the landscape. Such a measure, and the maps which result, could be incorporated into a larger, more comprehensive system for assessing wildlife habitat quality.

## LITERATURE CITED

- Gammon, P. and V. Carter. 1979. Vegetation mapping with seasonal color intrared photographs. Photogrammetric Engineering and Remote Sensing 45:87-97.
- Garrett, M. K. and V. Carter. 1977. Contribution of remote sensing to habitat evaluation and management in a highly altered ecosystem. In: Proceedings of Forty-second North American Wildlife Conference, p. 56-65.
- Leopold, A. 1936. Game management. Charles Scribner's Sons, N. Y., 481 p.
- Odum, E. P. 1971. Fundamentals of ecology. W. B. Saunders Company, Philadelphia, London, Toronto, p. 157.
- Pengelly, W. 1978. The changing history of wildlife management. In Proceedings of PECORA IV, Symposium on Application of Remote Sensing Date to Wildlife Management. Oct. 10-12, Sioux Falls, SD, p. 1-5.
- Pielou, E. C. 1977. Mathematical ecology. John Wiley and Sons, New York, 385 p.



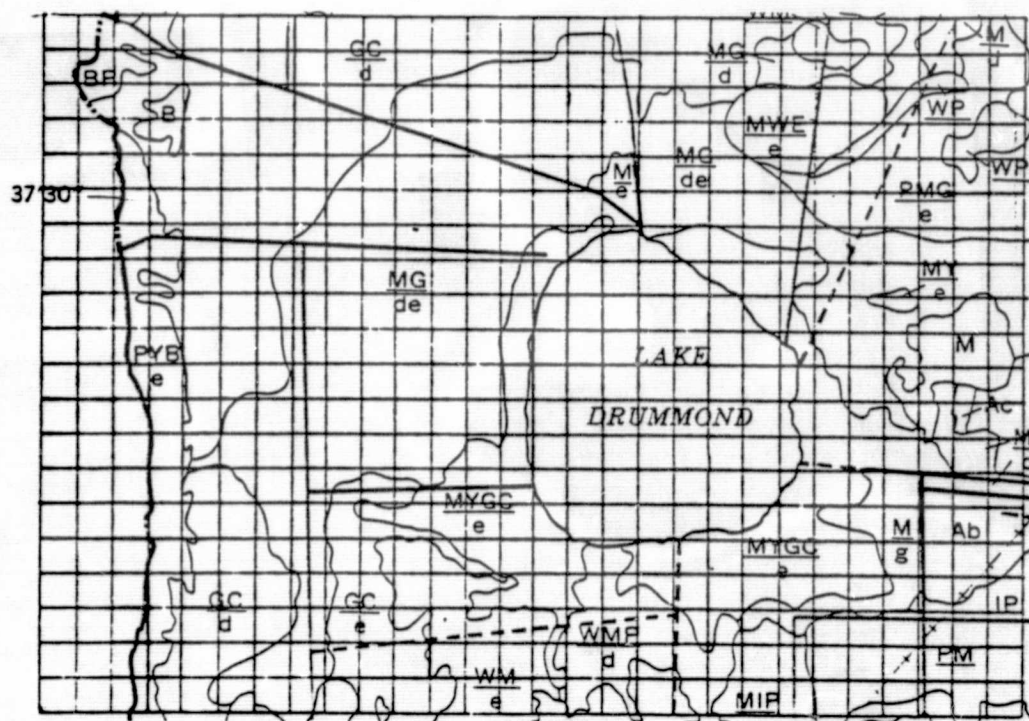


Figure 1. Digitally coded cellular map of a small portion of the Dismal Swamp vegetation/land cover map.

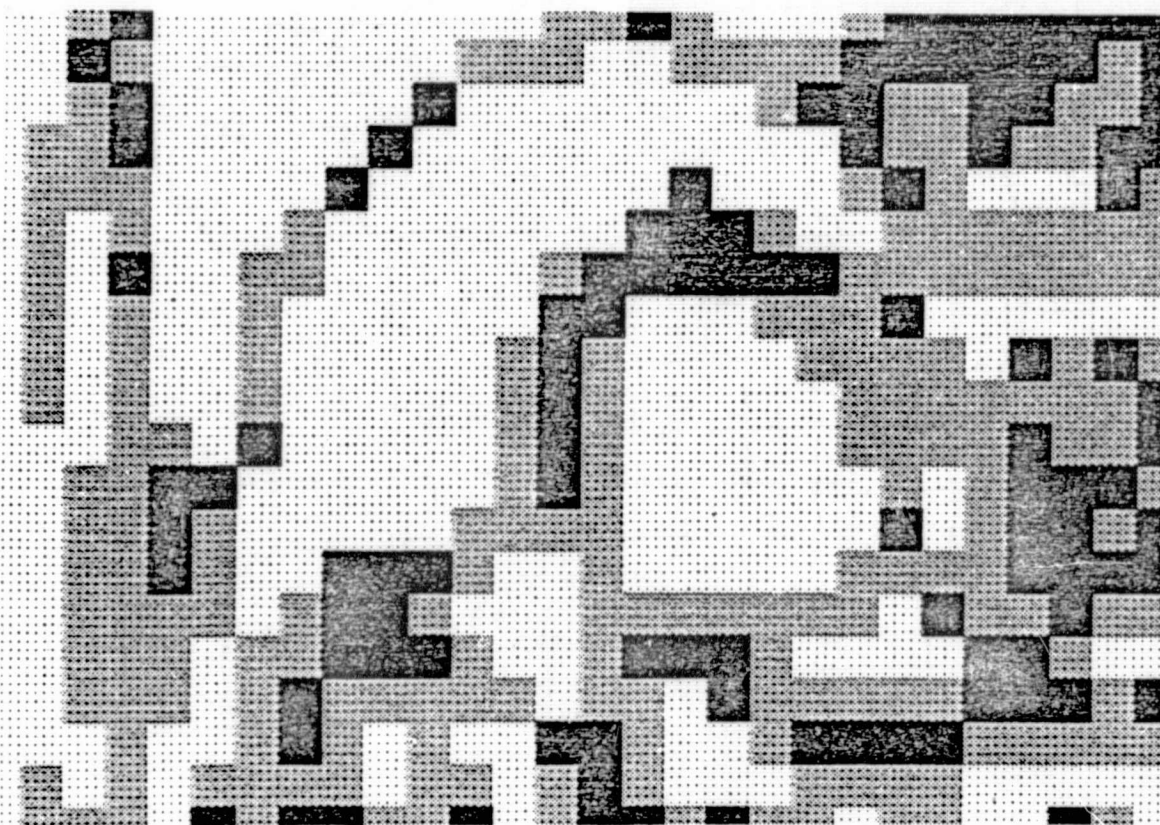


Figure 2. Interspersion map for the same area shown in Figure 1. "A" is low, ("IS" is  $>.3$  to  $.6$ ), "B" is intermediate ("SD" is  $0$  to  $.3$ ), and "C" is high ("SD" is  $>.6$  to  $1.0$ ) spatial diversity.

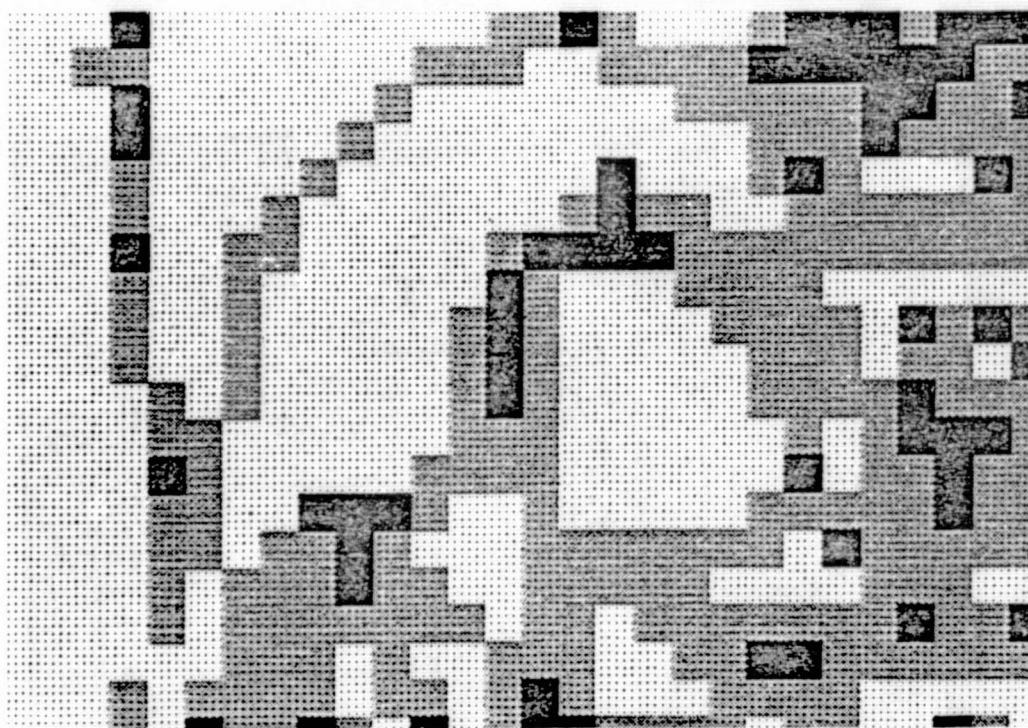


Figure 3. Juxtaposition map for the same area shown in Figure 1. "A" is low ("JX" is 0 to .3), "B" is intermediate ("JX" is  $>.3$  to .6) and "C" is high ("JX" is  $>.6$  to 1.0) juxtaposition.



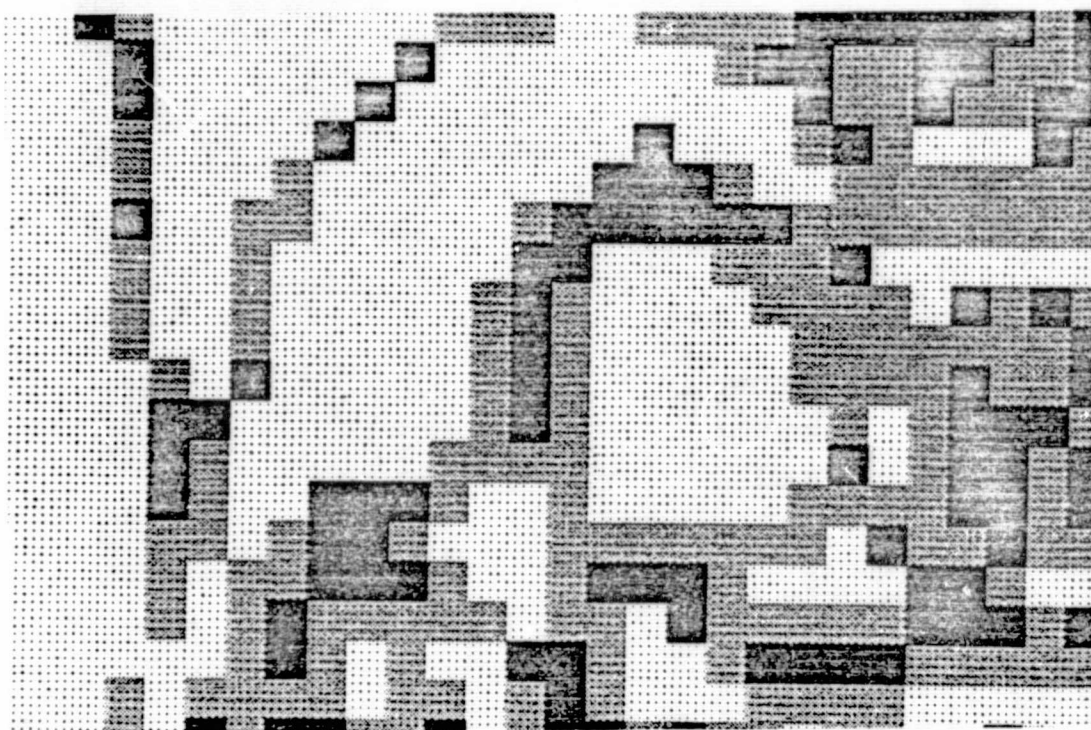


Figure 4. Spatial diversity map for the same area shown in Figure 1.

"A" is low quality ("I" is 0 to .3), "B" is intermediate ("I" is >.3 to .6) and "C" is high quality ("I" is >.6 to 1.0).