

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

AgRISTARS

RR-U3-04435

USDAES-83NEAR 327

E84-10029

CR 171 703
C.1

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made hereof."

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

Renewable Resources Inventory

January 1983

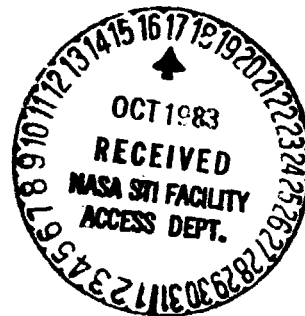
FINAL COOPERATIVE RESEARCH REPORT Project No. 1N04130

UPDATE AND REVIEW OF ACCURACY ASSESSMENT TECHNIQUES FOR REMOTELY SENSED DATA

Russell G. Congalton, Joel T. Heinen, and Richard G. Oderwald
Cooperative Agreement No. 13-1134

Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

Prepared for
U.S. Department of Agriculture, Forest Service
Nationwide Forestry Applications Program
Houston, Texas 77058

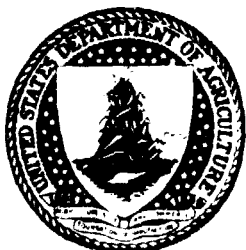


(E84-10029) UPDATE AND REVIEW OF ACCURACY
ASSESSMENT TECHNIQUES FOR REMOTELY SENSED
DATA Final Report (Virginia Polytechnic
Inst. and State Univ.) 41 p HC A03/MF A01

N84-10646

Unclass

CSCI 02C G3/43 00029



Lyndon B. Johnson Space Center
Houston, Texas 77058

1. Report No. RR-U3-04435		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Update and Review of Accuracy Assessment Techniques for Remotely Sensed Data				5. Report Date January 1983	
				6. Performing Organization Code	
7. Author(s) Russell G. Congalton, Joel T. Heinen, and Richard G. Oderwald				8. Performing Organization Report No. Remote Sensing Research Report 83-1	
9. Performing Organization Name and Address School of Forestry and Wildlife Resources VPI & SU Blacksburg, VA 24061				10. Work Unit No.	
				11. Contract or Grant No. Cooperative Agreement Number 13-1134	
12. Sponsoring Agency Name and Address Nationwide Forestry Applications Program Renewable Resources Inventory Project Houston, Texas 77058				13. Type of Report and Period Covered	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract An update and review of the research performed at Virginia Tech in the accuracy assessment of remotely sensed data during the past three years is given. This research included the use of discrete multivariate analysis techniques for the assessment of error matrices, the use of computer simulation for assessing various sampling strategies, and an investigation of spatial autocorrelation techniques.					
17. Key Words (Suggested by Author(s))				18. Distribution Statement	
19. Security Classif. (of this report) U.		20. Security Classif. (of this page) U.		21. No. of Pages	
				22. Price*	

PROJECT SUPPORTED BY

Nationwide Forestry Applications Program
Renewable Resources Inventory Project
Houston, Texas

Cooperative Agreement Number 13-1134

January 1983

Remote Sensing Research Report 83-1
Update and Review of Accuracy Assessment Techniques
for Remotely Sensed Data

by

Russell G. Congalton ^{1/}
Joel T. Heinen ^{1/}
Richard G. Oderwald ^{2/}

^{1/} Graduate Research Assistant, School of Forestry and Wildlife Resources, VPI & SU, Blacksburg, VA 24061.

^{2/} Associate Professor, School of Forestry and Wildlife Resources, VPI & SU, Blacksburg, VA 24061.

Table of Contents

	<u>Page</u>
List of Figures	i
Acknowledgments	ii
Abstract	iii
1.0 Introduction	1
2.0 Discrete Multivariate Analysis Techniques for Accuracy Assessment	2
2.1 Review of the Normalization Procedure	4
2.2 Review of the Test of Agreement Procedure	5
2.3 Review of the Multi-factor Comparison Procedure	6
3.0 Sampling Simulation for Accuracy Assessment	7
3.1 Preliminary Results	8
3.1.1 Intra-cluster Correlation Coefficients	9
3.2 Conclusions	11
3.3 Further Work	12
4.0 Spatial Autocorrelation Analysis	13
5.0 Conclusions	15
6.0 Literature Cited	16
Appendix I Multi-factor Comparison Example	18
Appendix II Spatial Autocorrelation Statistics for the Non-free Sampling Binary Classification Case	32

List of Figures

<u>Figure</u>	<u>Title</u>	<u>Page</u>
1	The form of a typical error matrix	3
2	A plot of average ROH vs. cluster size for each vegetation environment	10

ACKNOWLEDGMENTS

The authors would like to thank the Nationwide Forestry Applications Branch, especially Mr. James Bell and Dr. F. P. Weber, for their support and encouragement over the past three years.

ABSTRACT

An update and review of the research performed at Virginia Tech in the accuracy assessment of remotely sensed data during the past three years is given. This research included the use of discrete multivariate analysis techniques for the assessment of error matrices, the use of computer simulation for assessing various sampling strategies, and an investigation of spatial autocorrelation techniques.

1.0 Introduction

This report is an update and review of the research that was conducted at Virginia Tech in accuracy of remotely sensed data over the past three years. The majority of the report will briefly review the use of discrete multivariate analysis for assessing the accuracy of remotely sensed data. Wherever appropriate, a citation where more detailed information can be found will be given. The remainder of the report will discuss our continuing research in sampling simulation for accuracy assessment and the effects of spatial autocorrelation on accuracy. Wherever possible, preliminary results will be given.

2.0 Discrete Multivariate Analysis Techniques for Accuracy Assessment

Three discrete multivariate analysis procedures are used in the accuracy assessment of remotely sensed data. All three procedures operate on error matrices. An error matrix (Figure 1) is a square array of numbers set out in rows and columns which express the number of cells assigned as a particular land cover type relative to the actual cover type as verified in the field. The columns usually represent the reference data (ground verified) and the rows indicate either the Landsat classification or the photo interpretation.

ORIGINAL PAGE IS
OF POOR QUALITY

		REFERENCE DATA				
		A	B	C	D	E
LABORATORY DATA	A					
	B					
	C					
	D					
	E					

Figure 1: The form of a typical error matrix.

2.1 Review of the Normalization Procedure

The normalization procedure is a method of standardizing each error matrix so that a direct comparison of individual cell values is possible. It is an iterative process (Bishop et al. 1975) by which the rows and columns of the matrix are successively balanced until each row and column sums to a given value (marginal). Therefore, each cell value is influenced by the omission and commission errors for that particular land cover category. After normalization, the cell values in corresponding positions of two or more error matrices can be compared without regard for differences in sample size between matrices.

A FORTRAN computer program called MARGFIT can be used for performing the normalization process (Congalton et al. 1981). For details and examples of this technique see Congalton (1981).

2.2 Review of the Test of Agreement Procedure

The test of agreement procedure is a method of testing the similarity or agreement between two or more error matrices (Bishop et al. 1981). This measure of agreement, called KHAT, is based on the difference between the actual agreement of the classification (i.e., agreement between the remote sensor data and the reference data) indicated by the diagonal cell value and the chance agreement which is indicated by the row and column marginals.

KHAT values are calculated for each matrix and reflect how well the remote sensor data agrees with the reference data. A test can then be performed between two independent KHAT values in order to determine if they are significantly different, i.e., if one matrix is significantly different from another. The equations, more details, and examples can be found in Congalton (1981) and Congalton and Mead (1983). Also a FORTRAN computer program called KAPPA can be used to perform this test of agreement procedure (Congalton 1981).

2.3 Review of the Multi-factor Comparison Procedure

The multi-factor comparison procedure allows more than one factor affecting the classification accuracy to be examined at the same time. Appendix I contains a detailed explanation of this approach as well as a fully worked out example. For more examples and the APL computer program, CONTABLE, used to perform the computations see Congalton (1981).

3.0 Sampling Simulation for Accuracy Assessment

Monte Carlo computer simulation techniques were used to test the effects of various sampling schemes on the accuracy assessment of remotely sensed data. Three data sets of varying spatial complexity were used: a forested area, a rangeland area, and an agricultural area. Each area had two classified data sets associated with it. One of the data sets (usually photointerpretation) was assumed correct and the other was the Landsat classification. A difference image was then created for each area by comparing, pixel by pixel, the assumed correct data set with the Landsat classification. A difference image is a matrix of zeros and ones in which the zeros indicate agreement between the data sets and the ones indicate disagreement. In other words, the difference image is indicative of the pattern of error occurring in the classification. The population parameters (mean and variance) were computed from a total enumeration of each difference image and were used to compare with the sample statistics from the sampling simulations.

For a more detailed description of the data sets, as well as pictures of the difference images, see Congalton et al. (1982).

3.1 Preliminary Results

Simulations have been run for simple random sampling and cluster sampling. Testing of other sampling schemes is yet to be done.

3.1.1 Intra-cluster Correlation Coefficients

Intra-cluster correlation coefficient, ROH, is a measure of the homogeneity of the cluster. The more homogeneous the cluster, the greater the value of ROH. In order to maximize the given information within a cluster, one would wish the cluster to be as heterogeneous as possible. Therefore, one would try to make ROH approach zero. Figure 2 shows a plot of average ROH vs. cluster size for each of the vegetation environments.

10

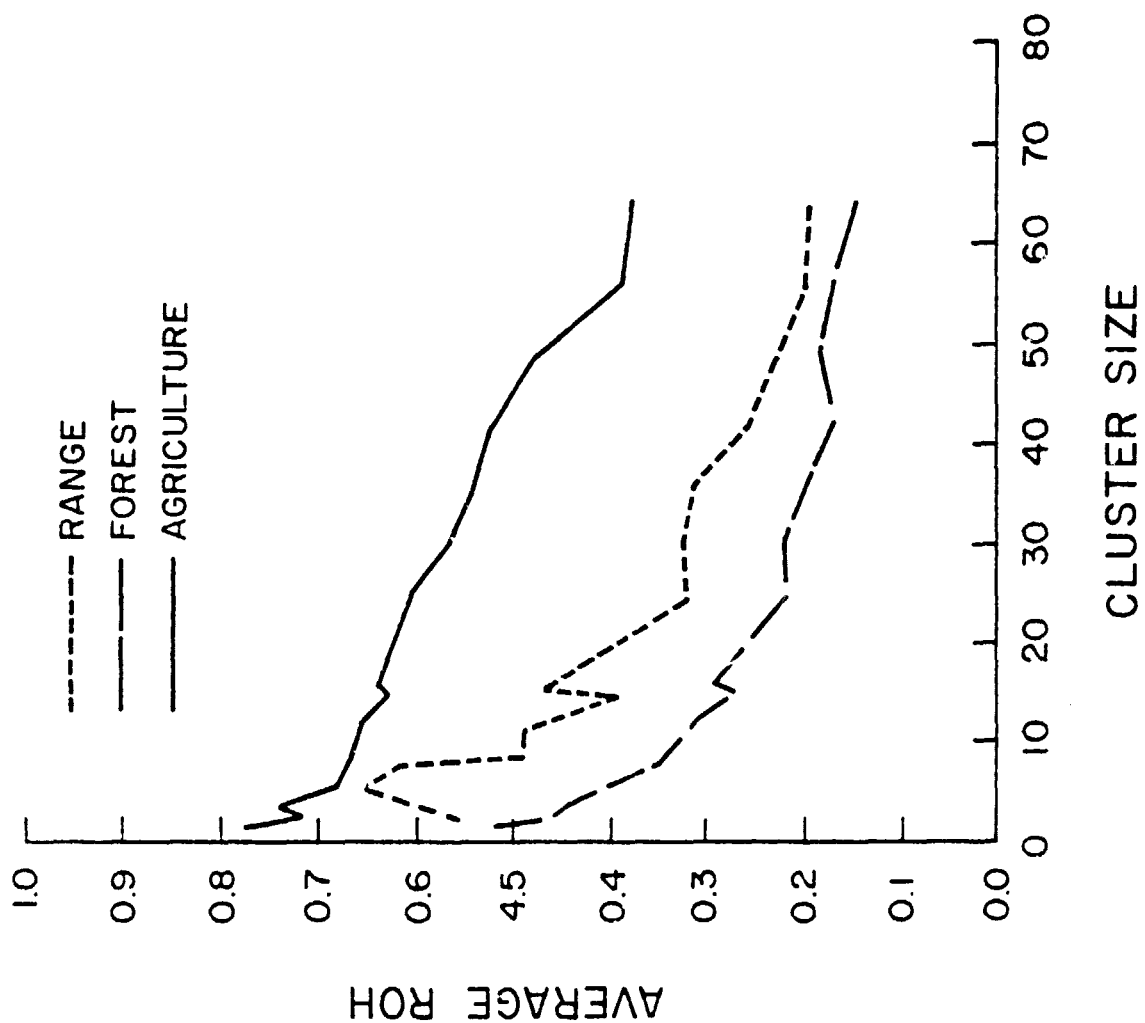


Figure 2. A plot of average ROH vs. cluster size for each vegetation environment.

3.2 Conclusions

Visual inspection of the difference images yielded different levels of homogeneity or spatial complexity in the three vegetation environments. As expected, the agricultural area was most homogeneous while the forested area was most diverse. This result was also demonstrated in the plot of ROH vs. cluster size. Given that a large ROH means greater homogeneity notice that for each cluster size, the agricultural environment had the largest ROH, while the forest had the smallest.

Also demonstrated in the plot of ROH vs. cluster size were some guidelines for cluster size determination. Despite the theoretical statement that ROH should go to zero, it is apparent from this plot that this is not always practically feasible. Note from the plot that between 0 and 20 pixels/cluster, ROH decreases rather quickly while after this point, the line levels off. It can be concluded from this result that large clusters taken to reduce ROH may not be gaining significantly in information while costing excessive time and money.

3.3 Further Work

Research is continuing at Virginia Tech in sampling simulation as well as in the spatial autocorrelation analysis techniques that will be discussed next.

4.0 Spatial Autocorrelation Analysis

Spatial autocorrelation analysis is a technique by which the pattern of a spatially distributed attribute can be investigated. In other words, if the presence of a given attribute in a certain location makes its presence in surrounding locations more or less likely, then the attribute is said to exhibit spatial autocorrelation (Cliff and Ord 1973).

Spatial autocorrelation analysis can be used to investigate the pattern of error in the difference images created from a Landsat classification and an assumed correct reference set. In this situation, the discrete binary classification applies. Each pixel has either been classified correctly or incorrectly and therefore a technique called join count statistics may be used to measure the spatial autocorrelation. A join is defined if any two pixels have a boundary of positive non-zero length in common (Cliff and Ord 1973).

In the case of a difference image, there are three possible joins, 0-0 (correct-correct), 1-1 (incorrect-incorrect), and 0-1 or 1-0 (correct-incorrect, incorrect-correct). The method that is used to test if the pattern of error in the difference image differs significantly from random is to use the fact that the join count statistics are asymptotically normal. The mean and variance (first and second moments) are obtained using the equations in Appendix II and are compared with the observed counts to test for significance.

There are two possible sampling schemes from which the statistics can be derived. These are free sampling (i.e., sampling with replacement) or non-free sampling (i.e., sampling without replacement). In the case of a difference image, non-free sampling is employed since we assume that each pixel has the same a priori probability of being right or wrong (Cliff and Ord 1973).

Combining the results of the spatial autocorrelation analysis with the results of the sampling simulations allows better interpretation of when to use which sampling scheme. By being able to examine the spatial pattern of the errors within the difference image, the results of the sampling simulation can be better explained.

5.0 Conclusions

The research conducted at Virginia Tech over the last three years has contributed significantly to the accuracy assessment of remotely sensed data. The application of discrete multivariate analysis to accuracy assessment has provided a new technique for analyzing accuracy. The research in sampling simulation and spatial autocorrelation is yet to be completed. However, it is felt that combining the results of sampling simulation with the knowledge gained from spatial autocorrelation will yield better recommendations for which sampling schemes to use when.

6.0 Literature Cited

- Bishop, Y., S. Fienberg, and P. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press:Cambridge, Massachusetts. 575 pp.
- Cliff, A. D. and J. K. Ord. 1973. Spatial Autocorrelation. Pion: London. 178 pp.
- Congalton, Russell G. 1981. The Use of Discrete Multivariate Analysis Techniques for the Assessment of Landsat Classification Accuracy. Unpublished M.S. Thesis. Virginia Polytechnic Institute and State University, Department of Forestry, Blacksburg, Virginia 24061. 111 pp.
- Congalton, Russell G. and Roy A. Mead. 1983. A quantitative method to test for consistency and correctness in photo interpretation. Photogrammetric Engineering and Remote Sensing. Vol. 49, No. 1. pp. 69-74.
- Congalton, Russell G., Richard G. Oderwald, and Roy a Mead. 1982. Accuracy of remotely sensed data: sampling and analysis procedures. AGRISTARS Report. Nationwide Forestry Applications Program. Cooperative Agreement No. 13-1134. 83 pp.
- Congalton, Russell G., Roy A. Mead, Richard G. Oderwald, and Joel T. Heinen. 1981. Analysis of forest classification accuracy. AGRISTARS Report. Nationwide Forestry Applications Program. Cooperative Agreement No. 13-1134. 85 pp.

Moran, P. A. P. 1948. The interpretation of statistical maps.
Journal Royal Statistical Society, Series B. Vol. 10.
pp. 243-251.

Appendix I

Multi-factor Comparison Example

**The Use of Two Model Fitting Procedures for
Determining Associations Between Four Spatial Variables**

ABSTRACT: This study describes the use of the log-linear and logit model fitting procedures which yield the best fitting model for determining associations between four spatially defined variables. These variables include (1) interspersions, (2) cover type, (3) aspect, and (4) elevation.

Introduction

A great deal of information has been generated concerning interactions of physical variables such as aspect and elevation on biotic variables such as cover type (Barbour et al. 1980). Knowledge of such interactions can be used as predictive tools in forestry and wildlife habitat analysis. Statistical procedures have been developed which can be used to determine associations between cross-classified categorical variables (Fienberg 1980; Bishop et al. 1975). Since most spatial variables may be readily grouped into categories (such as low, medium, and high elevations, or south vs. north facing slopes, etc.), these statistical procedures represent important analytic tools for resource managers.

The purpose of this study was to determine associations between four such spatially defined variables. These included interspersions,

cover type, aspect, and elevation. The interspersion index used in this study was described by Mead et al. (1981). It may be used as an index of habitat quality for any selected species. Thus, using the statistical procedures outlined here, associations between habitat quality, elevation, aspect, and cover type may be determined.

Study Area

The study area included 1542 acres situated in the northwest section of the Rampart Hills Quadrangle in the San Juan National Forest, Colorado. This area was chosen because recent cover type information was available and it has been used in previous habitat studies.

Methods and Procedures

Data Acquisition

The cover type information was provided by the Forest Service, and was originally derived from Landsat imagery. The one acre Landsat pixels were grouped into three acre sampling units for the purpose of cover type mapping. This work was completed by the Lockheed Corporation under Forest Service contract. The 19 categories present in the original work were collapsed into seven categories which represent important habitat components for mule deer (Odocoileus hemionus). This collapsing procedure followed recommendations of Forest Service personnel (Cook,

personal communication). Of these seven categories, four were present in the study area. These included (1) conifer, (2) oak, (3) aspen, and (4) other. Categories which do not form critical habitat components for the species of interest were grouped into the fourth category.

Interspersion was calculated for each cell (3-acre sampling unit) in the study area. The method of calculation was described by Mead et al. (1981) and by Heinen et al. (1981). Two categories of interspersion were used in this study. These included low (0 to 0.5) and high (>0.5 to 1.0).

A grid representing each cell in the study was drawn directly onto a copy of the U. S. Geological Survey 7-1/2 minute Rampart Hills Quadrangle. The elevation and aspect information for each cell was then obtained directly from this map by following the contour lines. Two categories for each variable were used. The categories used for the elevation variable included low (≤ 8920 ft) and high (> 8920 ft). This arbitrary cutoff point was chosen because this contour line roughly divided the study area into two equal parts. The use of only two elevational categories is justified because there is only an approximate 1400 foot range in elevation over the entire study area. In many cases the 8920 foot contour transected individual cells. In these cases, a judgment was made as to whether more than one-half of the cell was above or below the cutoff point, and then categorized accordingly.

Each cell was grouped into one of two categories of aspect. These included south and north facing slopes. These two categories were chosen because of their ecological significance (Spurr, Barnes, 1973), and also because the ridges in the study area are oriented approximately from southwest to northeast. This topography greatly facilitated the categorization of aspect for each cell.

The data set is presented in Table I. For the purposes of the statistical notation (explained below) the number assigned to each variable is important. These are (1) interspersions, (2) cover type, (3) aspect, and (4) elevation. Numbers are likewise assigned to each category within each variable. Variable 1 (interspersions) has two categories, (i) low and (ii) high, and variable 2 has four categories denoted as (i) conifer, (ii) oak, (iii) aspen, and (iv) other. Variables 3 (aspect) and 4 (elevation) each have two categories. In the case of the former these include (i) north-facing slopes, and (ii) south-facing slopes. In the case of the latter these include (i) low elevation (≤ 8900 ft) and (ii) high elevation (> 8900 ft). The number designated in each cell on Table I indicates the total number of cells in the study area which fall into that particular combination of variable categories. For example, a total of nine cells in the study area had low interspersions and were dominated by conifer on north-facing slopes at elevations at or below 8920 feet.

Table I. The number of cells in the study area described in each of 32 possible combinations of the four categorical variables.

Interspersion (1)	Cover Type (2)	Aspect (3)			
		North		South	
		Elevation (4) Low	High	Elevation (4) Low	High
Low (i)	Conifer (i)	9	0	15	4
	Oak (ii)	23	24	38	40
	Aspen (iii)	5	6	40	11
	Other (iv)	3	3	7	50
High (ii)	Conifer (i)	6	3	13	7
	Oak (ii)	10	21	23	33
	Aspen (iii)	10	12	23	13
	Other (iv)	3	13	9	27

Statistical Procedures

The logit and log-linear model fitting procedures were both used in this study. In each case the objective is to determine the simplest best fitting model which explains the data. Each procedure is outlined here.

Log-linear Model Fitting

The first step in this procedure is to test all uniform models. These may be defined as models which contain all n-way interactions where n ranges from one to the number of variables (Fienberg 1980). The uniform

order model is denoted by τ^* . If $\tau^* = 3$, for example, this indicates that all 3-way interaction terms are present in the model. All lower order terms are also present by default. Thus in the example above, all 2-way and 1-way terms would also be included in the model.

After all uniform order models are tested, one of the stepwise model fitting procedures may be used. These are the forward and the backward procedures. In the case of the former, the researcher chooses the uniform order model which provides a poor fit ($p < .05$) where the next higher uniform order model provides a good fit ($p > .05$). In this study, for example, $\tau^* = 1$ (no interaction terms) provided an extremely poor fit ($p < .005$), $\tau^* = 2$ provided a poor fit ($p < .025$), and $\tau^* = 3$ provided a good fit ($.5 > p > .25$). Thus $\tau^* = 2$ was chosen to begin the forward stepwise model building procedure. The next step involves adding the next higher order interaction terms to the model one by one and the resulting model which yields the highest p-value is chosen. This is repeated until all significant terms are included in the model. In each case, the criteria for testing models is based on the Likelihood Ratio (G^2) which is an asymptotic τ^2 distribution. The critical value for testing each model may therefore be obtained from a τ^2 table using the proper degrees of freedom.

The backward selection procedure begins with the uniform order model which provides a good fit ($p > .05$) where the next lower uniform order model provides a poor fit ($p < .05$). In this study the uniform order model

$\tau^* = 3$ was chosen to begin the backward procedure. Successive 3-way interaction terms are then dropped from the model. Each resulting model is again tested with the G^2 statistic.

Logit Fitting Procedure

When using the logit model fitting procedure, the researcher first assumes that there is one response variable, and all other variables are explanatory. The term denoting the interaction of all explanatory variables is therefore present in every model tested. In this study it was assumed that interspersions are a response of cover type, aspect, and elevation. Thus the term denoting the interaction of variables 2, 3, and 4 was present in every model tested. Choosing the proper uniform order model when using the logit procedure is similar to that described for the log-linear procedure as is the forward and backward model selection.

All model selection procedures described here were conducted using an interactive APL computer program written by Dr. S. K. Lee of the Department of Statistics, VPI & SU. See Fienberg (1980) and Bishop et al. (1975) for a more thorough description of these statistical procedures.

Results and Discussion

Using forward and backward stepwise model selections for both the log-linear and logit procedures, the best fitting model is as follows:

$$[1\ 2\ 3] \quad [2\ 3\ 4] \quad [1\ 4]$$

This indicates that variables 1 (interspersions), 2 (cover type), and 3 (aspect) all interact jointly. Variables 2 (cover type), 3 (aspect), and 4 (elevation) also interact jointly. The additional 2-way interaction term of variables 1 (interspersions), and 4 (elevation) is also a significant feature of the model. Thus, all possible 2-way interaction terms are present by default in this model.

In order to analyze the particular associations represented by each individual interaction term, tables were prepared for each term by summing across the variable(s) not present in that term (Tables 2, 3, and 4). In preparing the table for the $[1\ 2\ 3]$ interaction term (Table 2), for example, the raw data presented in Table I was summed across variable 4, which is not present in that term. The numbers obtained by this summation were then converted into proportions by dividing each into the sum of all numbers on the table. In this way, the type of association between variables may be readily determined by comparing the proportions.

Table 2. The observed frequencies describing the [1 2 3] interaction term.

Low Interspersion				High Interspersion			
	North Slope	South Slope			North Slope	South Slope	
Conifer	.03	.07	.10	Conifer	.04	.09	.13
Oak	.17	.28	.45	Oak	.14	.25	.38
Aspen	.04	.18	.22	Aspen	.10	.16	.26
Other	.02	.21	.23	Other	.07	.16	.23
	.26	.74			.35	.65	

Table 2 indicates a general trend toward lower interspersion on south than north facing slopes. Lower interspersion values also tend to be associated with oak stands, whereas higher interspersion values tend to be associated with conifer and aspen stands. However, aspen, as well as cells designated as other tend to be associated with low interspersion values on south-facing slopes.

Table 3. Observed frequencies describing the [2 3 4] interaction term.

Low Elevation				High Elevation			
	North Slopes	South Slopes			North Slopes	South Slopes	
Conifer	.06	.12	.18	Conifer	.01	.04	.05
Oak	.14	.26	.40	Oak	.17	.27	.44
Aspen	.06	.27	.33	Aspen	.07	.09	.16
Other	.03	.06	.09	Other	.06	.29	.35
	.29	.71			.31	.69	

Table 3 indicates that conifer and aspen are more prevalent at low elevations, while oak and the "other" category are more prevalent at higher elevations. The conifer and aspen are more common on south vs. north-facing slopes in these lower elevations. Cover type 4 (other) is more common at higher elevations on south vs. north-facing slopes. It must be pointed out that we may expect different trends for different conifer species. Spruce and fir, for example, may be expected to grow more readily at higher elevations on north-facing slopes. The data base used here, however, collapsed categories according to the habitat requirements of mule deer, and thus all conifers were included in one category. Trends for different conifer species could be readily determined using the same procedures had these categories not been collapsed.

Table 4. Observed frequencies describing the [1 4] interaction term

	Low Interspersion	High Interspersion	
Low Elevation	.28	.19	.47
High Elevation	.27	.26	.53
	.55	.45	

Table 4 indicates that higher interspersion is generally associated with higher elevations, and low interspersion is more common at low elevations.

The information obtained from the model building procedures presented here has utility as a predictive tool. Managerial decisions may be based in part on such information. For example, clearcuts or prescription burns could be placed more appropriately within specific elevational and slope regimes to achieve a generally higher interspersal index throughout and area if this is desirable.

Of particular interest in this scenario is the interaction of interspersal with the other variables. Knowledge of specific interactions can be used to stratify areas according to its habitat potential for a selected species resulting from the effects of cover type, elevation, and aspect. One important point is that the model which resulted in this case can only be applied to this area and to this species (mule deer). This is because of the cover type collapsing procedure. If such information is desired for other areas or species, a different collapsing procedure may be necessary.

One major advantage of this technique is that, after initial data generation, the model building procedure itself is a rather rapid process when using an interactive program. The entire process of forward and backward log-linear and logit model selection procedures took only 1-1/2 hours to complete on the computer terminal. This time would vary depending on the experience of the researchers with interactive programs, as well as on the number of factors to be analyzed in the model, but it is generally a rapid process.

Other spatial variables of interest could easily be tested. Some of these may include water availability, slope (steepness) or human disturbances. These could easily be categorized appropriately and used as additional dimensions in the multi-way classification in determining the effects of each on habitat quality. Such procedures are thus readily expandable and can easily test the simultaneous effects of any number of variables which may affect the aspects of habitat quality analyzed here.

Literature Cited

- Bishop, Y., S. Fienberg, and P. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press:Cambridge, Mass. 575 pp.
- Barbour, M. D., J. H. Burke, and W. D. Pitts. 1980. Terrestrial Plant Ecology. Benjamin Cummings Publishing Co.
- Cook, D. 1981. Personal Communication. Wildlife Biologist, U.S.F.S., San Juan National Forest, Colorado.
- Fienberg, S. E. 1980. The Analysis of Cross-Classified Categorical Data. MIT Press:Cambridge, Mass. 198 pp.
- Heinen, J. T., R. A. Mead, and J. L. Smith. 1981. Site-specific maps of juxtaposition and interspersion. S.A.F. National Workshop. In-Place Resource Inventories: Principles and Practices. Orono, Maine.
- Mead, R. A., T. Sharik, S. P. Prisley, and J. T. Heinen. 1981. A computerized spatial analysis system for assessing wildlife habitat from vegetation maps. Can. J. Rem. Sens. 7(1):34-40.
- Spurr, S. and B. Barnes. 1973. Forest Ecology (2nd Edition). The Ronald Press Co., New York, NY.

Appendix II

Spatial Autocorrelation Statistics for the Non-free Sampling Binary Classification Case*

*Moran, P.A.P. 1948. The interpretation of statistical maps.
Journal Royal Statistical Society, Series B. Vol. 10.
pp. 243-251.

Autocorrelation

Non-free sampling
binary classification

$$\text{like join} \quad (1-1) = 1/2 \sum_i \sum_j \delta_{ij} X_i X_j$$

$$\mu_1' (1-1) = \frac{An_1^{(2)}}{n^{(2)}}$$

$$\mu_2' (1-1) = \frac{An_1^{(2)}}{n^{(2)}} + \frac{2Dn_1^{(3)}}{n^{(3)}} + \frac{(A(A-1)-2D)n_1^{(4)}}{n^{(4)}} - \left[\frac{An_1^{(2)}}{n^{(2)}} \right]^2$$

$$\text{unlike join} \quad (1-0) = 1/2 \sum_i \sum_j \delta_{ij} (X_i - X_j)^2$$

$$\mu_1' (1-0) = \frac{2An_1n_2}{n^{(2)}}$$

$$\mu_2' (1-0) = \frac{2An_1n_2}{n^{(2)}} + \frac{4(A(A-1)-2D)n_1^{(2)}n_2^{(2)}}{n^{(4)}}$$

$$+ \frac{2Dn_1n_2(n_1+n_2-2)}{n^3} - 4 \left[\frac{An_1n_2}{n^{(4)}} \right]^2$$

Notation

$\mu_1'(X)$ = The first moment of X about the origin, the expected value of X .

$\mu_2(X)$ = The second moment of X about the mean, the variance of X .

$n^{(i)} = n(n-1) \dots (n-i+1)$

n = The total number of individuals in the population.

n_1 = The number of individuals in the population with the characteristic of interest.

n_2 = The number of individuals in the population without the characteristic of interest.

L_i = The number of individuals joined with the i^{th} individual.

$A = 1/2 \sum_i L_i$

$D = 1/2 \sum_i L_i (L_i - 1)$

$s_{ij} = 1$, if i^{th} and j^{th} areas are joined
0, otherwise

$x_i = 1$, if i^{th} area is correctly classified
0, otherwise