THE EFFECT OF THRESHOLDING A MAXIMUM LIKELIHOOD CLASSIFIER ON THE ACCURACY OF A LANDSAT CLASSIFICATION OF A FORESTED WETLAND

by

Jennie M. Agnello

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Forestry

APPROVED:

James L. Smith-Chairman

James B. Campbell

Steven D. Johnson

August, 1987

Blacksburg, Virginia

THE EFFECT OF THRESHOLDING A MAXIMUM LIKELIHOOD CLASSIFIER ON THE ACCURACY OF A LANDSAT CLASSIFICATION OF A FORESTED WETLAND

by

Jennie M. Agnello

James L. Smith-Chairman

Forestry

(ABSTRACT)

Although the maximum likelihood classifier is a popular classification technique, there is an inherent problem associated with the 100% classification of a scene. This is because there will inevitably be pixels within a study area that have a low probability of belonging to any of the predefined categories.

The focus of this research was to locate these low probability pixels and observe their affect on classification accuracy. This was done by performing supervised classifications at various threshold levels using two methods of classification training: combined category training site statistics and separated category training site statistics. In general, it was found that a majority of the scene was classified at very low probabilities but the accuracy of the resulting classifications was much greater than the low probabilities would suggest.

Other conclusions were that the separated training sites did represent the spatial variation within a category better than the combined training sites thus improving overall accuracy; however, spatially classifiying the unclassified pixels after thresholding did not improve accuracy over a standard classification for either training method.

ACKNOWLEDGEMENTS

I cannot claim this work to be totally my own; it is a result of the intellectual and emotional contributions of many people. First, I would like to express my deepest gratitude to Jim Smith who believed in my capabilities more than I did, and supported me both as a friend and an advisor. Second, thanks go to my Virginia Tech buddies who made it possible for me to face returning to school each of the six years I was here. Third, I owe my sanity and deepest love to . When I felt like I was drowning under the pressures of graduate work, he always managed to pull my head up from under water. Finally, I would like to dedicate this thesis to my parents. They have never said a word, but I know how much they sacrificed to get me where I am today. My parents supported me through every success and failure I encountered, and I could have never come this far without them. I love and admire them very much.

TABLE OF CONTENTS

INTRODUCTION AND JUSTIFICATION 1

LITERATURE REVIEW 5

The Pattern Classifier 7

The Maximum Likelihood Classifier 10

Accuracy Assessment 11

DATA AND PROCEDURES 15

Description of Study Area 15

Description of Spectral Data 18

Procedural Outline 21

Classification Procedures 22

Modification of a Maximum Likelihood Classifier . . 23

Logical Filling 25

Accuracy Assessment 28

The Spatial Distribution of Unclassified Pixels . . . 31

RESULTS AND DISCUSSION 32

Study Area and Training Site Statistics 32

Combined Training Site Classifications 37

Thresholding the Classifications 37

Logical Fill 45

Separated Training Site Classifications 48

Thresholding the Classifications 48
Logical Fill 53
Separated Versus Combined Category Training 56
Spatial Distribution of Unclassified Pixels 60

SUMMARY AND CONCLUSIONS 68

LITERATURE CITED 74

APPENDIX A. ERROR MATRICES-COMBINED TRAINING CLASSIFICATIONS AND FILLS 77

APPENDIX B. ERROR MATRICES-SEPARATED TRAINING CLASSIFICATIONS AND FILLS 97

VITA 107

LIST OF ILLUSTRATIONS

Figure 1. Two-dimensional Spectral Reflectance Patterns for Three Categories. 8

Figure 2. Decision Regions and Boundaries Separating Three Categories. 9

Figure 3. Vegetative Cover Map of the Great Dismal Swamp. 17

Figure 4. Anderson Level II and Level III Land Cover Classes for the Great Dismal Swamp. 19

Figure 5. Flow Diagram for the Original Bayesian Classifier 26

Figure 6. The Training Site Distribution of a Category Compared to the Real Distribution of that Category. 42

Figure 7. The Similiarity Between Two Categories Resulting in Confusion During Classification Even With Represenative Training Sites. . . 54

Figure 8. Overall Accuracies for Separated and Combined Training Site Threshold Classifications. . 57

Figure 9. Percent of the Scene Classified for the Combined and Separated Training Site Classifications. 61

Figure 10. Thresholding the Training Distribution of a Category. 62

Figure 11. Percent of Unclassified Pixels on the Boundaries for Combined and Separated Training Site Classifications. 63

Figure 12. The Percentage of Boundary Pixels Unclassified. 66

LIST OF TABLES

Table 1. Training Site Statistics. 34

Table 2. Category Statistics. 35

Table 3. Statistics of the Dismal Swamp Landsat Scene. 36

Table 4. Overall Accuracy and Category Accuracies for the Combined Training Site Threshold Classifications. 38

Table 5. Percent of the Scene Classified Overall and by Category for the Combined Threshold Classifications. 39

Table 6. Overall Accuracy and Category Accuracies Following the Logical Fills for the Combined Threshold Classifications. 46

Table 7. Overall Accuracy and Category Accuracies for the Separated Training Site Threshold Classifications. 49

Table 8. Percent of the Scene Classified Overall and by Category for the Separated Threshold Classifications. 50

Table 9. Overall Accuracy and Category Accuracies Following the Logical Fills for the Separated Threshold Classifications. 55

INTRODUCTION AND JUSTIFICATION

Land cover classification of remotely sensed data is an important tool for obtaining reliable and current information about the quantity and distribution of the earth's natural resources. For instance, the use of Landsat data is increasing in popularity in fields such as crop forecasting, mineral exploration, rangeland monitoring, and forest type mapping (Lillesand 1979). The digital format of the Landsat data allows for the classification of the image using computer assistance, which is much faster and less expensive than mapping large areas from the ground (Stanton 1975).

Consequently, research is focusing on developing more efficient and more accurate methods for use in the analysis of Landsat data (Fleming 1979). One of the most important variables the user of remotely sensed data must be attentive to is the expected accuracy of the classified product. This accuracy is influenced by factors such as scene characteristics, sensor characteristics, and processing techniques. The selection of processing techniques is one of the few ways an image processor can control the accuracy of the classified image. Emphasis should therefore be placed on a thorough understanding of the available image processing methods so as to maximize the accuracy of the final product and to minimize the cost of processing (Story 1984).

The first attempt at a large scale classification using remotely sensed multispectral data was in 1971 for the Corn Blight Watch Experiment, where a per-point maximum likelihood algorithm was employed during the classification process (Hixon 1980). Since then the maximum likelihood algorithm has increased in popularity because it is considered "the most accurate classifier on the average (Swain 1978)." The reason for this lies in the statistical nature of the maximum likelihood theory. First, the algorithm estimates the probability distribution for each category from sample data of known identity. Second, it calculates the statistical probability of an unidentified pixel belonging to all categories and assigns the pixel to the category possessing the highest probability (Lillesand 1979).

An important characteristic of the maximum likelihood classifier is its ability to classify every pixel in a data set into the most likely of the defined categories. However, there is an inherent problem associated with 100% classification. When classifying pixels into land cover types, there will inevitably be pixels in the study area that have very low probabilities of belonging to any of the predefined categories. These pixels should probably remain unclassified, but they are forced into inappropriate categories by the nature of the classifier. Low probabilities could be caused by using non-representative training sets to estimate training parameters, or by the presence of unidentified categories in the data set. Low probabilities could also occur during the classification of mixed pixels which usually possess inconsistent spectral signatures. Assigning pixels to categories which do not accurately describe them could be having a detrimental influence on the classified product and therefore warrants further study.

In order to study the effect of classifying low probability pixels on classification accuracy, an experiment was designed to prevent the classification of pixels having probabilties below certain threshold levels. It is hypothesized that the maximum likelihood classifier is misclassifying a significant amount of pixels due to the algorithm's inability to reject pixels with associated high probabilities of classification error.

Specific objectives of this research are:

1. To determine the effect of thresholding a maximum likelihood classifier on the number of pixels classified in a Landsat image of the Dismal Swamp National Refuge, the accuracy of the classified pixels, and the spatial distribution of the pixels not classified.

2. To determine if the method used to train the classifier (combined category training site statistics or separated category training site statistics) alters the effect of thresholding a maximum likelihood classification.

3. To determine if a combination of a thresholding maximum likelihood classifier and a post-processing logical filler can improve classification accuracy over a traditional maximum likelihood classification.

LITERATURE REVIEW

Thorough and accurate land use and land cover maps have become an important tool in the development of public policies and management activities concerning the earth's resources (Campbell 1983). Since the 1940's the accepted procedure for making these maps included the use of panchromatic, medium-scale aerial photographs; however, more recent efforts have been focused on small-scale aerial photographs and satellite images for mapping large areas (Lillesand 1979). The Computer-assisted classification of Landsat spectral data is fast and inexpensive compared to the respective manual method, yet the classification accuracies have been lower than desired (Sharp 1979).

Unfortunately, classification of forested wetlands using Landsat data has not been an exception. One explanation for the low classification accuracies could be because wetland areas are usually too small or too narrow for the Landsat sensor to register at an acceptable level of detail (Aldrich 1979). Another reason could be linked to the tremendous vegetative diversity and interspersion present in wetland areas. This lack of class homogeneity causes an overwhelmimg number of mixed pixels which are frequently misclassified due to their unusual or inconsistent spectral responses (Ranson 1975).

Recent studies performed on wetland areas have found the only way to improve classification accuracy (with the technology given) was to consolidate detailed categories into larger more general ones. For instance, an 80% overall classification accuracy was obtained for a wetland area in Delaware (Klemas 1975) and 70.9% for a wetland area in northern Indiana (Ernst-Dottavio 1981). Both of these studies classified Landsat MSS data into broad categories such as forestland and hardwood marsh. A comparison of vegetation classes in the Great Dismal Swamp was performed by Gammon and Carter (1979) using Landsat images taken in February and April and a temporal composite of both. Neither the February or April scene could by itself provide adequate classification accuracy, but comparisons between seasons and the temporal composite provided 70% classification accuracy.

The supervised classification techniques most often used in forested wetland classifications have become less popular because of the difficulty in locating homogeneous training sites in such a complex and therefore spectrally diverse environment. Unsupervised classification techniques have the advantage in this situation due to their ability to separate discriminable spectral classes without prior knowledge of the area and to find classes which might have previously been unknown (Townshend 1980).

THE PATTERN CLASSIFIER

Discriminable spectral response patterns are the basis for most of the techniques developed to classify spectral data into informational classes. The individuality of these patterns can be visualized by plotting each category's spectral response pattern for "n" number of bands, producing a n-dimensional cloud of points for each category (Figure 1). Conceptually, the algorithm's first task is to separate the multivariate space formed by the n bands into distinct decision regions, one for each category (Figure 2). The second task is to assign the unidentified pixels to the most appropriate class based on the decision region it falls into (Swain 1978).

Since spectral response similiarity between classes is a common problem, statistical pattern-recognition techniques have been developed to distinguish between classes more clearly. A statistical approach is valid for remote sensing applications for several reasons. First, statistical analysis can account for the variations in nature that can blur the spectral boundaries of categories and cause confusion between classes. Second, the "robustness" of statistics allows for some error in the actual identity of the training samples if the sample size is large enough. In other words, statistics can produce reliable results even when the training sets are less than truely representative descriptions of the catego-

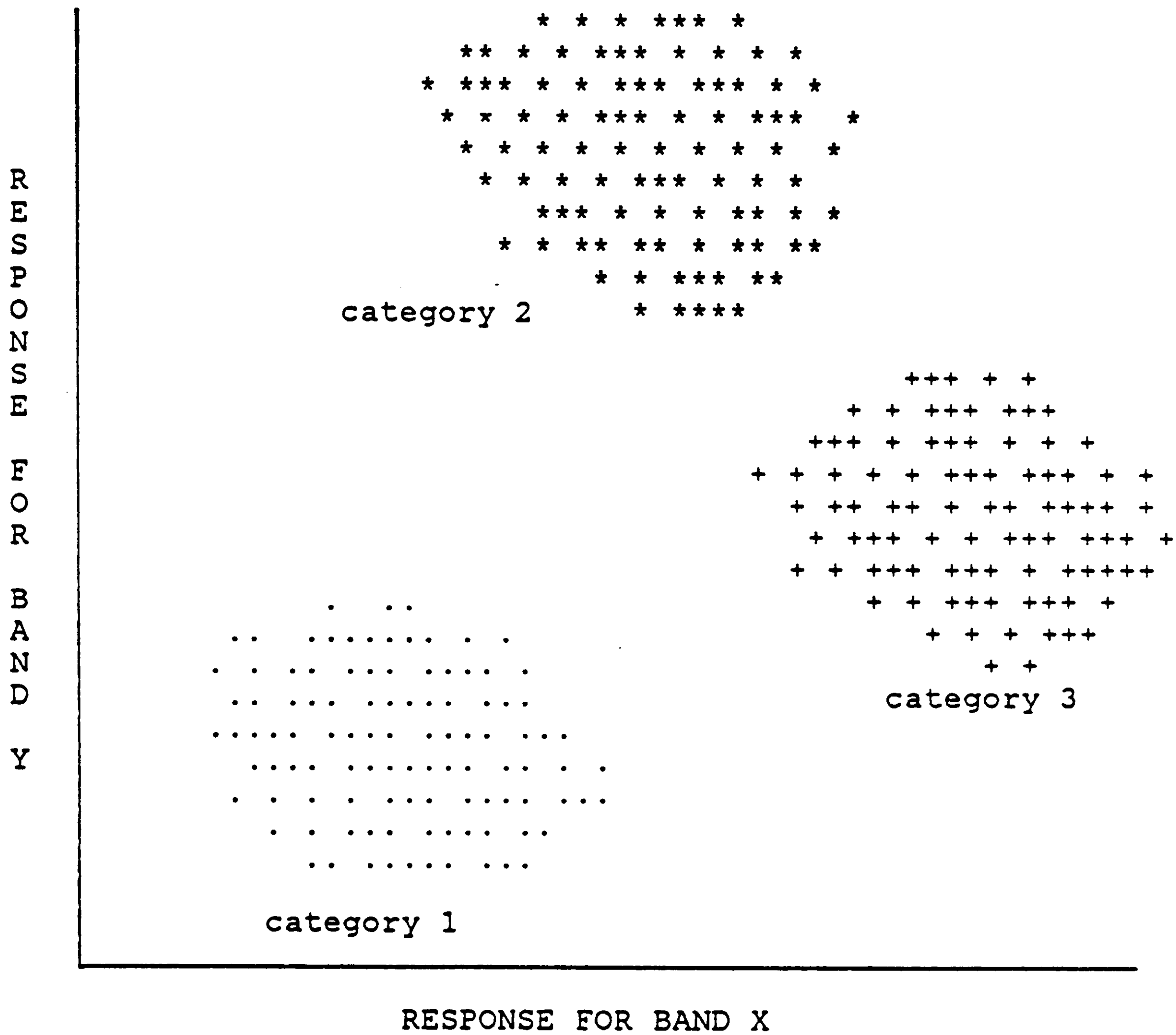


Figure 1. Two-dimensional Spectral Reflectance Patterns for Three Categories.

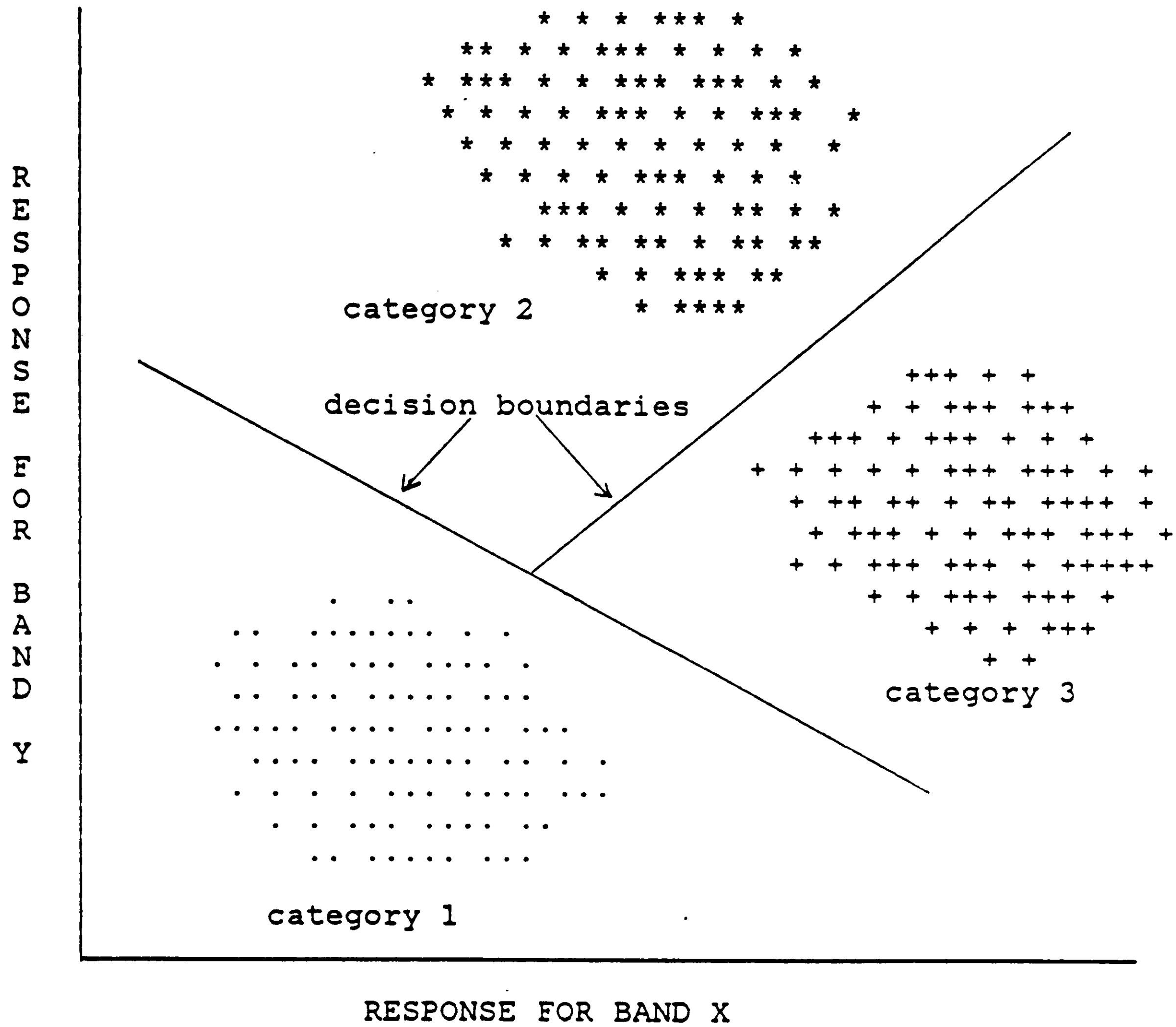


Figure 2. Decision Regions and Boundaries Separating Three Categories.

ry's spectral response. Finally, statistical pattern-recognition techniques permit decision region overlap by placing a pixel into the "most probable" class even when it falls into more than one decision region (Swain 1978).

Statistical classifiers usually assume a category's response pattern probability function to be normally distributed (Lillesand 1979) For remote sensing applications this assumption has been found to adequately describe most spectral response distributions. Even if the assumption is rather largely violated, the resulting classification accuracy is only slightly affected (Swain 1978).

THE MAXIMUM LIKELIHOOD CLASSIFIER

One of the most widely used statistical pattern-recognition technique is the maximum likelihood classifier. The basic procedure followed by this classifier can be divided into two general steps. First, the probability density function for each category is estimated from training areas of known identity. If the probability density functions are assumed to be normally distributed, they can be fully characterized by a category's mean vector and covariance matrix. Second,the probability of a given pixel belonging to each category is computed using the mean vectors and covariance matrices for the categories, and the pixel is assigned to the category with the highest probability (Lillesand 1979).

An important characteristic of the maximum likelihood classifier is its ability to assign every pixel in the scene to a class (Swain 1978). The obvious advantage in this is there are no pixels left unclassified and therefore more information about the scene is made available. The not so obvious disadvantage is that pixels which in reality do not belong to any of the predetermined categories are being forced into one of them by nature of the algorithm. These commission errors are usually larger for the maximum likelihood classifier than for other supervised classifiers because of the inherent "100% confidence interval" set for each category. This creates a situation where the range of acceptable values for each category is infinite (Mead 1977).

Another problem associated with the maximum likelihood classifier is that a large number of calculations need to be performed in order to classify one pixel. The intense computational factor of the algorithm greatly slows the classification process while increasing processing costs.

ACCURACY ASSESSMENT

Assessing the accuracy of a classified product is an important step in determining the usefulness and validity of that product. Various accuracy assessment techniques are available for comparing different classifications as well as different methods of classification. The range of accuracy assessment techniques can be divided into two groups: non-site specific and site specific.

Non-site specific accuracy is the oldest and one of the most widely used methods of accuracy estimation. The procedure usually begins by overlaying a classified map on a reference map and summing the total area in each category similiar to both maps. While this type of estimation is easy to perform, it provides information only about the overall agreement between the maps and not about the spatial location of the agreement and error. Where the errors are located is an important consideration in deciding the utility of the classified product (Campbell 1983).

The most common way to calculate non-site specific accuracy is to first sum the correctly classified pixels and then divide that sum by the total number of pixels classified. The result is referred to as the "overall percent correct" and is the simplest and one of the most often encountered measures of accuracy. This value provides a general estimate of the classified scene's accuracy but it does not indicate the distribution of the classification errors among the categories (Story 1986). Even though the overall percent correct provides little information in itself, Anderson (1976) states that in order for a land use and land cover map to remain useful to resource managers, the level of overall accuracy should not fall below 85%.

In contrast to non-site specific accuracy, site specific accuracy is a more complete assessment of a product's usefulness and is therefore prefered (Prisley 1982). The procedure for estimating this type of accuracy involves the comparison of a sample of classified pixels to the spatially located reference data. An error matrix with rows representing the classified categories and columns representing the reference data categories is then formed to tabulate the correct and incorrect classifications. The diagonal of this matrix displays the number of correctly classified pixels for each category while the number of commission and omission errors are located off the diagonal (Congalton 1983).

A more informative description of map quality was the classification accuracy of the individual categories. Categorical accuracy can be determined in two ways. In the first method, the number of correctly classified pixels for a category is divided by the total number of pixels from that category in the reference data. The resulting percentage is termed "producer's accuracy", and is the probability of a reference pixel being correctly classified by the classifier (Story 1986). Any misclassifications in this situation are errors of omission. The second type of class accuracy assessment is calculated by dividing the number of correctly classified pixels for a category by the total number of pixels assigned to that class by the classifier. The resulting percentage is called "user's accuracy", and is the probability of a pixel from the classified image actually representing the correct class on the ground. Pixels mis-classified under this method are errors of commission.

DATA AND PROCEDURES

DESCRIPTION OF STUDY AREA

The Great Dismal Swamp is 125,000 acres (50,590 hectares) of forested wetland located in southeastern Virginia and northeastern North Carolina. A donation of 53,000 acres (21,450 hectares) of the swamp was made to the U.S. Fish and Wildlife Service in 1973 to be managed as a National Wildlife Refuge.

Although the Great Dismal Swamp has been characterized as a forested wetland, it actually contains an extremely diverse vegetative land cover and understory. One cause for this diversity is the location of the swamp near the northern or southern limits of the ranges of many of the plant species. This situation creates an unusual environment where a wide variety of both deciduous and broadleaved and needle-leaved evergreen trees, and evergreen and deciduous shrubs, vines and herbaceous plants thrive. Recently, the swamps ecosystem has been greatly altered due to the affects of fire, timbering, ditching, road building, and changes in water availability resulting in the domination of "pioneer" plant species instead of the usual swamp climax plant species (Garrett and Carter 1977). For instance, the present vegetative communities include cypress-tupelo, maple dominated mixed hardwood stands, inkberry and bayberry shrubs, and pure stands of Atlantic whitecedar, loblolly and pond pine. These groups are often found in small parcels interspersed throughout the swamp (Kovalick 1983).

The vegetative communities of the Great Dismal Swamp are bound on the west by the Suffolk escarpment approximately 15 feet higher in elevation than the rest of the swamp, on the east by the Deep Creek swale, on the north by the Churchland flat, and on the south by the lower Pasquotank River drainage basin (Oaks and Coch 1973; Whitehead 1972). Most of the water for the swamp flows in from the west and out to the north, east, and south. Lake Drummond is located near the center of the swamp and supplies water for the operation of the the Dismal Swamp Canal (Garret and Carter 1977).

There are several reasons for choosing the Great Dismal Swamp to be the study area for this project. First, the spatial and spectral complexity of the land cover provides a thorough test of the classification ability of the maximum likelihood classifier. Second, a detailed map of the photo-interpreted land cover types was produced for 100% of the study area to serve as reference data (Figure 3). Third, results from this study can be compared to other classification studies of the same area.

The vegetation map of the Dismal Swamp reveals 43 dominant canopy types as well as 33 understory types with a classificaton accuracy for the canopy types of 93.8% (Gammon

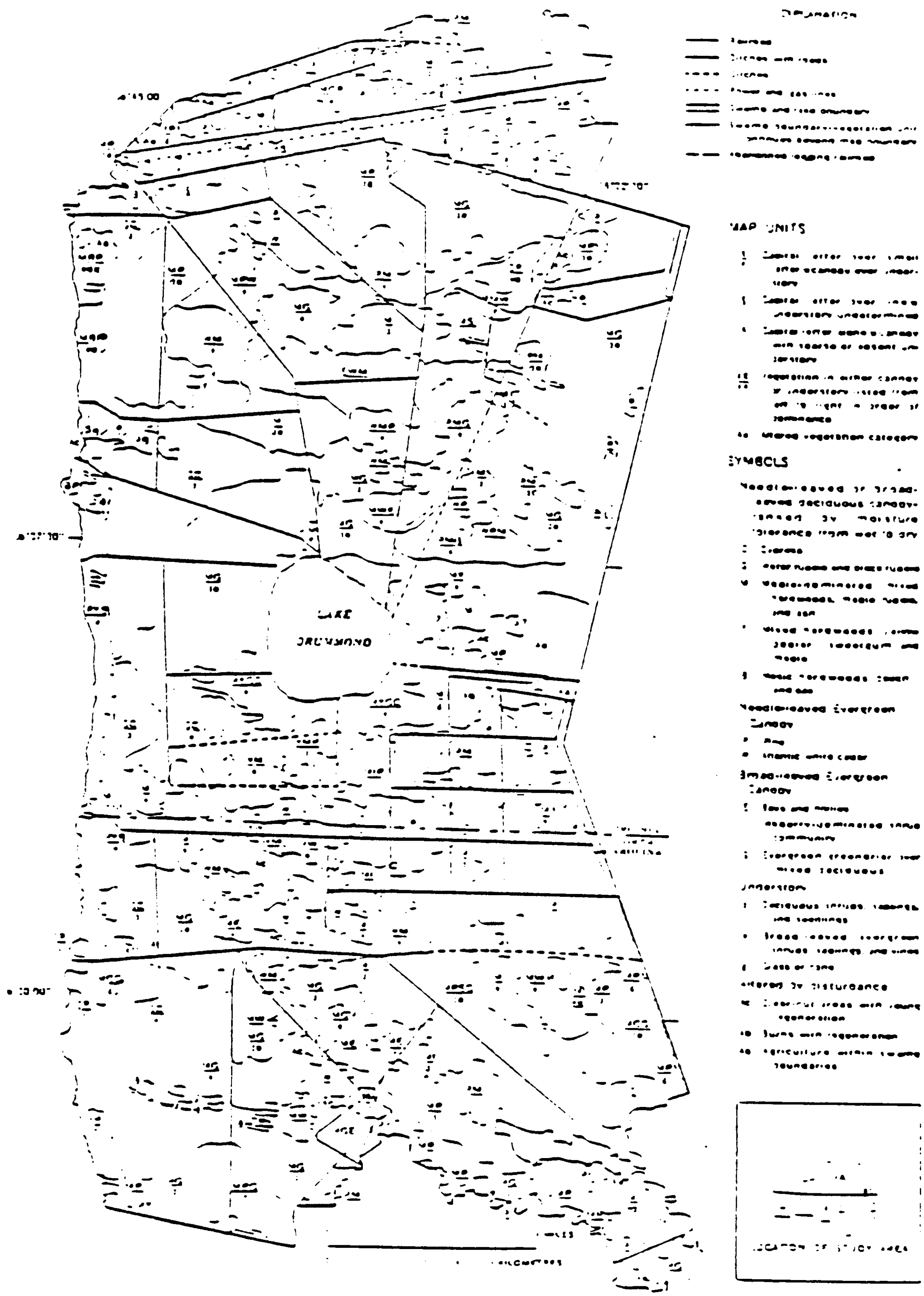


Figure 3. Vegetative Cover Map of the Great Dismal Swamp.

and Carter 1979). The map was digitized at the EROS Data Center in Sioux Falls, South Dakota and is in raster format with 810 rows and 420 columns registered to a 50-meter UTM grid. There are 340,000 total pixels in the digitized map with 233,681 pixels actually representing the swamp. Pixels located outside the swamp boundary were masked out by assigning them to zero (Kovalick 1983). The same reference data were used in classification studies performed by Prisley (1982) and Kovalick (1983), thus allowing for comparisons with this study.

The six land cover classes represented on the digitized vegetation map are listed in Figure 4. These classes were previously chosen by resource managers at the swamp for their usefulness in resource planning and were not chosen by the analyst of this study. Although detailed land cover class information was available, it would not have been practical to create a more specific class structure than an Anderson level II class structure given the coarse resolution (1.1 acres) of the Landsat data and the vegetative complexity of the swamp (Kovalick 1983).

DESCRIPTION OF SPECTRAL DATA

The spectral data to be used in this study came from the January, 1978 pass of the Landsat 2 satellite (scene ID: 21101-14331). Landsat's multispectral scanner (MSS) gathered

LEVEL II	LEVEL III
1. Coniferous evergreen	a. Pine b. Pine-deciduous mix c. Atlantic whitecedar d. Whitecedar-deciduous mix
2. Broadleaved evergreen	a. Inkberry shrub b. Evergreen vines c. Broadleaved evergreen
3. Deciduous-evergreen mix	a. Deciduous-broadleaved evergreen mix b. Deciduous-pine mix c. Deciduous-whitecedar mix
4. Deciduous	a. Deciduous with evergreen understory b. Deciduous with deciduous understory c. Deciduous (hydric species)
5. Agriculture	
6. Water	

Figure 4. Anderson Level II and Level III Land Cover Classes for the Great Dismal Swamp.

the data in four bands: green, red, near infrared, and far infrared with an approximate resolution cell size of 56 meters x 79 meters.

Preprocessing of the digital image was performed in three stages with the Interactive Digital Image Manipulation System (IDIMS) at the EROS Data Center (Prisley 1982). First,two iterations of a histogram normalization and smoothing routine were executed to restore the multi-striped image to a more natural represenation of the original scene. Second, the scene was registered to the same 50-meter UTM grid as the vegetation map using a first order nearest neighbor geometric registration and 29 geometric control points. Finally, a subscene containing only the Dismal Swamp area was taken from the Landsat scene by locating the boundaries of the swamp with the digitized version of the vegetation map and withdrawing that block from the rest of the scene.

The result was a four band destriped and resampled image of the Great Dismal Swamp made up of four matrices with 810 rows oriented east/west and 420 columns oriented north/south and a pixel size of 50 meters squared (0.62 acres). A copy of the Dismal Swamp subscene was brought to the Spatial Data Analysis Laboratory at the Virginia Polytechnic Institute and State University in Blacksburg, Virginia. Further processing was conducted using the General Image Processing System (GIPSY) on the VAX 11/785 minicomputer.

PROCEDURAL OUTLINE

The following procedure was implemented to observe the quantity and spatial distribution of pixels classified at low probabilities. To locate these pixels, a modified maximum likelihood classifier assigned pixels to a "rejected" class if their probabilities were below a user specified threshold level.

The outcome of the classification was dependent not only on the type of classifier employed, but also on the method of obtaining training statistics, so supervised training was performed in two ways. First, the statistics for the training sites were combined into their respective classes prior to classification. Second, statistics for the training sites were separated prior to classification and combined into the six general classes after classification. This was done to observe if classification accuracy could be improved by allowing more than one statistical set to represent a category.

For both the combined and individual training statistics, the category statistics were computed once and held constant for each classification of the Dismal Swamp at each threshold level. Threshold levels were varied from 0% to 95% probability of a pixel belonging to a category and were incremented as necessary to delineate a trend. By holding the category statistics constant, the differences observed between the classified images were due only to the change in threshold level.

Error matrices were obtained and analyzed at each threshold level. The spatial distribution of "rejected" pixels was observed by superimposing the classified image on the vegetative land cover map and noting the spatial location of the rejected pixels. Finally, to test the possibility of improving the accuracy of the classified image, a post-processing logical filler was used to assign the "rejected" pixels to categories based on their classified neighbors.

CLASSIFICATION PROCEDURES

The six land cover classes of the Great Dismal Swamp were represented by 52 training sites during the supervised classifications. The location, number, and size of these sites was chosen based on the spatial distribution and the spectral variability of these classes. The more spatially dispersed and spectrally variable a class was, the greater the number of training pixels were selected for that class. These parameters were empirically estimated by visually comparing the vegetation map of the Dismal Swamp to the Landsat image. To prevent the inclusion of mixed pixels in the class statistics, training sites will be kept to a size that will fit well within the boundaries of spectrally homogeneous areas (Kovalick 1983). Mixed pixels contain spectral responses from more than one class and therefore prevent the homogeneous representation of a single category (Campbell 1981).

The number of training pixels per category needed to perform a statistical classificaton such as a maximum likelihood classification can be no less than n+1, where n is the number of spectral bands in the data set. With fewer pixels, it would be mathematically impossible to calculate the variance and correlation for a category. In practice however, 10n to 100n pixels per category are usually chosen for training due to the general improvement in parameter estimates with an increase in sample size (Lillesand 1979).

MODIFICATION OF A MAXIMUM LIKELIHOOD CLASSIFIER

A Bayesian classification algorithm was chosen as the maximum likelihood classifier to be modified for use in this study. The algorithm performs a classification of a multiband image based on the assumption that the categories possess multivariate normal density functions. The input required for classification includes the multiband image to be classified along with the mean vectors and covariance matrices calculated for each category in the image from training sites of known identity or from clusters.

The mean vector and covariance matrix for each category are used in the following maximum likelihood discriminant function (category *a priori* probabilities are assumed to be equal for this study):

$$P(X|j)= \frac{1}{(2\pi)^{n/2} \, |A(j)|^{1/2}} \exp[-1/2(X-M(j))^{T} A(j)^{-1} (X-M(j))]$$

where:

- $P(X|j)$ is a multivarite normal probability function for the j-th category having a mean vector $M(j)$ and a covariance matrix $A(j)$,
- X is the pixel vector to be classified, and
- $|A(j)|$= determinant of $A(j)$.

If the *a priori* probabilities are assumed to be equal, then the chosen category, j, is the category for which $P(X|j)$ is a maximum (Cai 1983).

More commonly, the ln $P(X|j)$ is calculated instead of $P(X|j)$ which yields the following equivalent discriminant function:

$$\ln(P(X|j))= 1/2(n \ln(2\pi)+\ln|A(j)|-1/2(X-M(j))^{T} A(j)^{-1} (X-M(j))$$

where $A(j)$ is a symmetric matrix. In this equation, only the last term on the right side of the equal sign varies from pixel to pixel, thus, it is the only term that needs to be computed for each pixel. Hence, this form of the discriminant function is more efficient than the original form.

The maximum likelihood discriminant function classifies every pixel in a scene although there will be many pixels classified at very low probabilities of belonging to any of the preset categories. To circumvent this problem, the maximum likelihood algorithm was modified using a thresholding technique which compares the $\ln(P(X|j))$ of the category the pixel would normally be placed in to a user-specified minimum threshold value. If the $\ln(P(X|j))$ is less than the threshold value, the pixel will be assigned to a "rejected" class (Swain 1978). Figure 5 explains the logic of the original Bayesian classifier and the points in the algorithm modified for thresholding.

LOGICAL FILLING

A logical filling operation was performed on a scene after it was subjected to a maximum likelihood classifier. The filling operation was not expected to do well for images that did not have majority of the pixels classified. This is due to the algorithm's inability to correctly classify pixels with only a few previously classified neighbors to use in the decision process. It was, however, performed on scenes with 10% or more of the the scene classified in order to delineate a trend. This was done to test if a classification system consisting of a thresholding maximum likelihood algorithm and a post-processing logical filling operation could improve classification accuracy over a traditional maximum likelihood classification.

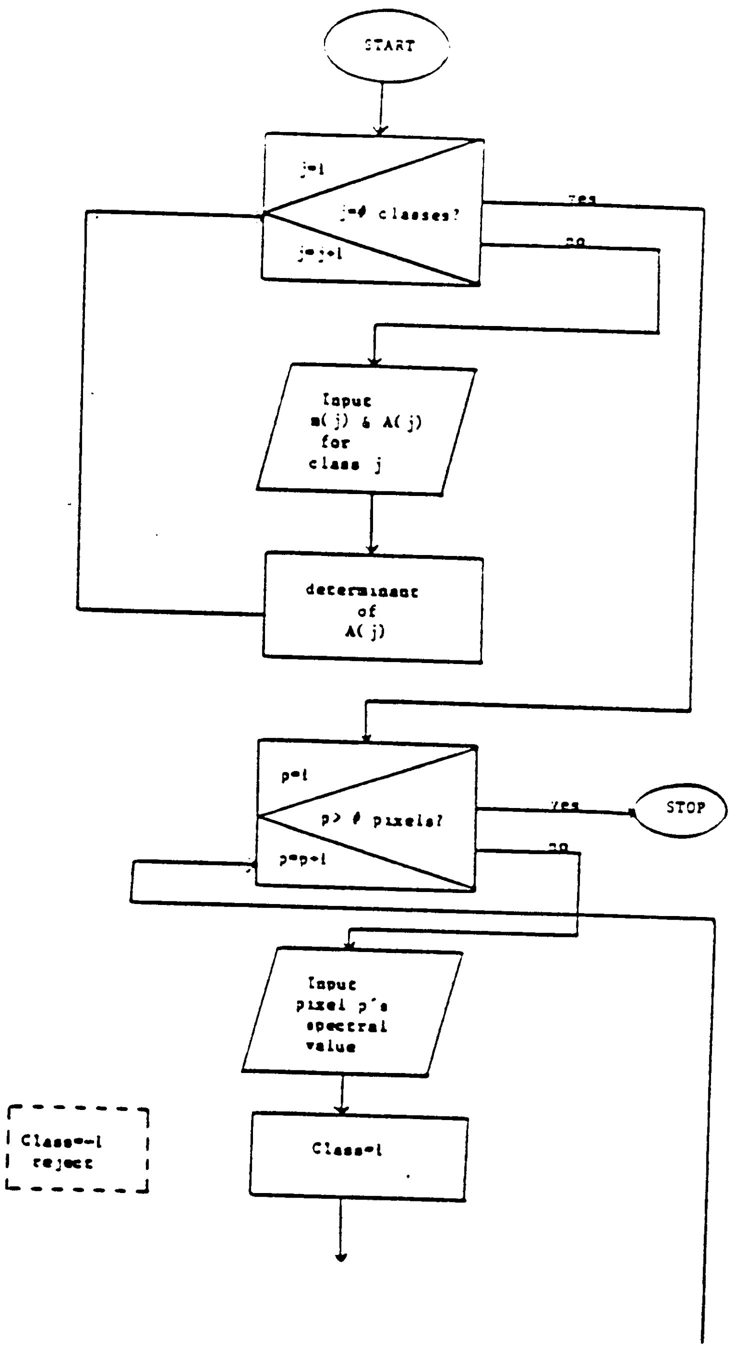


Figure 5. Flow Diagram for the Original Bayesian Classifier and Points of Modification for Thresholding.

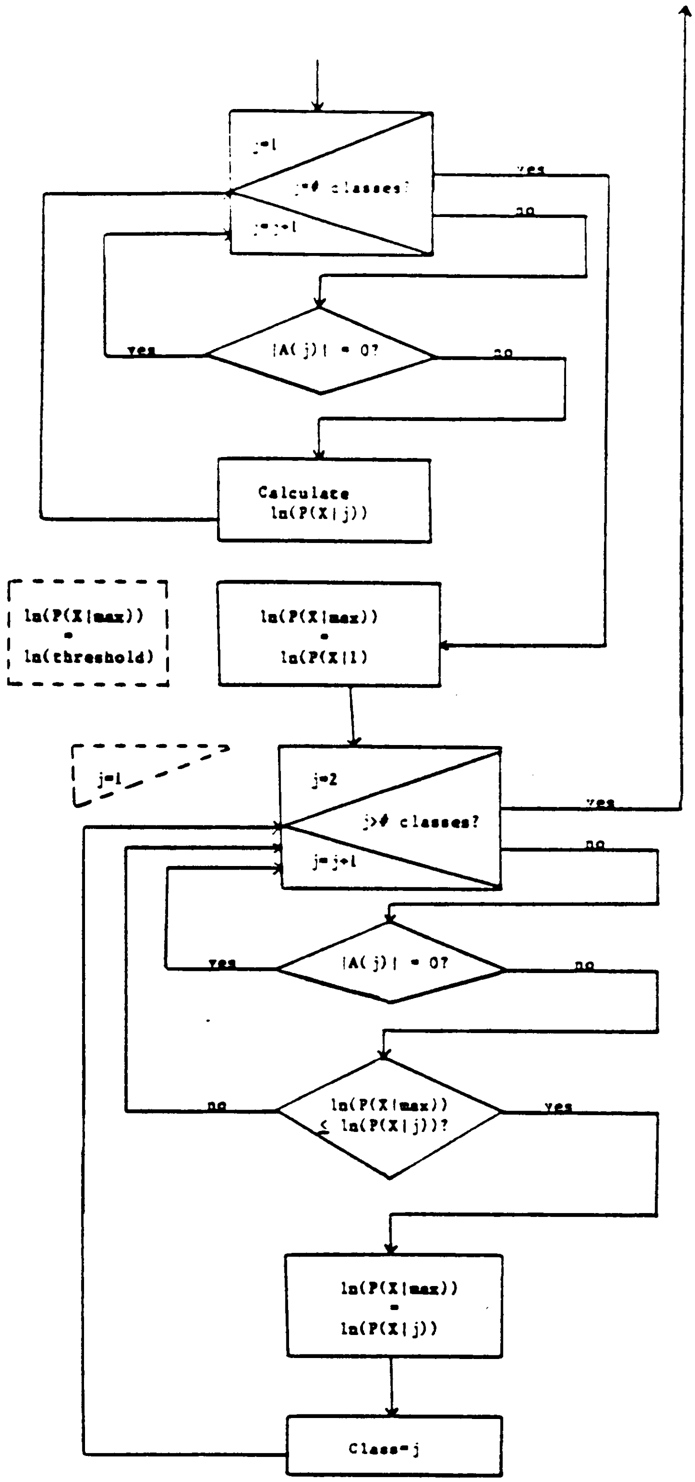


Figure 5 (continued).

The logical filling algorithm that was used for this purpose spatially classifies unclassified pixels based on a four or eight pixel neighborhood. It chooses a category label for the unknown pixel based on a majority vote of the categories in the neighborhood. The program can be repeated until all the unclassified pixels have been classified. It does not reclassify pixels that were already assigned to categories by the maximum likelihood algorithm. Options include limiting the number of iterations, alternating between 4 or 8 pixel neighborhoods at each iteration, or using either 4 or 8 pixel neighborhoods for all iterations. The eight pixel neighborhood option was chosen to allow a pixel to be classified based on input from all its neighbors.

ACCURACY ASSESSMENT

To assess the accuracy of the maximum likelihood classifications and the logically filled classifications, a quantitative comparison was made between the classified images and the digitized vegetation map of the Dismal Swamp. Since reference data were available for 100% of the classified area, accuracy assessments were computed without the need for statistical inference. An assumption was made that the vegetation map was 100% accurate.

Error matrices were created for every classification by comparing each pixel in the vegetation map to the corresponding pixel in the classified image. Error matrix construction is an important step in accuracy assessment because it condenses information about correctly classified pixels and commission and omission errors into a more easily manipulated format. The total number of rejected pixels was calculated by subtracting the number of pixels classified from the total number of pixels in the data set. Also, the program dealt with the large number of pixels involved in a classification by rounding cell entries to the nearest 10 pixels; thus, calculations involving the error matrices were only accurate to 10 pixels. In addition, the percent of the image classified was computed by dividing the number of pixels classified (the sum of the number of pixels in each cell of the error matrix) by the total number of pixels in the scene.

Overall and per class standard accuracies were calculated for the classified pixels from the scenes classified by the thresholding maximum likelihood algorithm and the scenes that were logically filled. User's accuracies for individual categories were calculated to help understand the overall accuracy of a classification. It was chosen over producer's accuracy because user's accuracy describes the reliability of a classification, which was thought to be more appropriate for this study than producer's accuracy which describes how well an area on the ground can be mapped.

The two major sections of comparison in this study were the maximum likelihood classifications and the logically filled classifications. Each section is further divided into the method used to train the maximum likelihood algorithm. Results from this study were summarized in graph and table format to observe any trends in regard to threshold level.

For maximum likelihood classification, tables were generated separately for the combined and the uncombined training site classifications to observe the effect of thresholding by observing:

1. the standard overall percent correct,
2. the user's accuracies for each class, and
3. the percent of the scene classified.

Next, the two training methods were compared by creating graphs with the x-axis again representing threshold levels and the y-axis representing the following:

1. the standard overall percent correct, and
2. the percent of the scene classified.

For the logically filled classifications, tables and graphs were also generated for the combined and the uncombined training statistics observing:

1. the standard overall percent correct, and
2. the user's accuracies for each category.

THE SPATIAL DISTRIBUTION OF UNCLASSIFIED PIXELS

The spatial distribution of the rejected pixels for each of the images classified by the thresholding maximum likelihood algorithm were inspected using two methods. First, an image containing only rejected pixels was displayed over the reference map to observe any clustering tendencies of the pixels that could suggest an unidentified category or unrepresentative training sites. Second, an image containing only boundaries between categories was created from the reference map. This boundary image was superimposed over the images of the rejected pixels to determine the percent of the rejected pixels occuring within a three pixel zone of the category boundaries. The purpose of this analysis was to test if the rejected pixels were occurring near boundaries implying that the low probability pixels were mixed pixels which inevitably appear in transition zones between categories.

RESULTS AND DISCUSSION

In order to study the effect of classifying low probability pixels on classification accuracy in a systematic manner, the analysis proceeded in the following sequence. First, statistics of the Dismal Swamp image were examined for the entire scene, and for individual categories to note how the characteristics of the scene might affect further processing. Second, the Dismal Swamp was classified several times using maximum likelihood classifier which allowed for thresholding. The majority of the analysis followed the classifications and dealt with the comparison of classification accuracies within and between training methods across various probability thresholds. After the classifications, a post-processing logical fill was performed to discover if accuracy could be improved over a traditional maximum likelihood classificaton. Finally, the spatial distribution of unclassified pixels was visually as well as quantitatively observed to provide insight into where and why the low probability pixels were occurring.

STUDY AREA AND TRAINING SITE STATISTICS

The supervised maximum likelihood classification used 52 training sites. Although the number of training sites seems large, the sites were small and well dispersed over the study area in order to better represent the spatial variation of the categories. Table 1 lists specific training site characteristics by category.

The mean and variance of the training sites averaged for each category are given in Table 2. Examination of these statistics shows the spectral signatures of the categories were similiar to one another. Observing the statistics for the Dismal Swamp scene in Table 3, the similiarity between categories is also evident in the low variabilty in spectral values throughout the 4 MSS bands. The image (before further processing) was dark, making it difficult to visually divide the area into categories. Many reasons exist for low spectral variability such as the season the image was taken in and the preprocessing techniques performed on the image. Another reason could be inherent in the definitions of the categories. For example, referring to Figure 4, the Coniferous evergreen category contains a Pine-Deciduous mix sub-category and so does the Deciduous-Evergreen mix category. In addition, the Deciduous-Evergreen mix category is comprised of Deciduous-Whitecedar stands and so is the Coniferous evergreen category. The theoretical difference between the sub-categories of two or more categories is the proportion of each species in a mix. Since the proportions are, in reality, on a continuum, an artifical boundary has to be made in order to create discrete categories. A problem

Table 1.

Training Site Summary for the Six Land Cover Categories

Land Cover Category	Number of Training Sites	Number of Training Site Pixels	Total No. of Pixels[1]	% of Total Represented by Training Sites
1. Coniferous Evergreen	14	3294	51463	6.40
2. Broadleafed Evergreen	7	1091	12002	9.09
3. Decidous-Evergreen Mix	14	3219	41279	7.80
4. Deciduous	12	6765	119061	5.68
5. Agriculture & Other	3	653	4719	13.84
6. Water	2	716	5157	13.88
Total	52	15738	233681	6.73

[1] From vegetative cover map.

Table 2.

Mean And Variance of Training Sites for Each Category By Band

Land Cover Category	Spectral Means by Band				Variance of Spectral Values by Band			
	4	5	6	7	4	5	6	7
1. Coniferous Evergreen	7.99	7.36	12.59	11.92	.09	.62	1.93	2.24
2. Broadleaved Evergreen	8.03	7.75	12.82	11.93	.03	.50	.86	1.31
3. Deciduous Evergreen Mix	8.04	7.91	12.00	11.18	.07	.29	.69	.91
4. Deciduous	8.29	8.44	10.65	9.39	.21	.31	.82	1.09
5. Agriculture and Other	9.12	10.90	14.70	13.15	.40	4.44	6.82	8.24
6. Water	7.03	6.03	4.67	.22	.03	.11	.38	.30

Table 3.

Summary of Overall Statistics for the Dismal Swamp Landsat Scene

	Band			
	4 Green	5 Red	6 IR	7 IR
Minimum Spectral Value	2	3	0	0
Maximum Spectral Value	26	39	42	38
Mean Spectral Value	8.56	9.06	12.9	11.7
Variance of Spectral Value	1.40	6.99	11.9	12.0

arises at the portions of the continuum directly on either side of the boundary because of the similiarity of spectral values in that close of a span. This creates a situation in which it is difficult to distinguish between categories if pixels fall in the overlapping areas. Since the category definitions were created by resource managers of the Dismal Swamp National Refugee, to meet their needs, they were unaltered for this study, although problems in discerning between some of the categories would surface.

COMBINED TRAINING SITE CLASSIFICATIONS

THRESHOLDING THE CLASSIFICATIONS

Using only the classified pixels, the overall percent correct for the entire scene (including training site pixels) and user's accuracies for individual categories were calculated (Table 4). One of the most interesting observations discovered during the thresholding process was that 25% of the scene was being classified at one percent probability or lower, and 75% at 10% probability or lower (Table 5). This means that a majority of the scene had a very low probability of belonging to any of the defined categories. Ultimately, it was not the probability a pixel was classified at but the accuracy of the resulting classification. The overall accuracy of the standard classification was 43%. This implied

Table 4.

Overall and Category Standard Accuracies for Only Classified Pixels at Various Threshold Levels for the Combined Training Site Method of Classification

Probability threshold (%)	Overall correct (%)	User's Accuracy by Category (%)					
		1	2	3	4	5	6
0	43	62	11	28	86	17	94
1	43	58	11	28	89	0	96
2	41	56	12	28	90	0	96
3	37	58	12	28	88	0	96
4	36	31	12	28	88	0	96
5	33	31	12	28	89	0	96
6	31	31	12	28	87	0	96
7	33	0	13	28	87	0	96
8	32	0	13	28	92	0	96
10	29	0	12	31	0	0	97
13	30	0	12	32	0	0	97
15	33	0	11	32	0	0	97
17	32	0	11	30	0	0	97
20	32	0	11	30	0	0	97
23	40	0	0	30	0	0	97
25	97	0	0	0	0	0	97
30	97	0	0	0	0	0	97
40	97	0	0	0	0	0	97
80	96	0	0	0	0	0	97

Table 5.

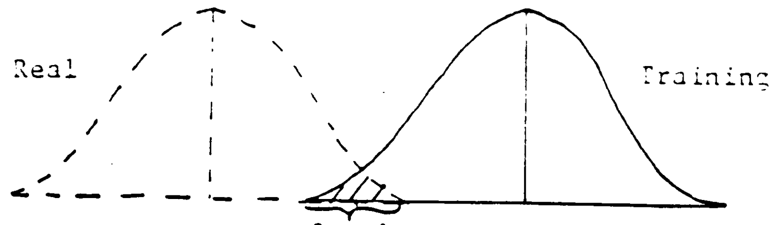
Percent of Pixels Classified Above Threshold Level, Overall and by Category, at Each Threshold Level for the Combined Training Site Method of Classification

Probability threshold (%)	% of pixels classified overall	% of pixels classified by category					
		1	2	3	4	5	6
0	100	100	100	100	100	100	100
1	75	61	92	98	70	0	79
2	66	50	83	95	53	0	76
3	61	48	81	93	37	0	76
4	53	10	65	93	35	0	71
5	50	10	65	91	25	0	71
6	48	10	65	91	19	0	71
7	44	0	56	91	19	0	71
8	38	0	41	91	12	0	71
10	25	0	28	66	0	0	66
13	18	0	23	44	0	0	66
15	16	0	13	44	0	0	66
17	13	0	13	31	0	0	66
20	13	0	13	31	0	0	66
23	9	0	0	31	0	0	66
25	1	0	0	0	0	0	66
30	1	0	0	0	0	0	66
40	1	0	0	0	0	0	66
80	1	0	0	0	0	0	28

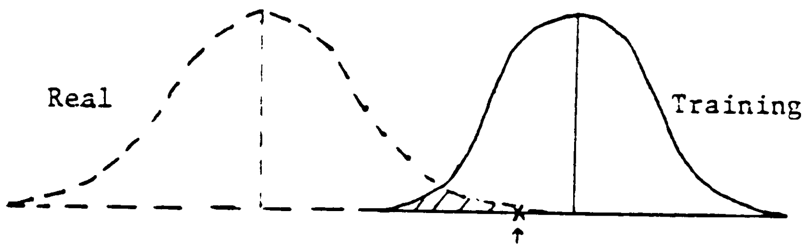
that a pixel could be correctly classified 43% of the time even though it had a much lower probability of belonging to the correct category during the classification. In other words, the actual probability a pixel was classified with was related to the accuracy of the classification but was not the same value. The actual values for the overall and category accuracies were low probably due to the similiarity in spectral signatures between categories as indicated in the low spectral variance of the scene.

Another interesting discovery was in the trend of the overall percent correct which first dropped as threshold level increased and later rose. The original hypothesis was that the overall percent correct would continually rise with an increase in threshold level because as the threshold level increased, the probability of correctly classifying a pixel would also increase. The reason for this was if the pixel was classified with a high probability, the reliability of the classification should also be high. An understanding of the trend came by observing the user's accuracies for the individual categories. In Table 4 it can be seen that the accuracy of category 1 (Coniferous evergreen) dropped with an increase in threshold, while the remaining category accuracies were fairly constant. Since approximately 22% of the scene was in category 1 (Table 1), its user's accuracy strongly influenced the overall accuracy; therefore, when its accuracy fell, so did overall accuracy. Overall accuracy did begin to climb at the 7% threshold because the number of pixels classified into category 1 was zero which removed its influence over overall accuracy (Table 5). The overall accuracy was also influenced by category 4 (Deciduous), which occurs in 50% of the classified scene. It dropped out of the classification at the 10% threshold level, and since it had a high accuracy, when it was dropped, the overall accuracy fell. The greatest jump in the overall accuracy was after the 23% threshold level when all but one of the categories had been dropped from the classification, and the overall accuracy became the user's accuracy of the remaining category (category 6 - Water) which was 97% (Table 4).

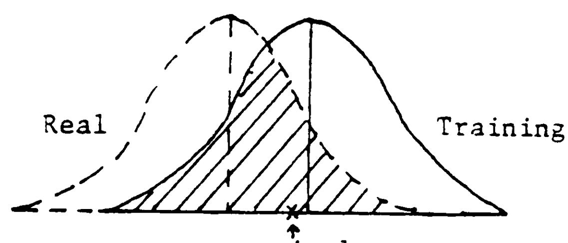
This explains the flucuations in the overall accuracy for the combined training site classifications, but it does not explain why the user's accuracy for category 1 fell. A possible answer lies in how close a category's training site distribution is to a category's real distribution. If the training site statistics are not close to the population statistics, the two distributions will overlap by a small amount. When the distributions overlap by only a small amount, the overlap is only in the tails of the distributions where probability is low. Since a pixel can only be correctly classified if it falls into the area of overlap, correct classification can only occur (in this case) at low training site probabilities (Figure 6a). In addition, the overlap might not include the area of the training distribution as-



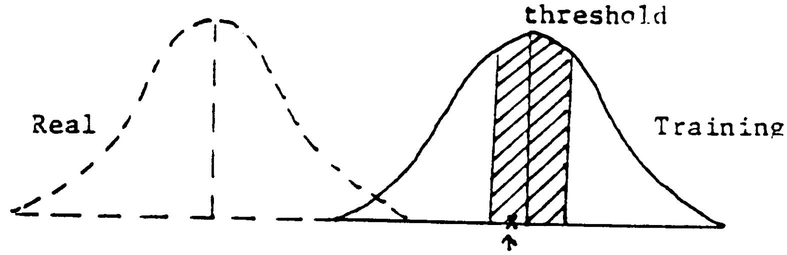
(a) **Training distribution and real distribution not similar so overlap only in tails.**



(b) **Pixel has high probability of belonging to training site distribution but low probability of belonging to real distribution.**



(c) **Pixel has high probability of belonging to both the training site distribution and the real distribution although probabilities are not equal.**



(d) **In shaded area, pixels classified into category although probability it belongs to category is low because far from the real distribution.**

Figure 6. The Training Site Distribution of a Category Compared to the Real Distribution of that Category.

sociated with high probabilities of correct classification. Even if a pixel has a high probability of belonging to the training distribution, it might not have the same probability of belonging to the real distribution (Figure 6b). That is to say, even though only pixels with high probabilities of belonging to the training site distribution are being classified, the reliability (user's accuracy) of those pixels is not necessarily high also. This is because the reliability of a pixel's classification is based on how close the pixel is to the real distribution, not how close it is to the category's training distribution. If the real and training distributions are very similiar, then a pixel classified with a high probability in the training site distribution will also have a high probability in the real distribution (Figure 6c).

The reliability of the classified pixels can be improved as well as degraded by thresholding the training distribution. As the threshold level increases, more of the area in the tails of the training distribution is removed; hence, some of the area of overlap between the real and training distributions is also removed. If the training is poorly performed, and the overlap is only in the tails of the two distributions, disallowing the classification of the low probability pixels (as defined by the training distribution) decreases the overall reliability of the category. This is because the pixels which are left are far removed from the real distribution and therefore have a low reliability of correct classification although they are in the high probability area of the training distribution (Figure 6d). In this case, increasing the threshold level actually decreases the overall reliability of the classified pixels by eliminating pixels with greater reliability which upgrade the average reliability of the category.

This is what is believed to have occurred during the classification of category 1. Its user's accuracy fell when the threshold level rose because less of category 1's real distribution overlaped with the training distribution. The pixels left were far removed from the real distribution because they had to be centered closer to the mean of category 1's training distribution to be classified. The remaining pixels's reliability continued to fall as the threshold level increased until at the 7% threshold, reliability became 0%. The reason the training distribution for category 1 was not close to its real distribution was probably a result of combining the training sites for the category which were spectrally different due to spatial location. This would yield a training distribution unlike any of the real distributions for category 1.

If classifier training had been performed better, and the training distribution had coincided with the real distribution more closely, then setting a threshold level on the classification should improve the reliability of the classified pixels because the threshold would now be in approximately the same position on the real distribution as on the training distribution. In this situation, if a classified pixel has a 95% chance of being correctly classified according to the training distribution, it has close to the same probability of being correctly classified in the real distribution (Figure 6c). An example of this situation occurred in category 6 (Water), where the reliability of the classification increased as threshold level increased (Table 4). If they had overlapped by a small amount, a 95% probability in the training site distribution would correspond to a much different probability in the real distribution (Figure 6b).

LOGICAL FILL

An attempt was made to improve the classification accuracy of the maximum likelihood classifier by first setting the minimum probability a pixel had to exceed to be classified and to then spatially classify the unclassified pixels with a logical filler. As Table 6 exhibits the only improvement in overall accuracy was at the 1% threshold level. The individual category accuracies at this level either dropped or remained the same so the increase in overall accuracy from 43% to 44% was probably due to rounding errors. Given that rounding errors did occur, there was no practical difference between the original classification and a 1% threshold and fill. For individual category accuracies,

Table 6.

Overall and Category Standard Accuracies for Only Classified Pixels at Various Threshold Levels for the Combined Training Site Method of Classification Followed by a Logical Fill

Probability threshold (%)	% of original scene unclassifed	Overall correct (%)	User's accuracy by category (%)					
			1	2	3	4	5	6
0	0	43	62	11	28	86	17	94
1	25	44	62	10	27	84	0	93
2	34	43	64	10	27	86	0	93
3	39	41	64	10	26	86	0	93
4	47	39	45	11	25	85	0	93
5	50	35	45	10	23	88	0	92
6	52	34	45	10	22	89	0	92
7	56	34	0	10	23	88	0	92
8	66	32	0	11	22	91	0	92
10	75	*	-	-	-	-	-	-
13	82	17	0	9	19	0	0	74
15	84	*	-	-	-	-	-	-
17	88	17	0	11	30	0	0	97
20	88	17	0	9	18	0	0	73
23	91	*	-	-	-	-	-	-
25	99	*	-	-	-	-	-	-
30	99	*	-	-	-	-	-	-
40	99	*	-	-	-	-	-	-
80	99	*	-	-	-	-	-	-

*Fill not performed for this threshold level.

there can occassionally be found thresholds where there was an improvement in accuracy over the standard classification but there was no one threshold level that did so for all the categories at once.

A problem arises when a certain threshold level is reached which inhibits the classification of all the pixels in a category. The logical fill algorithm, which classifies a pixel based on the spatial neighbors of that pixel, cannot redefine a missing category. If a category is missing from a classification, it will not be included in the logically filled classification either. An example of this happened when category 5 (Agriculture and other) was subjected to a 1% threshold and was entirely omitted from the classification and it was thus ommitted from the logically filled classifications that followed. Overall accuracy and category accuracies made the most dramatic fall after the percentage of the scene needing to be filled rose above approximately 40%. At this level, there are too few pixels surrounding most of the unclassified pixels to allow for a sound decision to be made on a pixel's identity.

In summary, increasing the probability a pixel must have in order to be classified does not consistently increase classification accuracy. This is due to the difficulty associated with locating training sites that are highly accurate representations of the real category distribution. Also, spatially classifying unclassified pixels after a thresholding does not improve accuracy.

SEPARATED TRAINING SITE CLASSIFICATIONS

THRESHOLDING THE CLASSIFICATIONS

The overall percent correct was calculated for the scene as well as individual category accuracies (user's accuracy) for a sampling of threshold levels for the separated training site classifications as for the combined training site classifications (Table 7). Fewer separated training site classifications were performed than for the combined training site classifications because the processing time needed to perform one separated training site classification was approximately six times greater than for the combined training site classifications. The threshold levels at which the classifications were performed did yield enough information to delineate a trend.

The probability the pixels were classified with were higher for the separated training site classifications than for the combined training site classifications (Table 7). Approximately 25% of the scene was classified at a 5% probability or less, and 75% of the scene was classified at a 90% probability or less (Table 8). The probabilities are greater for the separated training site classifications because the

Table 7.

Overall and Category Standard Accuracies for Only Classified Pixels at Various Threshold Levels for the Separated Training Site Method of Classification

Probability threshold (%)	Overall Correct (%)	User's Accuracy by Category (%)					
		1	2	3	4	5	6
0	50	47	12	27	87	30	94
5	51	47	13	28	89	0	96
10	50	48	12	27	89	0	97
15	51	48	14	27	89	0	97
30	51	49	15	27	90	0	97
60	45	36	14	27	90	0	97
80	40	26	14	27	93	0	97
90	43	0	0	27	93	0	97
95	40	0	0	27	92	0	97

Table 8.

Percent of Pixels Classified Above Threshold Level, Overall and by Category at Each Threshold Level for the Separated Training Site Method of Classification

Probability threshold (%)	Pixels classified (%)	% of pixels classified by category					
		1	2	3	4	5	6
0	100	100	100	100	100	100	100
5	77	76	79	81	76	0	69
10	51	64	59	80	65	0	65
15	50	50	35	71	52	0	65
30	48	37	23	71	45	0	64
60	47	25	14	68	29	0	57
80	35	20	14	68	23	0	27
90	26	0	0	68	19	0	27
95	24	0	0	68	14	0	27

training distributions were more accurate representations of the real category distributions than in the combined training site classifications. This resulted in a greater number of pixels in the higher probability area of the training distributions. The classification accuracy for a standard classification was 50% which is still low. Keeping training sites separate did create more accurate training sites than when combined but it was still difficult to correctly classify a pixel when the real distributions of the categories were similiar. The probabilities with which the pixels were being classified at were greater for the separated training site method, but it was still evident that pixels were being classified at lower probabilites than the classification accuracies would suggest. There was a relationship between the probability a pixel was classified with and the reliability of the classification, but the relationship was not direct.

Table 7 shows the trend of the overall percent correct over threshold level. It remained fairly constant until the 30% threshold level when it began to fall. The reason for the fall was the same as for the combined training site classifications; category 1's user's accuracy declined after the 30% threshold level while categories 2 through 6 remained constant. Until the 30% level, the overall percent correct is approximately 50%, and after this point it declined. Observing the percent of the pixels classified into category 1 across threshold level, category 1 dropped after the 80%

threshold from the classification so its accuracy influenced the overall accuracy for most of the classifications (Table 8).

It is interesting to note that the user's accuracy for category 1 did not become zero until after the 80% threshold level whereas it did so after the 6% threshold level in the combined training site method. This indicates that the training distributions representing category 1 overlapped the real distributions more so when using the separated training site classifications than when using the combined training sites (category 1 must of had many modes in reality). The same was true for the other categories where the user's accuracy did not fall to zero until a much later threshold than in the combined training site method. The threshold level at which user's accuracy became zero occurred when the section of the training distribution remaining after thresholding, was so far removed from the real distribution that the pixels that were classified had a reliability of zero.

Although the training distributions overlapped the real distributions more so for the separated training sites than for the combined training sites, the user's accuracies were still low for some of the categories. The reason for this was that even if a training distribution overlaps a real distribution to a great extent, the accuracy of the pixels could still be low because different categories could , in reality, have similiar spectral signatures or similiar training site statistics which could cause confusion during classification and reduce the reliability of the classified pixels. This seemed to be the case for categories 2 and 3 where the pixels classified spanned almost the entire range of threshold levels but the user's accuracy was still low.

For a category in general, if pixels are being correctly classified into this category for a large range of threshold levels, then the training site distributions for this category overlap the real distribution to a great extent. If the distributions do overlap but the accuracy of the classified pixels is still low, the category's real distribution must be similiar to another category's real distribution or training distribution causing confusion between the categories during classification (Figure 7).

LOGICAL FILL

The logical fill operation was also applied to the classifications produced using the separated training sites to see if accuracy could be improved over the standard maximum likelihood classification. Table 9 displays the results of the fills. It reveals no increase in accuracy except for at the 10% threshold level. This was most likely caused by rounding errors since the individual category accuracies were never greater than the corresponding accuracies in the original (0% threshold) classification.

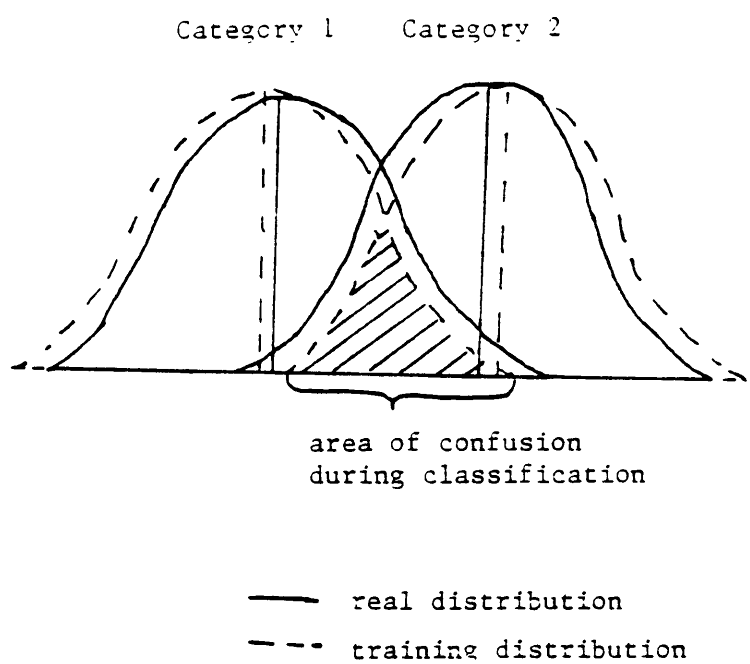


Figure 7. The Similiarity Between Two Categories Resulting in Confusion During Classification Even With Represenative Training Sites.

Table 9.

Overall and Standard Accuracies for Only Classified Pixels at Various Threshold Levels for the Separated Training Site Method of Classification and Logical Fill

Probability Threshold (%)	% of original scene unclassified	Overall correct (%)	User's accuracy by category (%)					
			1	2	3	4	5	6
0	0	50	47	12	27	87	30	94
5	23	50	45	12	27	86	0	93
10	49	51	47	10	27	83	0	93
15	50	50	46	11	25	82	0	93
30	52	48	44	12	24	80	0	93
60	53	47	44	9	23	83	0	94
80	65	*	-	-	-	-	-	-
90	74	*	-	-	-	-	-	-
95	76	*	-	-	-	-	-	-

*Fill not performed for this threshold level.

In summary, the separated training sites better represented a category's real distribution than did the combined training sites; however, confusion still existed between the categories resulting in low classification accuracies. This confusion was due to the similarity between the categories spectral signatures. Also, the thresholding maximum likelihood classifier and the spatial classifier did not improve classification accuracy over a standard maximum likelihood classification by a practical amount.

SEPARATED VERSUS COMBINED CATEGORY TRAINING

Comparing the separated and the combined training methods, it was discovered that the overall percent correct for the separated training site method of classification at the 0% threshold level was approximately seven percentage points greater than for the combined training site method of classification. For the range of threshold levels, the overall percent correct for the separated training site classifications was fairly constant, whereas the overall percent correct for the combined training site classifications was irregular (Figure 8).

The irregularity of the overall accuracy for the combined method was a result of the categories dropping out of the classification when a threshold level became too high. If the accuracy of the category prior to its drop was low,

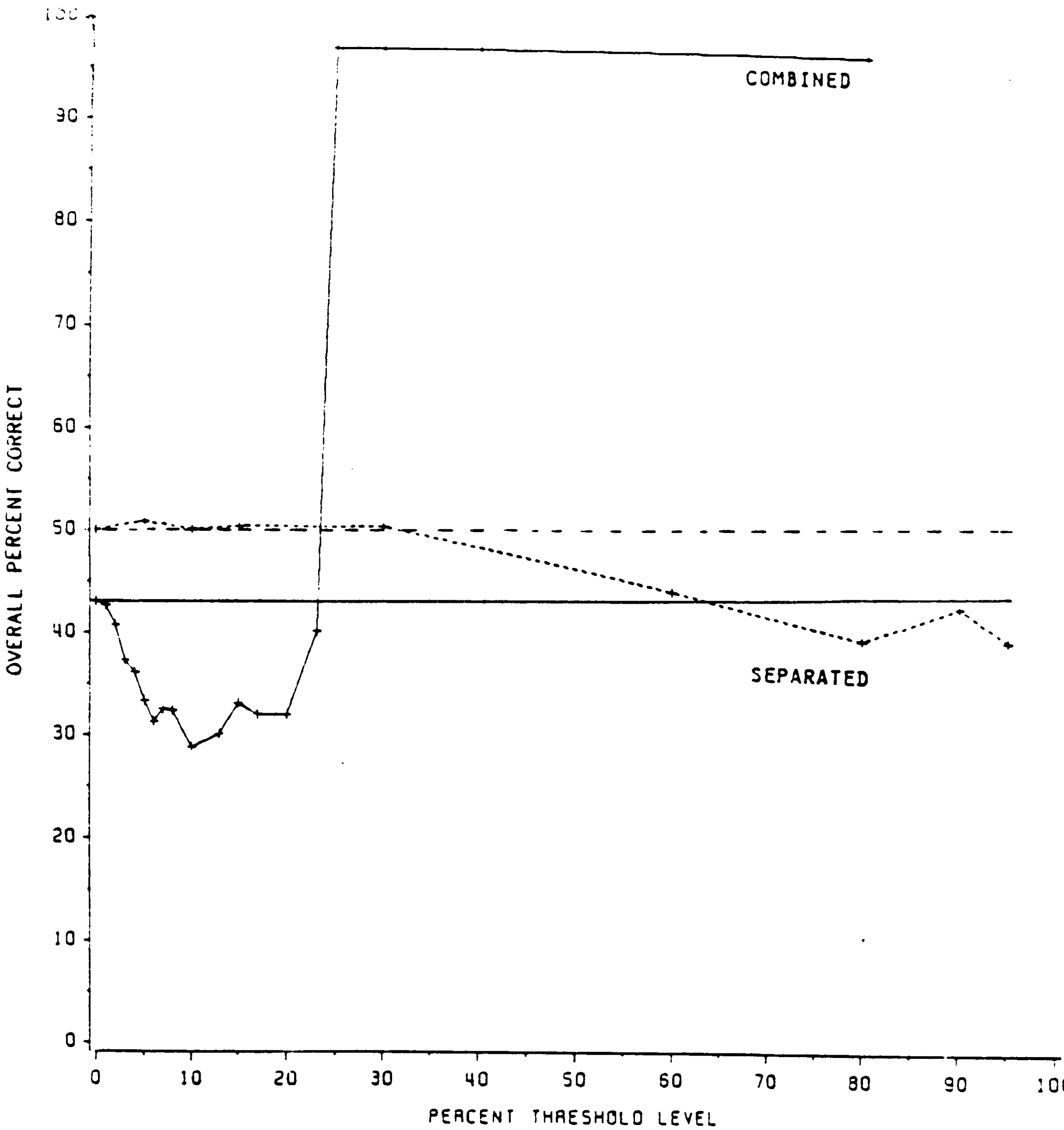


Figure 8. Overall Accuracies for Separated and Combined Training Site Threshold Classifications.

the overall accuracy of the scene would rise when it was eliminated. If the accuracy of the category prior to its drop was high, the overall accuracy of the scene fell after elimination. The amount of increase or decrease in overall accuracy depended on the percent of the entire scene the category was a part of; the more it occupied, the more influence it had on the overall accuracy. For example, after the 23% threshold level in the combined training site classifications, all the categories were dropped from the successive classifications and the overall accuracy of the classifications rose to 97% which was the user's accuracy of the only remaining category, category 6 (Water) (Figure 8).

Since categories did not drop out of the separated training site classifications until after the 80% threshold level (except category 5) (Table 8), the overall accuracies were only affected by fluctuations in the accuracy of individual categories. Individual category accuracies were fairly constant; hence, so were the overall accuracies (Figure 8).

Individual category accuracies at the 0% threshold level for categories 2, 3, 4, and 6 were similiar for both methods of training; however, category 5 had an accuracy 13% greater in the separated training site classification than in the combined classification, and category 1 had an accuracy 15% less in the separated classification than in the combined classification. The overall percent accuracy was improved by keeping the training sites for a category separated until after a classification, but confusion still existed between the spectral signatures of categories 1, 2, 3, and 5 as indicated by their low user's accuracies (Tables 4 and 7).

The greatest difference between the separated and combined training site classifications was that the user's accuracies for the categories in the separated classifications did not fall to 0% until after the 80% threshold level or later (except for category 5) while, in the combined classifications, they fell to 0% at approximately the 7% and 23% thresholds (except for category 5) (Tables 4 and 7). This indicates that the training site distributions for the categories in the combined training site method were not as close to the real distributions as the separated training site distributions were.

Support for this hypothesis also came from the comparison of the percent of the scene classified at various threshold levels between the methods of training. It is clear from viewing Figure 9, the number of pixels classified at each threshold level was increasingly greater, until approximately the 40% threshold, for the separated classifications than for the combined classifications. Since the training site distributions for the categories in the combined classifications were not close to the real distributions, classification into the category could only occur when a pixel was inside the section of the training distribution above the threshold. Because the cloud of pixels for a category was far from the section of the training distribution above the threshold, there were few pixels (if any) that could be classified at these higher threshold levels (Figure 10a). The training sites for the separated training site distributions were closer to the clouds of pixels for the categories so there were more pixels at higher threshold levels than for the combined training site distributions (Figure 10b).

SPATIAL DISTRIBUTION OF UNCLASSIFIED PIXELS

For each classified scene, the percentage of unclassified pixels appearing within 1 pixel on either side of the boundaries between categories (3 pixel boundary width) was calculated. By visually inspecting the combined training site classifications at various threshold levels, the unclassified pixels were observed at the 1% threshold to be located mostly on the perimeter of the swamp. At the 4% threshold, clumps of unclassified pixels were forming within the category boundaries. These clumps caused the percentage of unclassified pixels on the boundaries to decrease by increasing the number of internal unclassified pixels. The trend is illustrated in Figure 11.

Most of the clumping was located inside the Deciduous class (category 4) for both training method classifications. This was probably due to the definition of the category which resulted in its confusion with categories 1 and 3. All three

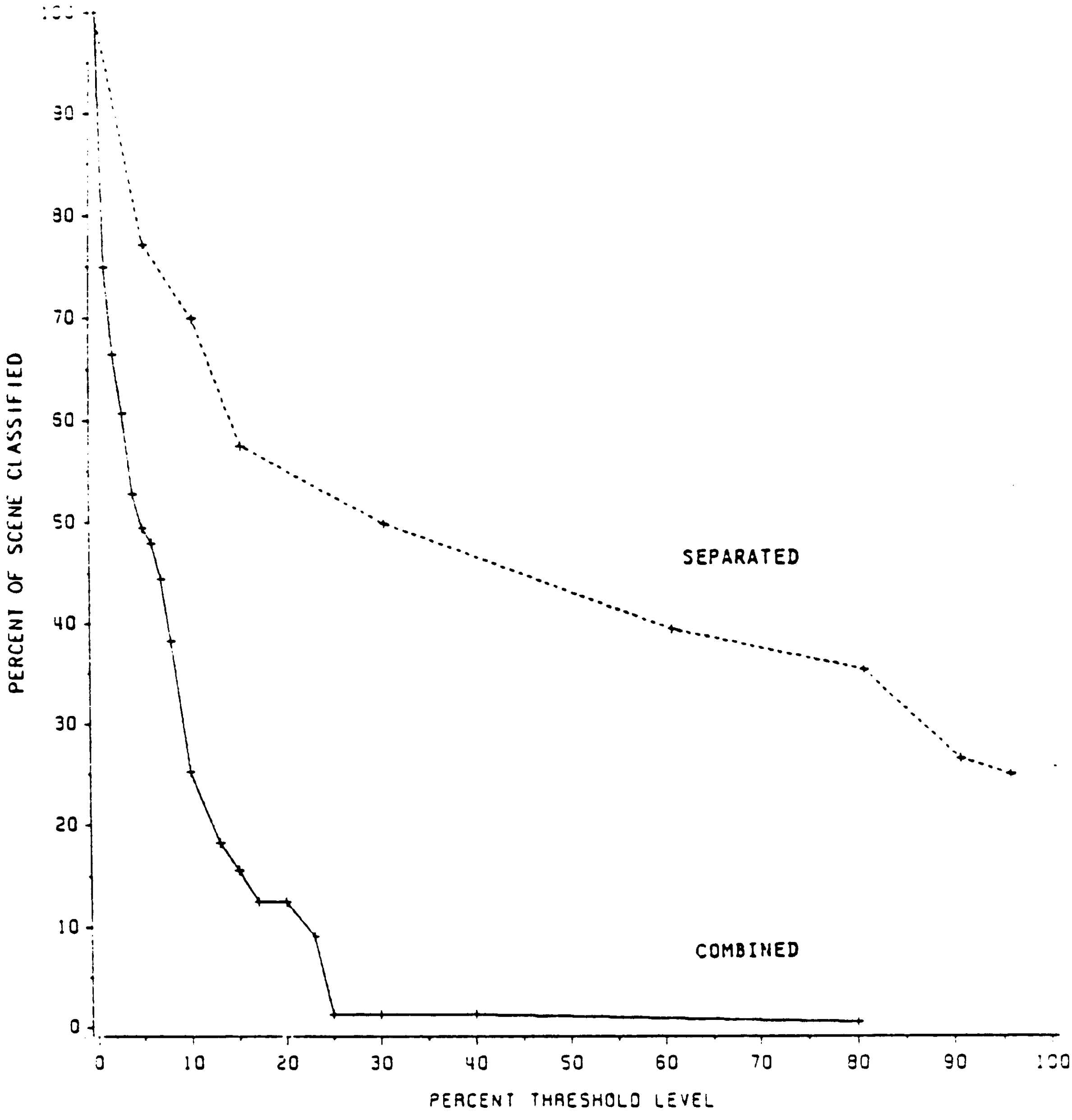
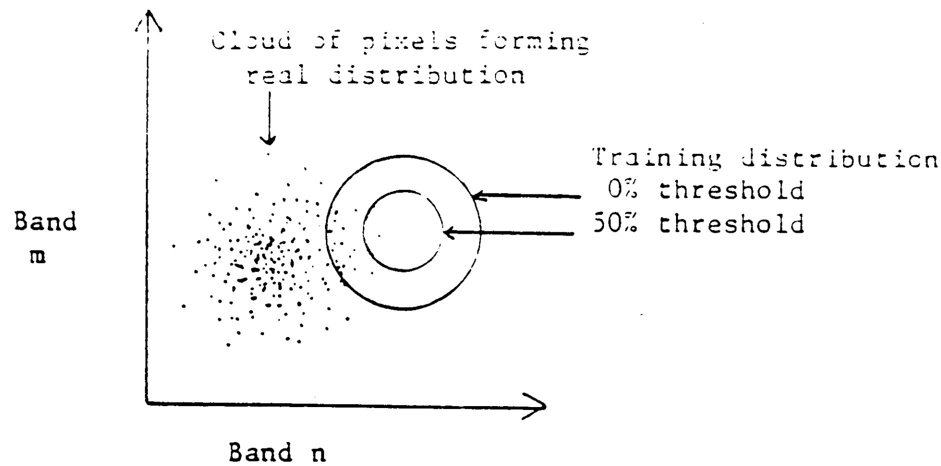
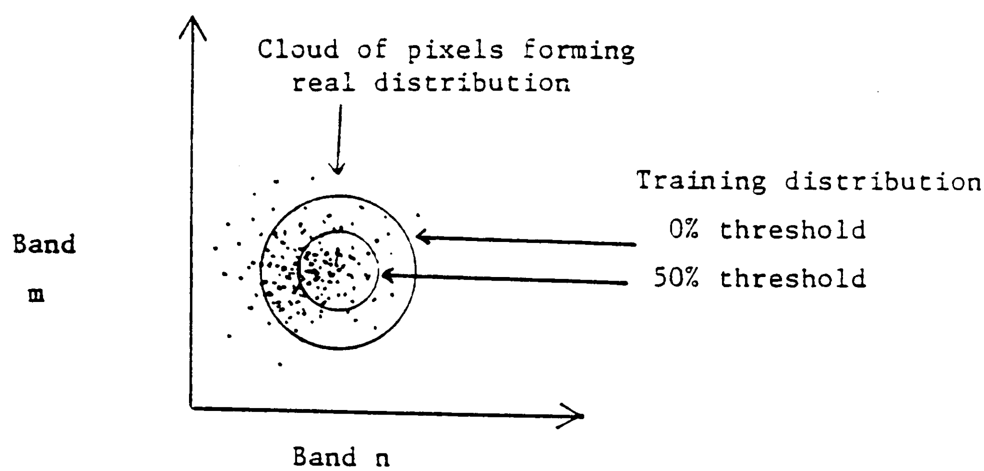


Figure 9. Percent of the Scene Classified for the Combined and Separated Training Site Classifications.



(a) Training distribution not close to real distribution so in the training distribution, very few pixels are in the high probability area.



(b) Training distribution for a category is close to real distribution so when threshold the training distribution, there are pixels in high probability area but in decreasing numbers.

Figure 10. Thresholding the Training Distribution of a Category.

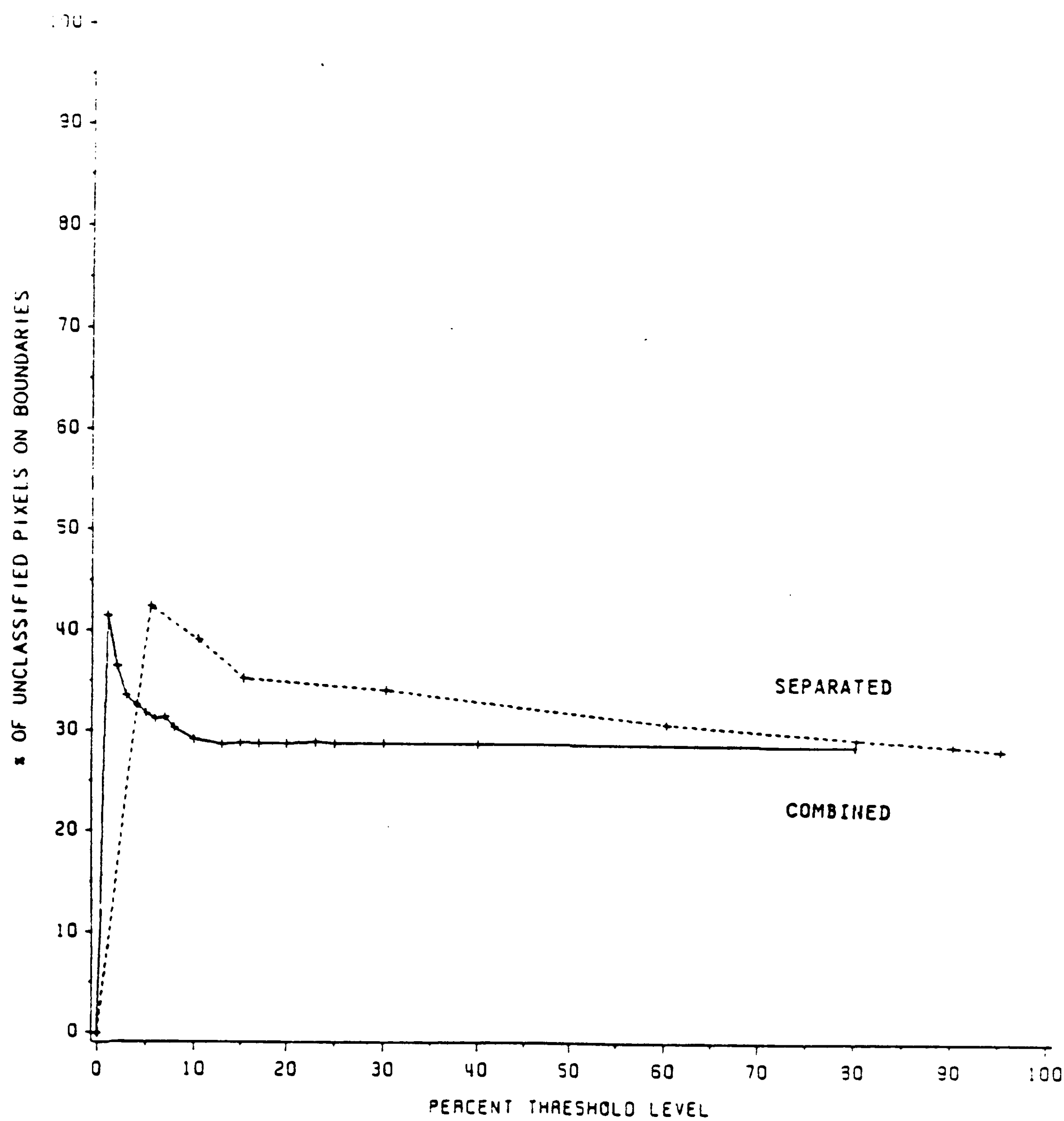


Figure 11. Percent of Unclassified Pixels on the Boundaries for Combined and Separated Training Site Classifications.

categories have Deciduous mix sub-categories. The clumps probably were areas which were assigned to category 4 when the reference data was created but were, in reality, similiar in vegetative composition to categories 1 and 3. Since they were similiar in composition, their spectral signatures were also similiar which caused confusion between the three categories during classification.

The separated training site classifications followed the same trend as the combined training site classifications except that the clumping of unclassified pixels for the separated classifications was not as pronounced in the early threshold levels (Figure 11). The reason for the lower level of clumping in the early thresholds, was that the spatial variation of the categories was better represented by separating the training sites for a category than by combining the training sites before the classification so there were less areas where groups of pixels did not fit well into any category.

In addition to the calculation of the percentage of unclassified pixels on the boundaries, the percentage of boundary pixels that were unclassified was also calculated. This value represents only the boundary pixels that were not classified, whereas the percentage of unclassified pixels on the boundaries represents the total number of unclassified pixels on the boundaries. The percentage of boundary pixels that were unclassified was calculated to isolate and observe the effect mixed pixels, which occur by nature on the boundaries between categories, have on the probability they were classified at. The trends for both training methods are given in Figure 12.

For both the training methods, it can be seen that over one-half of the boundary pixels were being classified at four percent probability or lower, and approximately three-fourths were classified at 10% probability or lower. The boundary pixels were being classified at lower probabilities than the rest of the scene was indicated by the falling percentage of unclassified pixels on the boundaries and the rising percentage of boundary pixels that were unclassified (over increasing threshold level). Since the mixed pixels (boundary pixels) had lower probabilities of belonging to the defined categories, they were removed from the classification at low thresholds thus rapidly increasing the percentage of boundary pixels that were unclassified and decreasing the percentage of unclassified pixels on the boundaries.

In general, the presence of mixed pixels on the boundaries between categories resulted in pixels being classified at lower probabilities than the internal pixels. This was due to the vegetative non-homogeneity of the mixed pixels causing unusual spectral responses for a category. The internal pixels were more homogeneous in vegetative composition; hence, so their spectral values had higher probabilities of belonging to a category.

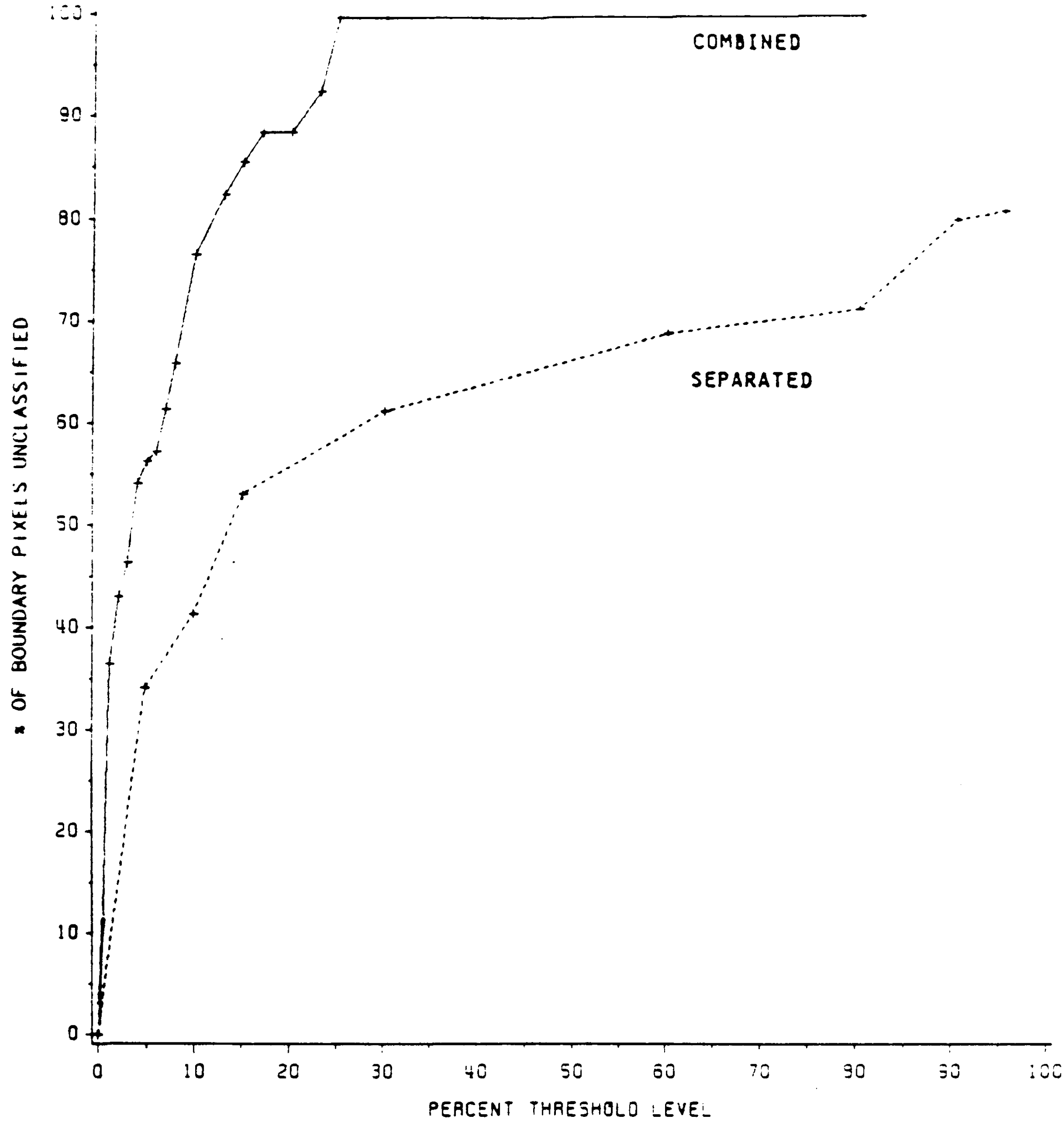


Figure 12. The Percentage of Boundary Pixels Unclassified.

It was noted previously that the ability of a pixel to be classified at a high threshold level did not necessarily correspond to a high reliability for that pixel due to the difference between its training distribution and its real distribution. Thus, the spatial distribution of the pixels at various levels of classification reliability cannot be inferred from the threshold levels at which the pixels were unclassified, unless there can be complete confidence in the accuracy of the training sites.

SUMMARY AND CONCLUSIONS

The purpose of this research was to determine the effect of thresholding a maximum likelihood classifier on the number of pixels classified in a Landsat scene of the Dismal Swamp National Refuge in Virginia-North Carolina , the accuracy of the classified pixels, and the spatial distribution of the unclassified pixels. Also, an attempt was made to improve classification accuracy over a traditional maximum likelihood classification by spatially classifying the unclassified pixels from the thresholded maximum likelihood classifications.

The procedure involved the classification of the Dismal Swamp scene at various threshold levels using statistics from training sites that were combined by category for one set of classifications and kept separate until after the classification for another set of classifications. An error matrix was created for each classification to determine the accuracy of the classified pixels, and the scene was compared to a boundary map to determine the spatial distribution of the unclassified pixels. The unclassified pixels were then spatially classified using a logical fill algorithm and an error matrix was again created to observe any changes in classification accuracy from the 0% threshold classification.

The accuracy assessment of the classifications was performed by calculating from error matrices, the overall percent correct for the entire scene and the user's accuracy (reliability) for each category. This was done for both the classified pixels from the thresholded classifications and for the entire scene after the logical fills. It was found that, in general, the majority of pixels in the Dismal Swamp scene were classified at very low probabilities, but the accuracy of the classification was much higher than the probability a pixel was classified at would predict although the two values were related. In addition, the separated training site method of classification had a greater overall accuracy as well as a greater number of pixels being classified at higher probabilities than did the combined training site classification method. This suggests that the probability at which a pixel was classified at might be correlated to the accuracy of its classification, but the values are not the same. More specifically, the results based on the combined and separated training site classification methods indicated that the reliability of the classified pixels could increase or decrease with an increase in the threshold level placed on a classification. This happened because the reliability of a pixel's classification was determined by the probability that pixel had of belonging to the population (real) distribution of its category, and not the probability it had of belonging to the category's training site distribution.

If the training site distribution and real distribution for a category coincided, then a high probability of classification accuracy as set by the training distribution, would correspond to a high probability of classification accuracy in the real distribution. In this case, setting a threshold level on a classification could improve the reliability of the classified pixels.

On the other hand, if the training site distribution did not coincide with the real distribution, then an increase in threshold level would decrease the reliability of the classified pixels. This would happen because the pixels centered around the mean of the training distribution would be farther away from the mean of the real distribution than the pixels at the tail of the training distribution. In this situation, the pixels at the tail of the training distribution actually have a greater reliability than the pixels toward the center of the training distribution because they are closer to the real distribution.

The actual values for the combined and separated training site classifications were observed and it was noted that keeping the training sites for a category separated until after a classification improved overall accuracy for the Dismal Swamp scene by seven percentage points. For resource managers; however, a classification accuracy of 50% is still too low to be used as a landcover map.

Post-processing the thresholded classifications using a logical filler did not improve the overall accuracy or category accuracies of a scene over the standard maximum likelihood classification by any practical amount. It did; however, decrease the accuracy for both training methods when the percent of the scene needing to be filled was above approximately 50%. Basically, the spatial classifier did as good a job as the spectral classifer for the Dismal Swamp scene if the percent of the scene needing to be spatially classified was above 50%.

The spatial distribution of the unclassified pixels formed increasingly larger clumps as threshold level was increased. The clumps were pixels that had low probabilities of belonging to a category based on the training site statistics. Inadequate classifier training caused these clumps of low probability pixels. This was because the spectral values of many areas in the scene did not fall near a training distribution since the training distribution was not representative of the category. Clumping occurred at a lower threshold level for the combined training site method because the training site distributions for its categories did not coincide well with the real distributions so many pixels were being classified at low training site probabilities and therefore were removed at low threshold levels. The difference in spectral values for a category from one area of the scene to another was best accounted for by keeping training site statistics for a category separated until after the classification.

It must be kept in mind when interpreting these results that they apply to a unique set of circumstances. These circumstances are as follows:

1. the Dismal Swamp Landsat scene was destriped and registered prior to classification at EROS Data Center on IDIMS using a histogram normalization and smoothing routine and a first order nearest neighbor geometric registration respectively,

2. a supervised classification scheme was employed,

3. GIPSY algorithms were used for classification which could differ from other image processing systems,

4. the resolution of the reference data was at Anderson level II,

5. the spectral data had a narrow range of values and low variability,

6. the study area was more spectrally and spatially diverse than other forested wetland areas of this size, and

7. the vegetative cover map used as reference data was assumed to be 100% correct. Errors in the reference data could affect classifier training as well as accuracy assessment. Since the vegetation map was found to be 94% accurate in a previous study, errors in this study caused by inaccurate reference data were considered to be minimal.

These circumstances limit the applicability of these results to other Landsat classifications; however, it is felt that these circumstances yielded some insight into the maximum likelihood classification process.

Extensions of this research could concentrate on using an unsupervised classification approach so that the problems with locating representative training sites would not occur. The statistics generated from the clustering of a section of the entire scene could be used to classify the entire scene, at various threshold levels, with a maximum likelihood classifier. Also, a digital image with greater spectral variabilty and more easily identifiable categories could be classified.

LITERATURE CITED

Aldrich, R.C. 1979.Remote sensing of wildland resources: a state of the art review.General Technical Report RM-71. Rocky Mountain Forest and Range Experimental Station, Fort Collins, Colorado. 56 pp.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. A land use and land cover system for use with remote sensor data: U.S.G.S. Professional Paper 964. 28 pp.

Cai, D.L. 1983. GIPSY documentation. Spatial Data Analysis Laboratory. Virginia Tech.

Campbell, J.B. 1981.Spatial correlation effects upon accuracy of supervised classification of land cover. Photogrammetric Engineering and Remote Sensing. Vol.47, No.3. pp. 355-363.

Campbell, J.B. 1983. Mapping the Land: Aerial Imagery for Land Use Information. Washington D.C.: Association of American Geographers. 96 pp.

Ernst-Dottavio, C.L., R.M. Hoffer, and R.D. Mrczynski. 1981. Spectral characteristics of wetland habitats. Photogrammetric Engineering and Remote Sensing. Vol.47, No.2. pp. 223-227.

Fleming, M.D., and R.M. Hoffer. 1979. Machine processing of Landsat MSS and DMA topographic data for forest cover type mapping.1979 Machine Processing of Remotely Sensed Data Symposium, Purdue University, Indiana. pp. 377-390.

Gammon, P.T., and V.P. Carter. 1976. Comparison of vegetation classes in the Great Dismal Swamp using two individual Landsat images and a temporal composite. Proceedings of Third Purdue/LARS Symposium on Machine Processing of Remotely Sensed Data. pp. 3D-1.

Gammon, P.T., and V.P. Carter. 1979.Vegetation mapping with seasonal color infrared photographs. Photogrammetric Engineering and Remote Sensing. Vol.45, No.1. pp.87-97.

Garrett, M.K., and V.P. Carter. 1977. Contribution of remote sensing to habitat evaluation and management in a highly altered ecosystem. Transactions of the 42nd North American Wildlife and Natural Resources Conference. pp. 56-65.

Hixson, M., D. Scholz, N. Fuhs, and T. Akiyama. 1980. Evaluation of several schemes for classification of remotely sensed data. Photogrammetric Engineering and Remote Sensing. Vol.46, No.12. pp. 1547-1553.

Klemas, V., D. Bartlett, and R. Rogers. 1975. Coastal zone classification from satellite imagery. Photogrammetric Engineering and Remote Sensing. Vol.41, No.4. pp. 499-513.

Kovalick, W.M. 1983. The effect of selected preprocessing procedures upon the accuracy of a Landsat derived classification of a forested wetland. M.S. Thesis, Dept. of Forestry, Virginia Tech, Blacksburg, VA. 109 pp.

Lillesand, T.M., and R.W. Kiefer. 1979. Remote Sensing and Image Interpretation. John Wiley and Sons, New York. 612 pp.

Mead, R.A., and M.P. Meyer. 1977. Landsat digital data application to forest vegetation and land use classification in Minnesota. Machine Processing of Remotely Sensed Data Symposium, Purdue University. pp. 270-280.

Oaks, R.O. Jr. and N.K. Coch. 1973. Postmiocene Stratigraphy and Morphology, Southeastern Virginia. Virginia Div. Min. Resour. Bull. 82. 135 pp.

Prisley, S.P. 1982. Use of ancillary data in a Landsat classification of a forested wetland. M.S. Thesis, Dept. of Forestry, Virginia Tech, Blacksburg, VA. 119 pp.

Ranson, K.J. 1975. Computer assisted classification of mixtures with simulated spectral signatures. M.S. Thesis, Colorado State University, Fort Collins, Colorado. 117 pp.

Sharp, J.M. 1979. Demonstrating the value of Landsat data a case for lowered expectations. Photogrammetric Engineering and Remote Sensing. Vol.45, No.11. pp. 1487-1493.

Stanton, F.L. 1975. Remote sensing-a valuable tool in the Forest Service decision making process. Proceedings of NASA Earth Resources Survey Symposium, June 1975, Houston, Texas. pp. 1197-1219.

Story, M.H., J.B. Campbell, and G. Best. 1984. An evaluation of the accuracies of five algorithms for machine classification of remotely sensed data. The Ninth William T. Pecora Memorial Remote Sensing Symposium, Sioux Falls, South Dakota. pp. 399-405.

Story, M.H., and R.G. Congalton. 1986. Accuracy assessment: a user's perspective. Photogrammetric Engineering and Remote Sensing. Vol.52, No.3. pp. 397-399.

Swain, P.H., and S.M. Davis. 1978. Remote Sensing: The Quantitative Approach. Mcgraw-Hill, New York. 396 pp.

Townshend, J.R., and C.O. Justice. 1980. Unsupervised classification of MSS Landsat data for mapping spatially complex vegetation. International Journal of Remote Sensing. Vol.1, No.2. pp. 105-120.

Whitehead, D.R. 1972. Development and environmental history of the Dismal Swamp. Ecological Monographs. Vol.42, No.3. pp. 301-315.

APPENDIX A. ERROR MATRICES-COMBINED TRAINING CLASSIFICATIONS AND FILLS

0% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
	1	1358.	2178.	821.	303.	485.	2.
	2	163.	674.	237.	39.	87.	0.
Reference	3	347.	1269.	1578.	527.	407.	0.
Data	4	303.	1788.	3054.	5584.	1152.	26.
	5	1.	21.	0.	15.	435.	0.
	6	2.	5.	10.	10.	42.	447.

Sum of diagonal elements = 100760
% of scene classified = 233681/233681 = 100
Overall % correct = 43.12

1% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
	1	772.	2064.	803.	149.	0.	0.
	2	104.	629.	234.	22.	0.	0.
Reference	3	243.	1204.	1561.	314.	0.	0.
Data	4	204.	1639.	2991.	4152.	0.	16.
	5	0.	17.	0.	8.	0.	0.
	6	1.	4.	10.	7.	0.	361.

Sum of diagonal elements = 74750
% of scene classified = 175081/233681 = 74.92
Overall % correct = 42.69

1% Threshold - Combined Training Sites
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
	1	1256.	2656.	932.	301.	0.	2.
	2	151.	759.	251.	39.	0.	0.
Reference	3	318.	1502.	1716.	591.	0.	0.
Data	4	269.	2164.	3306.	6131.	0.	36.
	5	26.	237.	10.	199.	0.	0.
	6	2.	10.	16.	17.	0.	471.

Overall % correct = 44.22

2% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
Reference Data	1	637.	1986.	776.	109.	0.	0.
	2	91.	607.	229.	19.	0.	0.
	3	212.	1135.	1519.	238.	0.	0.
	4	166.	1271.	2906.	3213.	0.	15.
	5	0.	17.	0.	3.	0.	0.
	6	1.	2.	10.	5.	0.	346.

Sum of diagonal elements = 63220
% of scene classified = 15511/233681 =
Overall % correct = 40.76

2% Threshold - Combined Training Sites
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
Reference Data	1	1175.	2718.	980.	271.	0.	2.
	2	135.	774.	255.	36.	0.	0.
	3	301.	1572.	1766.	489.	0.	0.
	4	231.	2333.	3491.	5816.	0.	36.
	5	5.	299.	34.	134.	0.	0.
	6	2.	8.	20.	15.	0.	472.

Overall % correct = 42.81

3% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
Reference Data	1	617.	1936.	759.	81.	0.	0.
	2	88.	585.	226.	14.	0.	0.
	3	197.	1109.	1498.	191.	0.	0.
	4	159.	1244.	2856.	2236.	0.	15.
	5	0.	16.	0.	2.	0.	0.
	6	1.	2.	9.	3.	0.	346.

Sum of diagonal elements = 52820
% of scene classified = 141871/233681 = 60.71
Overall % correct = 37.23

3% Threshold - Combined Training Sites
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
Reference Data	1	1180.	2733.	997.	234.	0.	2.
	2	135.	775.	260.	30.	0.	0.
	3	295.	1602.	1791.	440.	0.	0.
	4	231.	2527.	3745.	5367.	0.	36.
	5	8.	314.	20.	131.	0.	0.
	6	2.	9.	21.	11.	0.	474.

Overall % correct = 41.03

4% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
Reference Data	1	72.	1807.	759.	78.	0.	0.
	2	21.	499.	226.	13.	0.	0.
	3	88.	957.	1498.	188.	0.	0.
	4	53.	812.	2856.	2065.	0.	13.
	5	0.	12.	0.	1.	0.	0.
	6	0.	1.	9.	3.	0.	322.

Sum of diagonal elements = 44560
% of scene classified = 123531/233681 = 52.86
Overall % correct = 36.07

4% Threshold - Combined Training Sites
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
Reference Data	1	200.	3277.	1336.	332.	0.	2.
	2	35.	802.	325.	38.	0.	0.
	3	138.	1447.	2042.	501.	0.	0.
	4	75.	1604.	4504.	5689.	0.	35.
	5	0.	275.	45.	152.	0.	0.
	6	0.	5.	26.	12.	0.	472.

Overall % correct = 39.39

5% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

Classification Results

Reference Data	1	2	3	4	5	6
1	72.	1807.	738.	44.	0.	0.
2	21.	499.	224.	8.	0.	0.
3	88.	957.	1464.	133.	0.	0.
4	53.	812.	2790.	1507.	0.	13.
5	0.	12.	0.	0.	0.	0.
6	0.	1.	8.	2.	0.	322.

Sum of diagonal elements = 38640
% of scene classified = 115751/233681 = 49.53
Overall % correct = 33.38

5% Threshold - Combined Training Sites
****

Error Matrix After Fill

Classification Results

Reference Data	1	2	3	4	5	6
1	206.	3343.	1399.	196.	0.	2.
2	35.	809.	336.	20.	0.	0.
3	141.	1511.	2134.	342.	0.	0.
4	75.	1795.	5469.	4530.	0.	37.
5	0.	325.	98.	49.	0.	0.
6	0.	6.	28.	8.	0.	474.

Overall % correct = 34.89

6% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
Reference Data	1	72.	1807.	738.	37.	0.	0.
	2	21.	499.	224.	8.	0.	0.
	3	88.	957.	1464.	119.	0.	0.
	4	53.	812.	2790.	1154.	0.	13.
	5	0.	12.	0.	0.	0.	0.
	6	0.	1.	8.	2.	0.	322.

Sum of diagonal elements = 35110
% of scene classified = 111971/233681 = 47.92
Overall % correct = 31.36

6% Threshold - Combined Training Sites
****

Error Matrix After Fill

		Classification Results					
		1	2	3	4	5	6
Reference Data	1	206.	3355.	1418.	164.	0.	2.
	2	35.	809.	336.	20.	0.	0.
	3	141.	1523.	2158.	307.	0.	0.
	4	76.	1869.	5719.	4205.	0.	37.
	5	0.	327.	99.	47.	0.	0.
	6	0.	6.	29.	5.	0.	475.

Overall % correct = 33.61

7% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
	1	0.	1588.	738.	37.	0.	0.
	2	0.	441.	224.	8.	0.	0.
Reference	3	0.	776.	1464.	119.	0.	0.
Data	4	0.	679.	2790.	1154.	0.	13.
	5	0.	12.	0.	0.	0.	0.
	6	0.	1.	8.	2.	0.	322.

Sum of diagonal elements = 33810
% of scene classified = 103731/233681 = 44.39
Overall % correct = 32.59

7% THRESHOLD - COMBINED TRAINING SITES
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
	1	0.	3260.	1704.	181.	0.	2.
	2	0.	791.	385.	24.	0.	0.
Reference	3	0.	1457.	2353.	318.	0.	0.
Data	4	0.	1801.	5855.	4213.	0.	37.
	5	0.	320.	101.	50.	0.	0.
	6	0.	5.	30.	5.	0.	476.

Overall % correct = 33.52

8% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	878.	738.	19.	0.	0.
	2	0.	326.	224.	4.	0.	0.
Reference	3	0.	691.	1464.	46.	0.	0.
Data	4	0.	605.	2790.	777.	0.	13.
	5	0.	12.	0.	0.	0.	0.
	6	0.	1.	8.	1.	0.	322.

Sum of diagonal elements = 28890
% of scene classified = 89201/233681 = 38.39
Overall % correct = 32.39

8% THRESHOLD - COMBINED TRAINING SITES
****

Error Matrix After Fill

		Classification Results					
		1	2	3	4	5	6
	1	0.	3100.	1914.	130.	0.	2.
	2	0.	782.	402.	17.	0.	0.
Reference	3	0.	1424.	2540.	165.	0.	0.
Data	4	0.	1808.	6460.	3601.	0.	37.
	5	0.	327.	110.	35.	0.	0.
	6	0.	5.	31.	4.	0.	476.

Overall % correct = 31.66

10% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	564.	623.	0.	0.	0.
	2	0.	219.	207.	0.	0.	0.
Reference	3	0.	523.	1176.	0.	0.	0.
Data	4	0.	492.	1772.	0.	0.	11.
	5	0.	0.	0.	0.	0.	0.
	6	0.	1.	6.	0.	0.	304.

Sum of diagonal elements = 16990
% of scene classified = 58961/233681 = 25.23
Overall % correct = 28.82

13% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

Reference Data	Classification Results 1	2	3	4	5	6
1	0.	488.	475.	0.	0.	0.
2	0.	170.	170.	0.	0.	0.
3	0.	430.	814.	0.	0.	0.
4	0.	363.	1045.	0.	0.	11.
5	0.	0.	0.	0.	0.	0.
6	0.	1.	5.	0.	0.	304.

Sum of diagonal elements = 12880
% of scene classified = 42721/233681 = 18.28
Overall % correct = 30.15

13% THRESHOLD - COMBINED TRAINING SITES
****

Error Matrix After Fill

Reference Data	Classification Results 1	2	3	4	5	6
1	0.	3092.	2052.	0.	0.	3.
2	0.	738.	462.	0.	0.	0.
3	0.	1410.	2718.	0.	0.	0.
4	0.	2683.	9050.	0.	0.	173.
5	0.	332.	140.	0.	0.	0.
6	0.	5.	21.	0.	0.	490.

Overall % correct = 16.89

15% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	236.	475.	0.	0.	0.
	2	0.	85.	170.	0.	0.	0.
Reference	3	0.	259.	814.	0.	0.	0.
Data	4	0.	227.	1045.	0.	0.	11.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	5.	0.	0.	304.

Sum of diagonal elements = 12030
% of scene classified = 36311/233681 = 15.54
Overall % correct = 33.13

17% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
Reference Data	1	0.	236.	318.	0.	0.	0.
	2	0.	85.	122.	0.	0.	0.
	3	0.	259.	546.	0.	0.	0.
	4	0.	227.	806.	0.	0.	11.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	4.	0.	0.	304.

Sum of diagonal elements = 9350
% of scene classified = 29181/233681 = 12.49
Overall % correct = 32.04

20% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	236.	318.	0.	0.	0.
	2	0.	85.	122.	0.	0.	0.
Reference	3	0.	259.	546.	0.	0.	0.
Data	4	0.	227.	806.	0.	0.	11.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	4.	0.	0.	304.

Sum of diagonal elements = 9350
% of scene classified = 29181/233681 = 12.49
Overall % correct = 32.04

20% THRESHOLD - COMBINED TRAINING SITES
****

Error Matrix After Fill

		Classification Results					
		1	2	3	4	5	6
	1	0.	2774.	2367.	0.	0.	5.
	2	0.	709.	492.	0.	0.	0.
Reference	3	0.	1377.	2751.	0.	0.	0.
Data	4	0.	2586.	9144.	0.	0.	176.
	5	0.	228.	244.	0.	0.	0.
	6	0.	4.	21.	0.	0.	491.

Overall % correct = 16.91

23% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	0.	318.	0.	0.	0.
	2	0.	0.	122.	0.	0.	0.
Reference	3	0.	0.	546.	0.	0.	0.
Data	4	0.	0.	806.	0.	0.	10.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	4.	0.	0.	302.

Sum of diagonal elements = 8480
% of scene classified = 21081/233681 = 9.02
Overall % correct = 40.23

25% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	0.	0.	0.	0.	0.
	2	0.	0.	0.	0.	0.	0.
Reference	3	0.	0.	0.	0.	0.	0.
Data	4	0.	0.	0.	0.	0.	10.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	0.	0.	0.	302.

Sum of diagonal elements = 3020
% of scene classified = 3121/233681 = 1.34
Overall % correct = 96.76

30% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

	Classification Results 1	2	3	4	5	6
1	0.	0.	0.	0.	0.	0.
2	0.	0.	0.	0.	0.	0.
3	0.	0.	0.	0.	0.	0.
4	0.	0.	0.	0.	0.	10.
5	0.	0.	0.	0.	0.	0.
6	0.	0.	0.	0.	0.	302.

Sum of diagonal elements = 3020
% of scene classified = 3121/233681 = 1.34
Overall % correct = 96.76

40% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	0.	0.	0.	0.	0.
	2	0.	0.	0.	0.	0.	0.
Reference	3	0.	0.	0.	0.	0.	0.
Data	4	0.	0.	0.	0.	0.	10.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	0.	0.	0.	302.

Sum of diagonal elements = 3020
% of scene classified = 3121/233681 = 1.34
Overall % correct = 96.76

80% Threshold - Combined Training Sites
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	0.	0.	0.	0.	0.
	2	0.	0.	0.	0.	0.	0.
Reference	3	0.	0.	0.	0.	0.	0.
Data	4	0.	0.	0.	0.	0.	4.
	5	0.	0.	0.	0.	0.	0.
	6	0.	0.	0.	0.	0.	128.

Sum of diagonal elements = 1280
% of scene classified = 1331/233681 = .57
Overall % correct = 96.17

APPENDIX B. ERROR MATRICES-SEPARATED TRAINING CLASSIFICATIONS AND FILLS

0% THRESHOLD - SEPARATED TRAINING SITES

****

Error Matrix After Thresholding

Classification Results

		1	2	3	4	5	6
	1	3179.	543.	981.	268.	174.	2.
	2	601.	262.	273.	41.	24.	0.
Reference	3	1354.	458.	1684.	514.	118.	0.
Data	4	1458.	955.	3264.	5831.	371.	28.
	5	95.	8.	24.	40.	305.	0.
	6	10.	3.	10.	15.	20.	458.

Sum of diagonal elements = 117190
% of scene classified = 233681/233681 = 100
Overall % correct = 50.15

5% THRESHOLD - SEPARATED TRAINING SITES

****

Error Matrix After Thresholding

Reference Data	Classification Results 1	2	3	4	5	6
1	2446.	500.	834.	176.	0.	0.
2	497.	234.	243.	28.	0.	0.
3	1093.	409.	1483.	352.	0.	0.
4	1115.	666.	2823.	4698.	0.	13.
5	51.	5.	0.	12.	0.	0.
6	4.	1.	8.	7.	0.	322.

Sum of diagonal elements = 91830
% of scene classified = 180181/233681 = 77.11
Overall % correct = 50.97

5% THRESHOLD - SEPARATED TRAINING SITES

****

Error Matrix After Fill

Reference Data	Classification Results 1	2	3	4	5	6
1	3148.	658.	1030.	309.	0.	2.
2	601.	293.	266.	40.	0.	0.
3	1341.	546.	1660.	580.	0.	0.
4	1565.	992.	3133.	6184.	0.	32.
5	370.	20.	2.	80.	0.	0.
6	12.	4.	14.	20.	0.	467.

Overall % correct = 50.29

10% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
	1	2273.	403.	803.	145.	0.	0.
	2	476.	170.	236.	22.	0.	0.
Reference	3	1020.	328.	1458.	307.	0.	0.
Data	4	978.	571.	2803.	4018.	0.	11.
	5	12.	1.	0.	9.	0.	0.
	6	3.	1.	8.	6.	0.	304.

Sum of diagonal elements = 82230
% of scene classified = 163661/233681 = 69.91
Overall % correct = 50.24

10% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
	1	3102.	576.	1122.	344.	0.	2.
	2	652.	228.	275.	45.	0.	0.
Reference	3	1270.	490.	1734.	634.	0.	0.
Data	4	1376.	944.	3270.	6258.	0.	32.
	5	211.	23.	10.	227.	0.	0.
	6	11.	3.	16.	20.	0.	467.

Overall % correct = 50.56

15% THRESHOLD - SEPARATED TRAINING SITES

****

Error Matrix After Thresholding

Classification Results

		1	2	3	4	5	6
Reference Data	1	1842.	353.	614.	113.	0.	0.
	2	370.	127.	181.	16.	0.	0.
	3	812.	243.	1243.	247.	0.	0.
	4	793.	212.	2643.	3280.	0.	11.
	5	0.	0.	0.	5.	0.	0.
	6	3.	0.	8.	5.	0.	304.

Sum of diagonal elements = 67960
% of scene classified = 134251/233681 = 57.45
Overall % correct = 50.62

15% THRESHOLD - SEPARATED TRAINING SITES

****

Error Matrix After Fill

Classification Results

		1	2	3	4	5	6
Reference Data	1	3104.	659.	1023.	358.	0.	2.
	2	691.	187.	275.	48.	0.	0.
	3	1332.	462.	1717.	617.	0.	0.
	4	1508.	369.	3885.	6111.	0.	33.
	5	122.	9.	42.	299.	0.	0.
	6	11.	1.	20.	16.	0.	468.

Overall % correct = 49.58

30% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
Reference Data	1	1395.	200.	614.	96.	0.	0.
	2	289.	88.	181.	14.	0.	0.
	3	603.	146.	1243.	216.	0.	0.
	4	541.	159.	2643.	2872.	0.	10.
	5	0.	0.	0.	4.	0.	0.
	6	1.	0.	8.	4.	0.	302.

Sum of diagonal elements = 59000
% of scene classified = 116291/233681 = 49.76
Overall % correct = 50.73

30% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
Reference Data	1	2581.	497.	1550.	516.	0.	2.
	2	663.	159.	325.	53.	0.	0.
	3	1219.	360.	1923.	626.	0.	0.
	4	1240.	336.	4167.	6131.	0.	33.
	5	112.	9.	44.	306.	0.	0.
	6	4.	1.	27.	14.	0.	470.

Overall % correct = 48.20

60% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
	1	677.	76.	581.	54.	0.	0.
	2	173.	49.	176.	9.	0.	0.
Reference	3	542.	93.	1196.	145.	0.	0.
Data	4	485.	130.	2551.	1874.	0.	8.
	5	0.	0.	0.	2.	0.	0.
	6	1.	0.	7.	2.	0.	271.

Sum of diagonal elements = 40670
% of scene classified = 91011/233681 = 38.95
Overall % correct = 44.69

60% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Fill

		Classification Results 1	2	3	4	5	6
	1	2826.	212.	1753.	355.	0.	0.
	2	746.	74.	334.	47.	0.	0.
Reference	3	1374.	205.	2037.	512.	0.	0.
Data	4	1382.	300.	4568.	5626.	0.	30.
	5	158.	11.	40.	263.	0.	0.
	6	6.	0.	30.	11.	0.	469.

Overall % correct = 47.21

80% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Thresholding

		Classification Results 1	2	3	4	5	6
	1	393.	76.	581.	37.	0.	0.
	2	133.	49.	176.	6.	0.	0.
Reference	3	527.	93.	1196.	72.	0.	0.
Data	4	466.	130.	2551.	1498.	0.	4.
	5	0.	0.	0.	1.	0.	0.
	6	1.	0.	7.	1.	0.	128.

Sum of diagonal elements = 32640
% of scene classified = 81271/233681 = 34.79
Overall % correct = 40.16

90% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
Reference Data	1	0.	0.	581.	30.	0.	0.
	2	0.	0.	176.	5.	0.	0.
	3	0.	0.	1196.	62.	0.	0.
	4	0.	0.	2551.	1302.	0.	4.
	5	0.	0.	0.	1.	0.	0.
	6	0.	0.	7.	1.	0.	128.

Sum of diagonal elements = 26260
% of scene classified = 60451/233681 = 25.87
Overall % correct = 43.44

95% THRESHOLD - SEPARATED TRAINING SITES
****

Error Matrix After Thresholding

		Classification Results					
		1	2	3	4	5	6
	1	0.	0.	581.	23.	0.	0.
	2	0.	0.	176.	5.	0.	0.
Reference	3	0.	0.	1196.	49.	0.	0.
Data	4	0.	0.	2551.	949.	0.	4.
	5	0.	0.	0.	1.	0.	0.
	6	0.	0.	7.	1.	0.	128.

Sum of diagonal elements = 22730
% of scene classified = 56691/233681 = 24.26
Overall % correct = 40.09

The vita has been removed from the scanned document