USE OF ANCILLARY DATA IN A LANDSAT CLASSIFICATION OF A FORESTED WETLAND

by

Stephen Post Prisley

Thesis submitted to the Graduate 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:

Dr. Roy A. Mead, Chairman

Dr. James A. Burger

Dr. Rickard H. Hokans

June, 1982
Blacksburg, Virginia

ACKNOWLEDGEMENTS

This thesis is the result of a great expenditure of time and effort on the part of many people. I greatly appreciate the many who have provided financial, technical, administrative, as well as moral support and encouragement. Primarily, I am indebted to my major professor, Dr. Roy Mead, and my supervisors in the USGS Water Resources Division, Virginia Carter and Pat Gammon. The guidance and financial support I received from them made this work possible. I also owe a great debt of thanks to Wayne Miller of the EROS Data Center Applications Branch, who was an indispensible cooperator, a fine tutor, and a great friend. I sincerely appreciate the support of my graduate committee, Dr. Rickard Hokans and Dr. James Burger, and of Mary Kieth Garrett of the Great Dismal Swamp National Wildlife Refuge.

I owe an immeasurable debt to many excellent friends, roommates, fellow graduate students, backpacking buddies, and raquetball partners for their comraderie, understanding, and support. I wouldn't have lasted here without it. I especially appreciate the many hours of statistical consul-

tation, the endurance of my complaints, the numerous editorial suggestions, and mostly the close friendship of Dr. James L. Smith.

Finally, I owe what I am to my family. Their contribution has lasted for the past 24 years, and cannot be easily forgotten. I believe I am also endebted to a man who has had a great influence on my life, in ways he never knew while he lived. Therefore, it is to the memory of my grandfather, Charles Arnett Taylor, that I dedicate this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Land cover classification and mapping has long been a concern of those involved in natural resources management. Cover-type maps are an invaluable source of information on the distribution of natural resources on the earth's surface, and are used in all phases of resource planning and management. Modern remote sensing techniques have provided a means of classifying and mapping large areas much faster and at less expense than older ground-based methods (Stanton, 1975).

Use of satellite sensors (such as the Landsat multispectral scanner) may be more efficient than use of aerial photography when the area to be mapped is very large, a high level of detail is not needed, or repetitive coverage is desired. Cover-type classifications of Landsat data have been performed by manual interpretation of visual products, or more commonly, computer-aided analysis of digital data. A variety of techniques have been used for computer classification of Landsat data, and results have been variable. Often, however, cover-type classifications have been less accurate than hoped for, or to a lower level of detail than needed. The 85% threshold for acceptable accuracy suggested by Anderson, et. al. (1976) is often not reached except in general level II classifications.

Recent research efforts have used additional data to supplement the four bands of Landsat MSS data in an attempt to improve the accuracies of computer classifications. Most of these studies have been performed in areas in which there is an obvious relationship between the cover types of interest and some ancillary variable(s). A typical example is the distribution of forest cover types along gradients of slope, aspect, and elevation in the Rocky mountains (Hoffer, et. al., 1975). Little work has been done in areas where no such clear trends exist; consequently, the full spectrum of these procedures has not been examined.

To further evaluate the utility of ancillary data in Landsat classifications, a project was initiated to classify the vegetation of the Great Dismal Swamp using Landsat and supplemental data. The Great Dismal Swamp National Wildlife Refuge was chosen as a study site for several reasons. First, recent studies have indicated that mapping of vegetation communities in the swamp using multitemporal Landsat

MSS data alone could not be done at the desired level of detail to an acceptable degree of accuracy (Gammon, et. al., 1979). A detailed map of photo-interpreted vegetation types has been produced for the Dismal Swamp. This map can serve as a "ground truth" or reference data source for the swamp, and allows low-cost classifier training and accuracy assessment for the entire area. Several data sets are available for variables that may be related to vegetation distribution and are consequently suitable for incorporation as ancillary data in a classification process. Finally, the area is ecologically complex, and is unlike any of the study sites used in similar previous research. Therefore, the Great Dismal Swamp offers a stringent test of the contribution of ancillary data to a Landsat classification.

It is hypothesized that the use of vegetation-related ancillary variables in a layered classification of Landsat data for the Dismal Swamp will result in accuracies higher than those obtainable from classification of spectral data alone.

The goal of this research project is to evaluate the use of vegetation-related ancillary variables for improving the performance of Landsat classification of the vegetation of the Great Dismal Swamp. Specific objectives are:

1) To use ancillary variables in a layered procedure to divide individual Landsat spectral classes into more accurate vegetation classes.

2) To test for improvement in classification accuracy of the layered procedure over single-stage unsupervised and spectral stratification classifications.

LITERATURE REVIEW

The practice of land cover mapping has progressed a great deal in the past several decades. Early mapping efforts involved extensive field work and required considerable investment of time and money. Since the 1940's, the use of panchromatic, medium-scale aerial photographs has been an accepted technique, and has recently been supplemented by the use of small-scale aerial photographs and satellite imagery (Lillesand and Kiefer, 1979). For mapping of extensive areas into broad land cover classes, digital processing of Landsat spectral data can be efficient and cost-effective. However, results have often been less accurate than hoped for (Sharp, 1979), or to a lower level of detail than needed (Fleming and Hoffer, 1979).

Recent attempts to improve the performance of digital Landsat classifications have utilized ancillary data sets to supplement the four channels of Landsat spectral data (Hutchinson, 1982). A variety of data types and classification algorithms have been used and most studies have produced promising results (Johnson and Rohde, 1980).

Ancillary Data Types

The first supplementary data sets used in Landsat cover-type classifications were composed of additional spectral data. Landsat data from several dates have been used in multitemporal classifications to discriminate objects whose spectral characteristics vary over time (Gammon, et al., 1979). Also, use of other remote sensors has provided coverage of additional spectral wavelength bands for use in classification algorithms. For example, thermal imagery (Price, 1981), and radar imagery (Daily, et al., 1979; Clark, 1981) have been combined with Landsat visible and infrared data to improve land cover classifications.

The next step taken to increase the information available to a computer classification algorithm was the consideration of the spatial context of the spectral data. This was done using classification algorithms such as those reported by Hornung and Smith (1974) and Kettig and Landgrebe (1976). One common technique was to use edge detection algorithms to identify spectrally homogeneous areas on the landscape which were then classified as units (Bryant, 1979). This eliminated the speckling effect resulting from classification of each pixel independently of its neighbors. Another method of incorporating information on the spatial arrangement of spectral data in machine classifications was the use of a texture channel. A texture channel, as described by Logan, et al. (1979), is composed of the standard deviation of the spectral values of the pixel's eight nearest neighbors. Line and sample (row and column) coordinates of pixels have also been used as ancillary channels to introduce spatial constraints in classification algorithms (Logan and Strahler, 1980). Such additional channels of data were then included in typical algorithms for classification of multispectral data.

Recently, non-spectral ancillary data layers have also been used in computer-assisted Landsat classifications. The logic of the classification process dictates that the ancillary variables to be used should be defined in a spatial context, be relatively independent of each other, and be strongly related to the distribution of the cover types of interest. Many such data are available. Examples include magnetic data in geologic mapping (Anuta, et al., 1976), and terrain data in vegetation mapping (Hoffer, et al., 1979). Even political or ownership boundaries may supply useful information that cannot be derived from spectral data alone (Davis and Friedman, 1979). In addition, transformations of primary ancillary variables may be used to create secondary variables more directly related to the cover types of interest. For example, spatial differentiation of elevation yields slope and aspect values, which may be combined to form solar radiation indices (Anuta, 1976).

In many cases, selection of the proper variables for use in a digital Landsat classification may be based on a previous knowledge of the effect of certain factors on the distribution of cover types. Otherwise, sample data from locations for which the cover type and ancillary variables are known may be used to formulate statistical models which describe the relationship between the ancillary variables and the cover types. For example, discriminant analysis (Tom, et al., 1978; Shasby, et al., 1980) and linear regression techniques (Hoffer, et al., 1979) were used with variable selection procedures to choose variables and develop such models. The appropriate models and ancillary variables may then be incorporated into cover-type classification algorithms.

Classification Techniques

Once the ancillary variables have been chosen, they must be put in computer-readable form and spatially registered to the Landsat data. Then, a variety of techniques are available for incorporation of this data in digital classifications of land use or land cover types (Johnson and Rohde, 1980). The techniques used to date may be categorized as radiance correction, single-stage, layered, or modified probabilities classification methods.

Radiance Correction

The first approach to be considered is the use of topographic data to remove variation in spectral response due to terrain. This technique involves mathematical adjustments to the raw data to simulate the radiance values that would have been recorded by the sensor if all pixels were on flat land. This is intended to remove spectral variation between similar cover types occurring on different topographic positions. According to Fleming and Hoffer (1979), the procedure succeeded in reducing spectral variation, but did not result in increased classification accuracy. Similarly, a terrain effect correction did not appear highly significant in improving the classification performed by Tom, et al. (1978). A variation of this procedure has been employed by Woodcock, et al. (1980). The technique was termed "differential illumination compensation" and involved the use of terrain data after spectral classification to divide each spectral class into shaded and unshaded portions. This doubled the number of classes and reduced spectral variability within each class. The classes were then edited and aggregated to form the vegetation types of interest. It is not known what effect this portion of the classification procedure had on the results, as it was one of several modifications of the control classification. These results have indicated that adjustment of radiance values may not be the best use of topographic data in a Landsat classification.

Single-stage Classification

Another method of incorporation of ancillary data in a digital classification scheme is a typical supervised or unsupervised procedure using additional channels of data. Such single-stage classifications with ancillary data are characterized by the consideration of spectral and ancillary data at the same step in the classification algorithm. Both supervised and unsupervised single-stage classifiers typically use a sample of the entire data set to "train" the computer to recognize cover type classes. Then, in a classification step, all other pixels are assigned to the class they most likely belong to. A simple mathematical treatment of these procedures has been presented by Swain and Davis (1978), and a general discussion may be found in Lillesand and Kiefer (1979).

During the training phase of unsupervised classifications, the computer identifies spectrally similar groups or clusters of pixels in the sample. Only after all clusters are defined are decisions made as to the cover type represented by the spectral clusters. Each cluster is defined by a mean vector and variance-covariance matrix. For example, using four bands of Landsat spectral data, the location of a cluster in the feature space is described by the four mean values for the data points assigned to it, one mean for each spectral band. The dispersion of the cluster is described by a four-by-four variance-covariance matrix. Clusters are split or joined in the clustering algorithm dependent on user-defined criteria such as minimum distances between cluster centers or maximum standard deviations within clusters. Varying these criteria will usually result in varying numbers of clusters defined by the computer. The procedure remains the same when ancillary data are added, except that the algorithm then deals with more than four data dimensions. For example, addition of three ancillary channels to four Landsat channels will result in clusters defined by a vector of seven means, and a seven-by-seven dispersion matrix.

After definition of each cluster in the training phase, the remaining pixels are labeled in the classification phase with the number of the cluster they most likely belong to. This labeling or classification is based on a variety of possible decision rules: minimum distance, maximum likelihood, parallelipiped, Bayesian, and others. Unsupervised single-stage classifications have been employed with ancillary and Landsat data sets by Anuta, et al. (1976), Hoffer, et al. (1979) and Logan and Strahler (1980).

Supervised classifications differ from unsupervised in that the user chooses the data points to be assigned to clusters, not the computer. This method employs a training set of pixels, for each of which is known the true cover type and the four Landsat radiance values. The computer then calculates the mean vector and variance-covariance matrix for all pixels assigned to each cover type. Labeling of pixels whose cover type is unknown proceeds as described earlier, using the statistics computed in the training step and an appropriate decision rule. Again, introduction of ancillary data merely increases the dimensionality of the data for the algorithm to work with. Applications of this procedure include work done by Miller, et al. (1978), Strahler, et al. (1978), and Fleming and Hoffer (1979).

Layered Classification

Another procedure for classification of Landsat and ancillary data is the layered approach. The logic of this technique has been described by Swain, et al. (1975) and Swain and Hauska (1977). Layered classifiers (also called decision-tree or multi-stage classifiers) involve multiple steps using a different data type and resulting in different numbers of classes at each step. In work done by Shasby, et al. (1980), the first classification step involved the identification of 48 spectral classes using a modified clustering algorithm and four-band Landsat data. Next, spectral classes which represented more than one cover type were divided on the basis of discriminant analysis of four terrain channels of data. In a wildland resource inventory in Arizona (Johnson and Rohde, 1980), elevation data were used in a second layer reclassification of vegetation types that were classified together in the first step. For example, creosote bush was found to occur on lower elevations than sagebrush, and therefore could be discriminated with elevation data, while the two were inseparable based on spectral data alone. Ernst and Hoffer (1980) used a layered procedure to improve the classification of wetland habitats. In their study, some of the spectral classes formed in the first step using Landsat data were split during the second step into upland and lowland classes using digitized soils units as an ancillary channel.

Both of the above examples used spectral data at the first step in the process. An alternative is to separate the data into classes based on ancillary data first, then perform typical spectral classifications on each class independently. This is essentially a pre- classification stratification, and may be used to reduce the variability of the spectral data to be classified (Hutchinson, 1981). Other classifications using the layered technique are discussed by Fleming and Hoffer (1979), Linden, et al. (1980), and Miller, et. al. (1980).

Modified Probabilities Classification

A combination of the single-stage and layered procedures has been developed which employs modified prior probabilities in the classification step. While Scholz and Weismiller (1980) called this a layered classification, the use of a similar technique by Strahler, et al. (1978) was termed a probabilistic approach, and was treated in a single-stage fashion. This procedure differs from both single-stage and layered techniques enough to be considered separately.

The first step in a modified probabilities classification is the determination of prior probabilities of class membership based on some type of ancillary data. For example, in the work done by Scholz and Weismiller (1980), a study area was divided into three physiographic classes using an ancillary data channel. A sampling of the area provided the proportions of each cover type within each physiographic class. These proportions were used as prior probabilities for the inclusion of a pixel into a given cover type, assuming its physiographic class was known. Next, the statistics for spectral classes are calculated using a typical supervised or unsupervised approach. Finally, the spectral statistics and prior probabilities are used in a single-step classification of the spectral and ancillary data. A pixel is assigned to the class for which it has the highest probability of membership, where this probability is a function of the mean vector and covariance matrix of the class, the prior probability vector for the class, and the spectral and ancillary values for the pixel.

This technique should not be considered a single-stage approach because the spectral and ancillary data layers are used in totally different manners: the former in a maximum-likelihood decision rule, the latter as an indicator of prior probabilities. However, neither is it a conventional layered approach since all data types are used in the same equation in the classification algorithm. Rather, it is a hybrid of the two procedures, and exhibits characteristics similar to each.

Accuracy Assessment

A necessary step in the classification process is assessment of the accuracy of the final product. This permits evaluation of the usefulness of the product, as well as comparison with classifications done using other data sources or techniques.

Classifier performance may be evaluated using either site-specific or non-site-specific techniques. Non-site-specific accuracy assessment involves only the comparison of total acreages of various cover types as classified by the computer with known acreages of cover types from a reference source (Meyer, et al., 1975). Site-specific accuracy assessment, however, provides information on locations and distributions of classification error among cover types, and is generally the preferred technique. This procedure typically involves comparison of a sample of classified pixels with aerial photographs or manual observations of the corresponding locations on the ground. The number of correctly classified pixels, as well as omission and commission errors, are then reported in a contingency table or error matrix (Lillesand and Kiefer, 1978). Rows of these matrices generally represent the cover types of the pixels as assigned during the Landsat classification, and columns represent the cover types of the pixels as recorded in the "ground truth" or reference data.

Several techniques exist for evaluating site-specific or non-site-specific classification accuracy. The simplest procedure involves merely computing the overall percentage of correctly classified pixels. This is done by summing the diagonal elements of an error matrix and dividing by the total number of pixels classified.

In recent years, statistical analysis techniques have been applied to classification accuracy assessment. For example, a simple linear regression of known cover types on classified cover types may be used to indicate the degree of agreement between the ground truth and the computer classification. In this method, high correlation coefficients indicate higher classification accuracies. Analysis of variance (Rosenfield, 1981) has also been proposed as a means of statistically evaluating classifier performance. A technique employing discrete multivariate analysis has been suggested by Congalton (1981). This procedure indicates whether a classification has produced significantly better results than a "random" classification of the same data to the same groups. In a random classification, each cell entry is equal to the product of the marginal probabilities for that row and column. The kappa statistic uses the difference between diagonal entries from the observed error matrix and the corresponding diagonal entries from a randomly classified error matrix. Kappa may vary from -1 to +1, although only the worst classifications will result in a kappa value less than zero. High classification accuracies correspond to kappa values near one. Kappa values for different classification procedures may be compared using a t-test (Congalton and Mead, 1981) to check for similar degrees of accuracy. This is a flexible and powerful technique, and eliminates the effect of different marginal frequencies of cover types on the expression of classification accuracy.

Comparison of Classification Techniques

Hoffer (1975) reported that consistent improvements in accuracy were not observed in a study using unsupervised single-stage classification of Landsat and ancillary data. In all other studies, improvements in overall classification accuracy were noted, ranging from 7% (Scholz and Weismiller, 1980) to 27% (Strahler, et al., 1978). No one technique reviewed proved consistently better or worse than others, and all showed some limitations. However, most authors concluded that use of ancillary data in Landsat classifications showed great promise.

Several difficulties were encountered in a number of these studies. Most problems involved violations of the assumptions inherent in the classification algorithms. The first problem encountered was in the sampling of the study area data for unsupervised clustering or supervised training procedures. Proper classification relies on the assumption that the sample used for training or clustering encompasses the total variability in each channel of the data for each class. Hoffer (1975) reported that difficulties arose when the sample was representative of the variation in the spectral data, but not the variation in the topographic data. This problem was resolved in another study which employed a topographically-stratified random sample to better represent the topographic variation (Hoffer, et al., 1979).

Many of the algorithms in use for classification assume that the data used are parametric; typically, that they follow a multivariate normal distribution. This is generally an acceptable assumption for multispectral data on land use or land cover types that exhibit fairly consistent spectral characteristics. However, many of the ancillary data used show no such distribution. For example, water bodies may occur at almost any elevation and may not show a characteristic elevation distribution that can be adequately described by a mean and variance. In such cases, nonparametric techniques such as those described by Tom, et al. (1978) may be more appropriate.

An additional difficulty with data distributions is the spread of the data in the various channels. Even if data are normally distributed, problems may emerge when the data show a higher variance in one channel than in others. This causes many clustering algorithms to subdivide the data more finely in that dimension than in the rest. Hence, one data channel may influence the clustering procedure more than its relative importance warrants. Strahler (1981) recommends scaling data such as texture and terrain to alleviate this problem.

An important factor to be considered in comparison of techniques is the computer processing time required for the algorithm. In single-stage classifiers, incorporation of additional data channels could increase the number of calculations in a geometric progression. Tom, et al. (1978) discuss the effect of ancillary data channels on processing time and algorithm efficiency (defined as number of correctly classified pixels per CPU second expended). Their research indicates that a point is reached at which continued addition of ancillary channels does not increase accuracy enough to warrant the cost. The layered classifier uses different data channels at different steps, which eliminates the manipulation of unnecessary data at each step. This results in considerably faster processing with no significant difference in accuracy when compared to single-stage classifiers.

In many cases, land-cover classification using Landsat spectral data has not yielded sufficiently accurate or detailed results. Consequently, ancillary spectral and non-spectral data have been incorporated in the classification process. Several algorithms for dealing with additional channels of data have been developed, and results have been encouraging. At present, the computational efficiency, the logical decision approach, and the fewer violated assumptions involved in the layered classifiers lend them an advantage over single-stage techniques.

DATA AND PROCEDURES

Description of the Study Area

The Great Dismal Swamp is a 50,590 ha. (125,000 ac.) forested wetland situated on the Virginia-North Carolina border in the mid-Atlantic coastal plain. The swamp is presently managed as a National Wildlife Refuge by the U.S. Department of Interior Fish and Wildlife Service. It is the site of many biological, geological, and hydrological studies, consequently, many types of data are available from several sources for ongoing research.

The Great Dismal Swamp is bounded on the west by the Suffolk Scarp, a Pleistocene beach ridge roughly 15 feet higher in elevation than the major portion of the swamp. From the scarp, the land surface of the Dismal Swamp slopes downward to the east at a gradient of about 0.2m/km (1 foot/mile). Two drainages cutting through the Suffolk Scarp on the west provide an inflow of surface water to the swamp, while outflow occurs by means of the Dismal Swamp Canal and the Northwest and Pasquotank rivers. A network of over 40 canals and ditches throughout the swamp creates a complex pattern of drainages and barriers to water flow.

The majority of the swamp resides on organic soils ranging in depth from a few centimeters to 3.7m (12 feet). Part of the variation in peat depth is due to the repeated occurence of fires in the Dismal Swamp. Extensive fires have been documented since the early 1800's and have been responsible for the loss of up to 1.5 meters or more of peat soil in certain areas.

The Great Dismal Swamp lies near the northern or southern extremes of the range of several plant species and consequently exhibits a great diversity of vegetation types (Garrett and Carter, 1977). The major vegetation communities in the Great Dismal Swamp include cypress-gum swamp types, maple-dominated mixed hardwood stands, inkberry and bayberry shrub communities, and pure stands of Atlantic white-cedar, loblolly, and pond pines. Gradations and mixtures of the above types occur in many combinations and locations throughout the swamp. In addition, disturbances to the swamp by ditching, logging, and fires are believed to have significantly altered the previous distribution of these vegetation types. The result is a highly intermixed assortment of communities and ecotones which makes mapping and subsequent management of vegetation and wildlife habitat a complicated process.

Preparation of Data

The creation of a spatial data base containing Landsat and ancillary data types was necessary prior to any cover-type classification. The spatial data base structure allows easy access to any data type for any given location on the ground, and a site-specific overlay of any combination of data types. The first step in the construction of this data base, therefore, was the registration of all data to a common coordinate system. A grid format based on the Universal Transverse Mercator (UTM) projection was chosen for its simplicity and adaptability. All data sets were registered to this common base using first order coordinate transformation equations. To simplify data processing and lessen the computer storage requirements, all data were scaled to integer values between 0 and 255, inclusive. This is because computer representation of values in that range is based on an eight-bit byte, allowing 2^8 or 256 quantized levels.

The data base consists of multiple layers of data, each in image (matrix) format. Each pixel (cell) in each image contains the value of some variable for a unique 50-meter square on the ground. The entire study area is covered by a matrix of 810 rows and 420 columns, with rows oriented east/west and columns oriented north/south.

The first data set to be entered into the spatial data base was a vegetation type map of the Dismal Swamp (Figure 1). This map had been produced from photo-interpretation of color-infrared aerial photographs and contained over 70 detailed classes of overstory and understory vegetation (Carter and Gammon, 1976). The map was digitized at the EROS Data Center and registered to the 50-meter UTM grid of the data base. Each unique vegetation class was assigned a code number, with areas outside the swamp boundary assigned a zero. It was decided that rather detailed classes would be sought during the classification, since broad level I classes (Anderson, et. al., 1976) would not be as meaningful to resource managers. Therefore, the 76 vegetation classes were collapsed to form a 15-class level III and a 6-class level II classification structure (Figure 2). These vegetation images were subsequently used in classifier training, cluster labeling, and accuracy assessment.

The spectral data chosen for this study came from the January 27, 1978 overpass of the Landsat C satellite. All preprocessing of Landsat data was performed on the Interactive Digital Image Manipulation System (IDIMS) at the EROS Data Center in Sioux Falls, South Dakota. Sensor miscalibration caused a striping in the image that was corrected

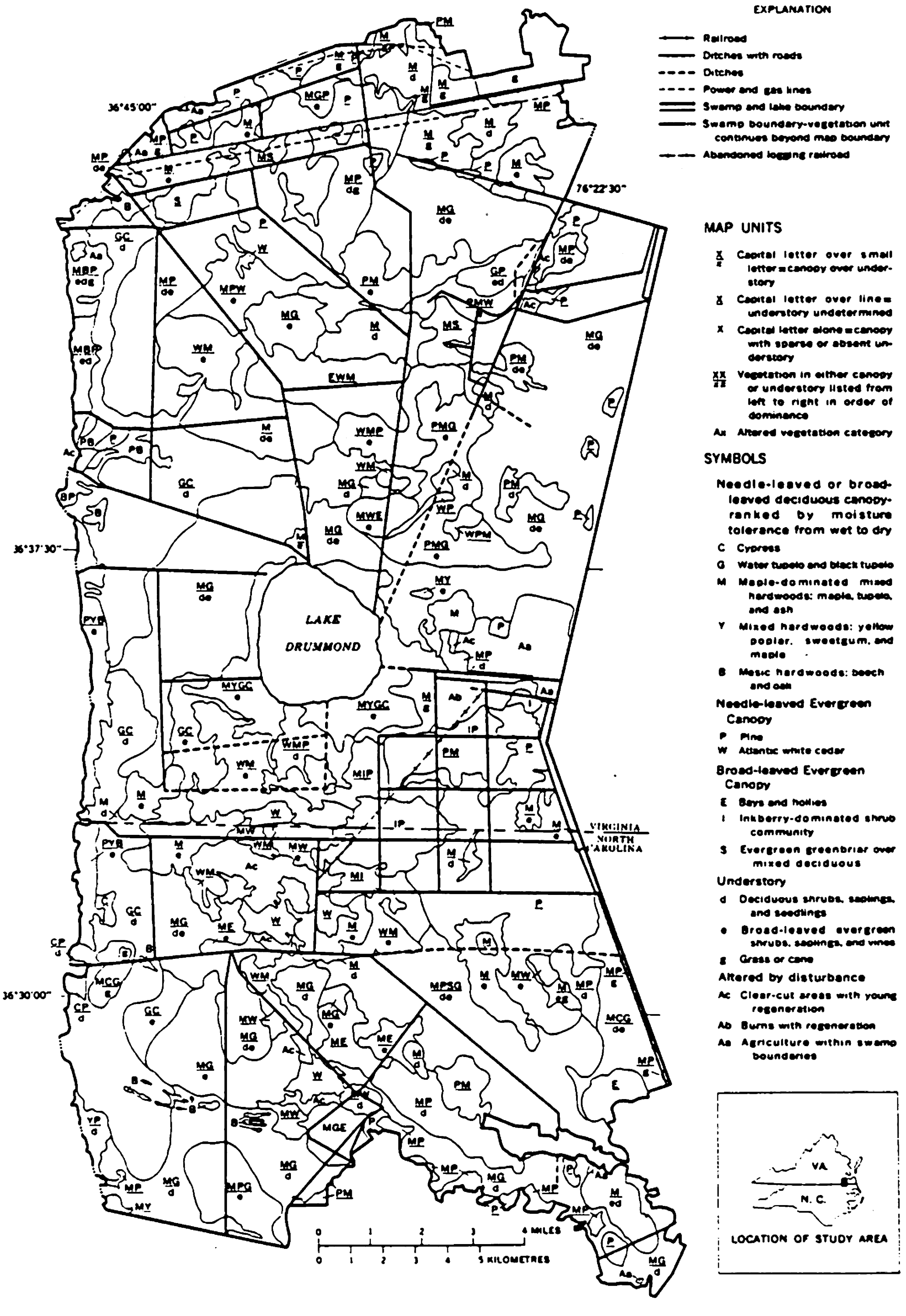


Figure 1. Great Dismal Swamp Vegetative Cover Map

Level III	Level II
1. Pine	1. Coniferous evergreen
2. Pine-deciduous mix	
3. Atlantic whitecedar	
4. Whitecedar-deciduous mix	
5. Inkberry shrub	2. Broadleaved evergreen
6. Evergreen vines	
7. Broadleaved evergreen	
8. Deciduous-broadleaved evergreen mix	3. Deciduous-evergreen mix
9. Deciduous-pine mix	
10. Deciduous-whitecedar mix	4. Deciduous
11. Deciduous with evergreen understory	
12. Deciduous with deciduous understory	
13. Deciduous (hydric species)	
14. Agriculture and other	5. Agriculture and other
15. Water	6. Water

Figure 2. Vegetation Categorization of the Great Dismal Swamp.

using two iterations of a histogram normalization and smoothing routine (Johnson and Rohde, 1980). Twenty-nine geometric control points were then located on the Landsat scene and on orthophotoquad maps. The image and ground coordinates of these points were used to derive coefficients for a first-order nearest neighbor geometric registration. Finally, the swamp boundary, digitized from the vegetation map, was used to mask out all pixels outside the swamp. These pixels were set equal to zero to avoid classifying the urban and agricultural areas not within the study area. The result of the above procedures was four bands of spectral data for the Dismal Swamp, in the form of four 810 by 420 matrices of 50-meter square pixels.

A texture image (Logan et. al., 1979) was added to the spatial data base to incorporate information on the variability of spectral data within a spatial neighborhood. Logan suggests that texture measures may be considered indicative of cover type homogeneity. Large, pure stands of vegetation would be expected to have low spectral neighborhood variability, and consequently low texture values. Edges between cover types or stands of mixed vegetation types may be expected to exhibit high spectral variability in a spatial neighborhood, and result in high texture values. To create this texture image, the four-band spectral data was reduced to two bands using principal components techniques. Next, each pixel in the texture image was assigned the value of the standard deviation of the first principal component of each of its eight nearest neighbors. These values were then scaled to integer numbers between 0 and 255.

Garrett and Carter (1977) state that "the single most important environmental factor in the (Dismal) swamp is the water regime." Hydrologic factors such as ground water levels, soil depth, and drainage conditions are often observed to "exert a strong influence on distribution (of vegetation types)", (Carter, et. al., 1978). Therefore, hydrologic variables seem natural choices for ancillary data types in the classification of a wetland. Consequently, two data sets were included in the digital data base to supply information on important hydrologic parameters. These were land surface elevations and organic soil depths.

In wetland ecosystems such as the Dismal Swamp, microtopography may be a very significant factor in the water regime, especially as it pertains to flooding. Regeneration of wetland types such as Atlantic white-cedar may depend greatly on the effect of microrelief on moisture conditions in the seedbed (USDA, 1965). Digital elevation data such as may be obtained from the U.S. Geological Survey or Defense Mapping Agency are of too low a precision to adequately model the topography of the swamp; however, extensive topographic surveys of the Dismal Swamp have provided elevation data at a precision of 0.01 foot (0.021mm). These data were used by Caruso and Paschal (1980) to create a ground surface model of the swamp. Their work included the digitization of all points for which elevation was known, interpolation of these values in unsurveyed portions of the swamp, and the placement of these values in a north/south - east/west oriented grid of 200-meter square cells. This work was performed using the Unitech Contour Plotting System (CPS-1) on a Honeywell computer. Further data editing and reformatting was done on the Honeywell Multics system and a Harris minicomputer, both at the USGS National Center in Reston, Virgina. All that remained to be done before incorporation of these data into the data base was to resample the 200-meter cells to reduce the cell size to 50 meters. Again, pixels outside the swamp boundary were set to zero to insure uniformity with other layers in the data base.

The organic soils of the Dismal Swamp may be considered both causes and effects of the vegetation communities that exist on them. While the peat soils are a result of the accumulation of litter from the forest canopies above them, they also have an important influence on maintenance of soil moisture conditions favorable to wetland forest types. Although the depth of these soils has been altered significantly in some areas, it was still believed that organic soil depth may have a correlation with the distribution of certain vegetation types, and thus was included as an ancillary variable in the data base. A soil survey conducted in the swamp has provided the depth of organic soil at over 200 locations throughout the southern portion of the study area. In addition, peat depths were recorded for the borings of numerous water table wells scattered through the entire swamp. In the same manner as the elevation data, these peat depths were digitized, interpolated, and registered to the 50-meter grid cells. Due to the smaller amount of data in the northern part of the swamp and limitations in the interpolation algorithm, certain areas of the swamp totally lack peat depth data.

The structure and content of the spatial data base for the Dismal Swamp is summarized in Figure 3.

Pixel Size:	50 meters by 50 meters (0.4 hectares, 0.89 acres)
Image Size:	810 rows by 410 columns (320400 total pixels, 233681 pixels within boundary of study area; study area size is 93472.4 hectares)
Location:	UTM Zone 18; northwest corner of study area at 4071300 m north, 360200 m east
Contents:	Landsat band 4 (green) Landsat band 5 (red) Landsat band 6 (near infrared) Landsat band 7 (far infrared) Vegetation level IV (76 classes) Vegetation level III (15 classes) Vegetation level II (6 classes) Spectrally-derived texture Elevation Peat Depth Swamp boundary

Figure 3. Structure and Contents of the Great Dismal Swamp Spatial Data Base.

Classification Procedures

Three classification procedures were examined in this study to evaluate the utility of ancillary data in a cover-type classification of Landsat data for the Dismal Swamp. These included a standard unsupervised approach, a two-stage unsupervised procedure employing spectral stratification (Rohde, 1978),and a three-stage layered procedure incorporating ancillary data in the third stage. All of the above procedures were accomplished using the IDIMS system and statistical programs on a Burroughs-6700 at the EROS Data Center.

A typical unsupervised classification of the spectral data (classification A) was performed as a control classification for later comparison with the layered procedure. The first step in this procedure was the clustering of a systematic sample of data (27% of all non-zero pixels) which resulted in 29 computer classes (Figure 4). The remainder of the data were then placed into these classes using a maximum likelihood algorithm. Finally, visual comparison of the classified image with the vegetation cover type map led to the assignment of each of the 29 computer classes to one of the 15 level II vegetation categories. A check of this assignment scheme was performed by computerized crosstabulation of the sample from the classified image and the vegetation image. The final assignment of computer classes to vegetation types is shown in Figure 5.

A second classification scheme (classification B) was developed involving a two-step spectral stratification as described by Rohde (1978). In the first step of this classification procedure, spectral data were split into four classes using an unsupervised classification algorithm (Figure 6). These broad classes represent an approximate classification of water, evergreen, deciduous, and mixed classes. These four classes were then clustered and classified independently into one, nine, five, and five classes, respectively, using an unsupervised procedure. The first step of this two-step spectral classification reduced the variability of the data to be classified, and allowed independent manipulation of the clustering criteria in the second step. Ideally, this procedure results in the extraction of subclasses of more detail from the few broad first-stage classes. The assignment of the twenty computer classes derived from the above procedure was performed as before, using visual comparison of the displayed classified image with a vegetation map, and crosstabulation of a sample from each image. The class assignments are given in Figure 7.

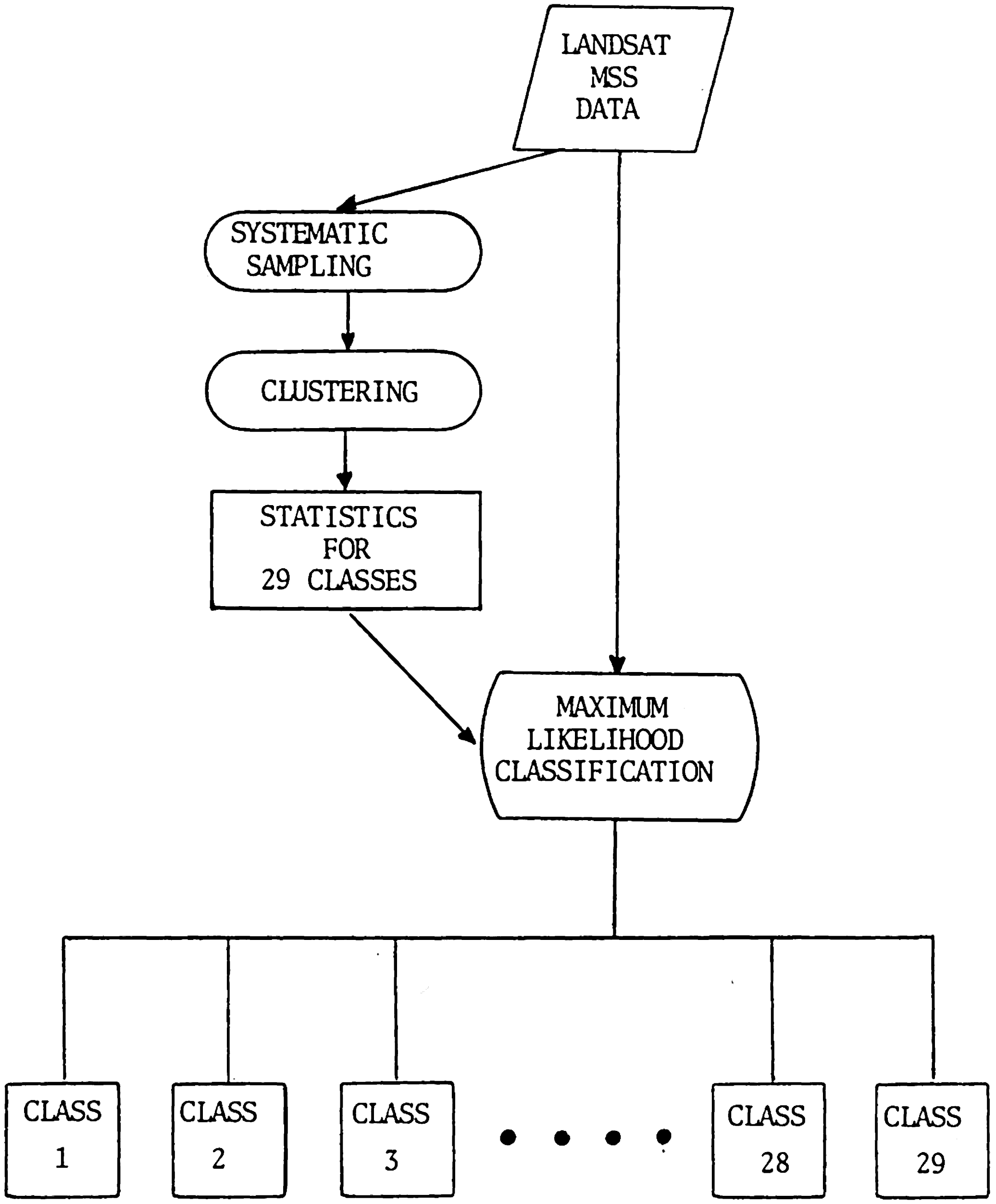


Figure 4. Flow chart of single-stage unsupervised classification.

Spectral Classes	Vegetation Groups (Level III)
8, 12, 14, 16, 19	1 (Pine)
22, 28	2 (Pine-deciduous mix)
3, 9	9 (Deciduous-pine mix)
5, 21, 24, 29	11 (Deciduous with evergreen understory)
2, 11	12 (Deciduous with deciduous understory)
4, 7, 10	13 (Deciduous-hydric species)
13, 15, 17, 18, 20, 23, 25 26, 27	14 (Agriculture and other)
1, 6	15 (Water)

Figure 5. Assignment of Spectral Classes to Vegetation Groups, Single-Stage Unsupervised Classification.

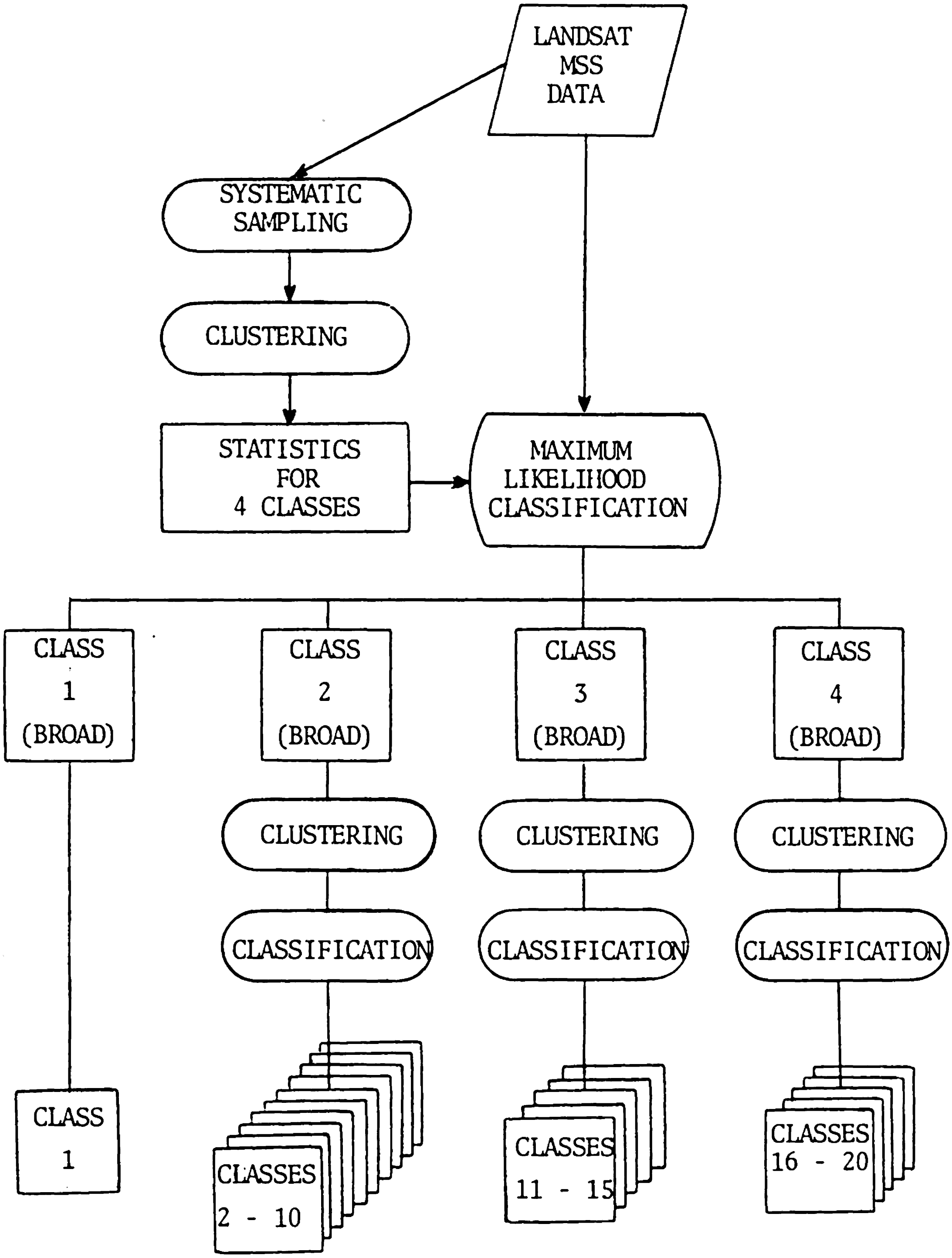


Figure 6. Flow chart for two-stage spectral stratification classification.

Spectral Classes	Vegetation Groups (Level III)
14, 15	1 (Pine)
11, 16, 19	2 (Pine-deciduous mix)
8	4 (Whitecedar-deciduous mix)
12, 13	9 (Deciduous-pine mix)
2, 9	11 (Deciduous with evergreen understory)
4, 5	12 (Deciduous with deciduous understory)
3, 6, 7, 10	13 (Deciduous-hydric species)
17, 18, 20	14 (Agriculture and other)
1	15 (Water)

Figure 7. Assignment of Spectral Classes to Vegetation Groups, Two-Stage Spectral Stratification Classification.

The final classification scheme (classification C) used the ancillary data sets of elevations, peat depths, and texture values at the third stage of a layered procedure. Where confusion existed in the twenty computer classes derived in the spectral stratification procedure, discriminant analysis of ancillary variables in a third step provided a further breakdown of computer classes into component vegetation types (Figure 8).

For example, crosstabulation of a sample of pixels from computer class K with the vegetation image indicated that class K was composed mainly of vegetation types A, B, and C. In the spectral stratification procedure, the entire set of pixels in class K was assigned to only one of the three vegetation types, say B. Therefore, there existed error whenever a member of class K belonging to vegetation types A or C was assigned to B. To reduce this error, class K was split using ancillary data in the layered procedure into three subclasses that would more accurately correspond to vegetation types A, B, and C. In all cases, there was no intuitive basis for this further splitting, because an exact relationship between the ancillary variables and the cover types was not known. However, use of linear discriminant analysis is appropriate for such a task, and has been documented in a similar situation by Shasby, et. al. (1980).

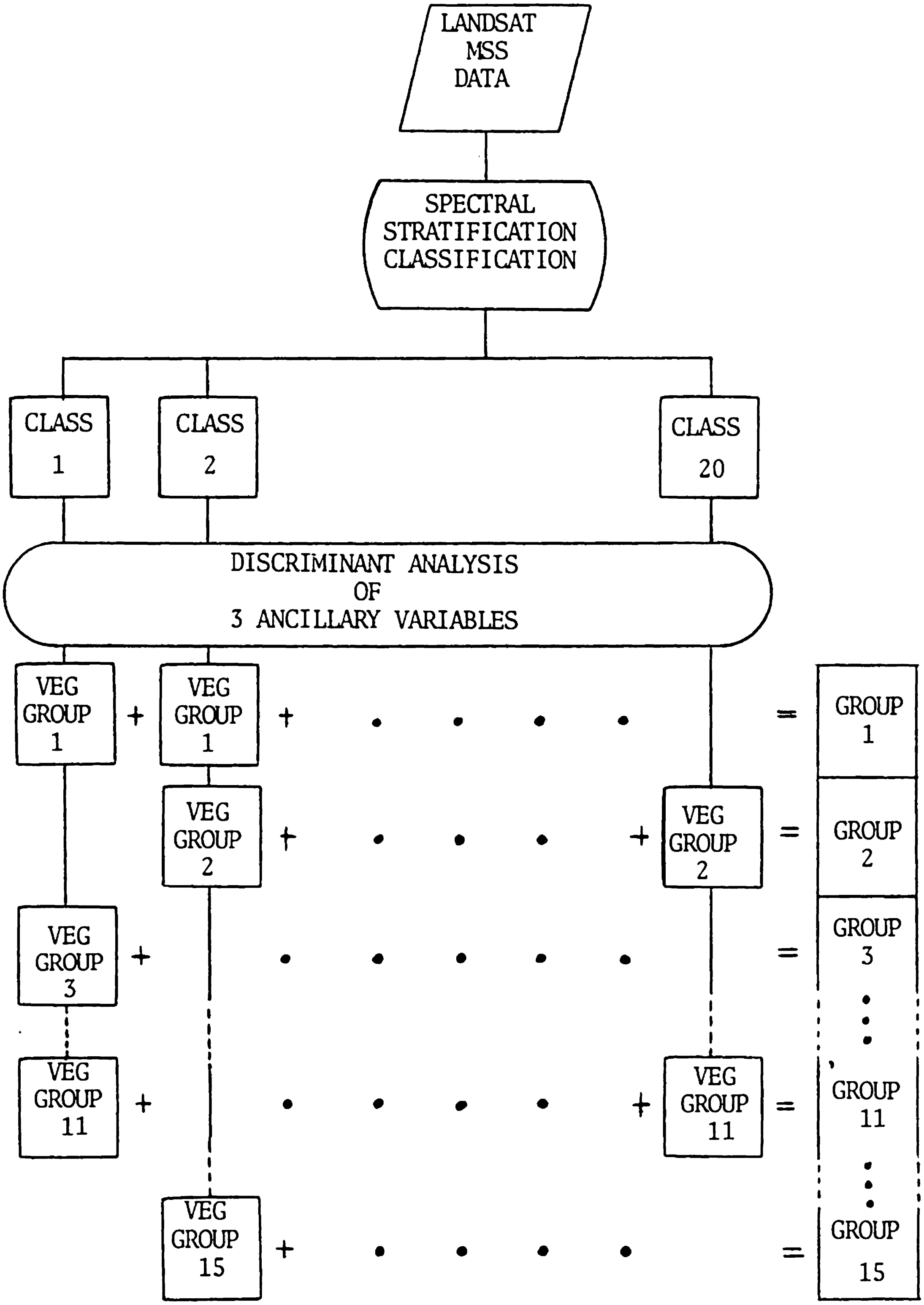


Figure 8. Flow chart of three-stage layered classification of spectral and ancillary data.

Discriminant Analysis Procedures

Discriminant analysis is a powerful multivariate statistical tool, and has been used extensively for remote sensing problems (see chapter 12, Manual of Remote Sensing, 1975). In this project, it is used in a classification sense to assign cases (pixels) to groups (vegetation types) using discriminating variables (elevations, peat depths, and texture values) assumed to be associated with cover type. Discriminant analysis involves the use of a set of observations of known group (cover type) identity to estimate coefficients for discriminant functions. Observations of unknown group identity may then be classified into groups based on their score from the discriminant functions (Tatsuoka, 1971). These classification functions are equivalent to those used in maximum likelihood classification algorithms. Therefore, the discriminant analysis procedure of coefficient calculation and subsequent classification is identical to a supervised maximum likelihood classification. In both cases, a sample of pixels of known identity are used to build statistical models describing the groups they belong to, in terms of the measurements made on each pixel. Then, both procedures may classify unknown observations into groups based on the sample statistics.

The use of discriminant analysis also provides valuable information for further interpretation of the data. Among these are tests to evaluate the success of the discriminating functions, comparisons of the relative importance of the discriminating variables, and descriptions of the locations of group centers in n-dimensional space (Klecka, 1975).

Therefore, discriminant analysis was used in this project to further discriminate "confused" spectral classes, and to gain additional insight into the interaction of the ancillary variables with the vegetation cover types. Specifically, answers to several questions were sought prior to the use of the ancillary variables in classification.

1. How well do the ancillary variables perform in subdividing of the various spectral classes into component vegetation groups (i.e. in the presence of a previous spectral classification)?

2. How well do the ancillary variables alone discriminate between vegetation groups in the absence of any spectral classification?

3. What is the effect of the four spectral bands used as discriminating variables with and without the ancillary variables?

4. Was the second step in the spectral stratification necessary, or, does subdivision of the four broad spectral classes from step one into vegetation groups using ancillary variables perform any worse than a similar subdivision of the twenty spectral classes from step two?

5. What interpretations can be drawn about the relationships of the ancillary variables to vegetation type distribution?

In order to answer these questions, four discriminant analysis procedures were performed, using a sample of pixels from the Dismal Swamp data base. A stratified sample of pixels was taken from the image classified by spectral stratification. Each of the twenty classes represented a stratum, from which a sample was allocated proportional to the number of pixels in the class. A thirty pixel minimum was imposed to prevent the creation of singular covariance matrices later in the procedure. Also, no more than 500 pixels were chosen from any class. A total of 5039 pixels were selected, representing roughly a two percent sampling intensity. For each pixel in the sample, the spectral class number, vegetation cover type, elevation, peat depth, texture, and spectral data were recorded.

The DISCRIMINANT routine from the SPSS statistical package (Nie, et. al.,1975) was used to conduct discriminant analyses of the sample data. Part of this analysis was performed on a Burroughs 6700 mainframe at the EROS Data Center, and the remainder on an IBM 370 at VPI & SU in Blacksburg, Virginia.

Some general conventions were applied to all the discriminant analyses of the sample data. First, variance-covariance matrices of ancillary data for the vegetation groups were not assumed equal, so pooled dispersion matrices were not used. Second, missing values for elevation and peat depth data eliminated many pixels in the sample from consideration in the discriminant analyses. These missing data reduced the number of observations that could be used, which in several cases led to creation of singular (noninvertable) variance-covariance matrices. When a matrix could not be inverted, the pooled within-groups dispersion matrix was used. Third, variables were entered into the analyses using a stepwise variable screening procedure, based on an F-test for significance. Fourth, in the classification of observations, prior probabilities proportional to group size were included. Finally, observations in the "training" sample were classified into vegetation groups, and error matrices were compiled as indicators of the relative performance of the various procedures. Kappa statistics were then calculated for each error matrix and compared using a t-test (Congalton and Mead, 1981).

Four discriminant analyses (hereafter referred to as procedures one through four) were conducted to answer the aforementioned questions as follows. First, to analyze the effectiveness of the ancillary variables in subdividing spectral classes, procedure 1 involved discriminant analysis performed independently on each of the twenty classes using a subset of sample pixels belonging to that class, and the three ancillary variables as discriminating variables. For example, all pixels assigned during the spectral classification to class five were considered together in one discriminant analysis run. It was noted that component vegetation groups comprising class five included groups 9, 11, 12, and 13. Therefore, these four vegetation types were identified as the groups into which class five was to be split. It can be seen that this procedure involved twenty seperate discriminant analysis runs, and exemplified the three-step layered classification with spectral data at the first and second step, and ancillary data at the third.

Second, to evaluate the ability of the ancillary variables to discriminate between vegetation groups in the absence of a spectral classification, procedure 2 used the entire sample of 5039 pixels together in one DISCRIMINANT run. This procedure essentially ignored any previous spectrally-based identification of observations, and classified them solely on the basis of the three ancillary variables.

Procedure 3A involved discriminant analysis of the entire sample together using the four-band Landsat radiance values and the three ancillary variables simultaneously as discriminating variables. In procedure 3B, the entire sample was analyzed using the four spectral bands alone as discriminating variables. These procedures, therefore, represent single-stage supervised classifications of the Landsat data with and without ancillary data.

Procedure 4 was conducted using data from the four broad spectral classes independently. All observations belonging to a given class were subjected together to a discriminant analysis using the three ancillary data types as discriminating variables. This procedure, therefore, involved four runs of the DISCRIMINANT routine, and represented a two-stage classification with spectral data at the first, and ancillary data at the second step.

The stepwise variable selection technique chooses at each step the variable which will provide the most additional "discriminating power", in terms of the greatest increase in Wilks' lambda (Klecka, cited in Nie, et. al., 1975). Therefore, the order in which the discriminating variables entered the analyses may be seen as an indication of their relative strength of association with vegetation cover types. Also, analysis of variance of the ancillary variables may be used to locate and interpret the distribution of the vegetation types along the gradients or axes of the ancillary variables. This may then lead to suggestions as to the biological meaning of ancillary variables in relation to vegetation distributions.

The above discussion has presented the various discriminant analyses performed on the sample data set. Only one procedure, however, was applied to the classification of the entire data set for the Dismal Swamp. This was the method depicted in Figure 8, and tested on the sample in the first procedure outlined above. The results from the discriminant analysis indicated which of the variables were useful in splitting each of the twenty spectral classes into component vegetation groups. The only data needed for the classification algorithm to perform this splitting are the vector of ancillary variable means for each vegetation group, the variance-covariance matrix for each group, and any prior probabilities to be included in the algorithm. Therefore, these data were compiled from the discriminant analyses and used in a maximum likelihood classification of the entire Dismal Swamp data set.

Accuracy Assessment

It should be noted that due to extensive areas lacking peat or elevation data, many pixels remained unclassified. Those that were classified were then crosstabulated with the vegetation image in an error matrix. This produced a 100% enumeration of misclassifications and correctly classified pixels, and eliminated the need for statistical comparisons of accuracy percentages. Kappa values were calculated for each classification of the data set to reflect the improvement over a random classification of the same data.

RESULTS AND DISCUSSION

Discriminant Analysis of Sample Data

The discriminant analyses of the sample data provided some insight into the relationship between ancillary variables, spectral data, and the vegetation cover types. The results from the use of discriminant analysis of the 20 spectral classes in procedure 1 are given in Table 1. Discriminant analysis of each class produced a seperate error matrix. The diagonals of these 20 error matrices are given in the rows of Table 1. The column margin represents the classification accuracy for a given vegetation type, aggregated across all spectral classes. The row margin indicates the overall accuracy obtained within each of the spectral classes. Table 2 presents a single overall error matrix for procedure 1. This was obtained by summing the 20 individual error matrices from the 20 classifications of spectral classes. Discriminant analysis of the 20 spectral classes independently resulted in an overall accuracy of 58.44 percent.

Table 1. Proportions of Correctly Classified Pixels in Discriminant Analysis Procedure 1.

Vegetation Groups^2

Spectral Classes	1	2	3	4	5	6	7	8
1	-	-	-	0/1	-	-	-	-
2	0/13	3/30	3/5	0/1	-	5/5	0/17	0/8
3	0/3	0/2	2/2	8/9	0/1	-	0/3	-
4	-	3/7	0/1	6/17	0/1	0/1	1/10	3/4
5	-	2/10	0/1	0/1	-	-	2/5	0/2
6	-	0/1	0/1	0/1	-	-	-	-
7	0/2	0/3	-	0/2	-	-	0/1	0/2
8	0/1	2/5	5/19	113/129	-	-	1/13	0/8
9	1/3	0/3	0/2	0/7	-	-	0/4	0/2
10	-	-	1/2	-	-	-	-	-
11	0/1	8/11	-	2/2	-	-	0/2	-
12	3/3	0/19	0/10	18/28	0/2	0/1	12/27	10/23
13	14/41	7/39	4/4	6/25	10/17	3/3	16/30	4/25
14	155/174	3/19	1/5	1/5	28/36	1/2	24/26	1/3
15	10/19	0/6	0/2	8/9	0/2	-	1/12	-
16	-	14/14	-	-	-	-	-	-
17	1/2	9/9	0/1	-	-	-	0/2	-
18	-	0/1	-	-	-	-	-	-
19	-	14/14	-	-	-	-	-	-
20	0/1	2/2	-	-	-	-	-	-
Total	184/263	67/195	16/55	162/237	38/59	9/12	57/152	18/77
%	69.96	34.36	29.10	68.35	64.61	75.00	37.50	23.38

Table 2. Classification Results for Discriminant Analysis Procedure 1.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	184	16	2	9	9	1	2	5	28	1	31	2	4	4	0	298
2	1	67	0	0	0	0	3	2	5	0	18	0	1	0	0	97
3	1	2	16	5	1	0	2	3	1	0	6	2	1	1	0	41
4	9	6	23	162	4	0	27	12	6	7	20	10	15	0	0	301
5	20	22	2	8	38	0	3	5	1	2	2	1	1	0	0	105
6	1	1	0	0	0	9	2	0	6	0	1	0	1	0	0	21
7	5	1	0	7	1	0	57	1	1	2	6	4	3	0	0	88
8	1	5	0	2	1	0	4	18	16	0	7	1	6	0	0	61
9	10	13	2	4	0	1	12	4	181	3	38	24	24	9	0	325
10	1	0	1	1	2	0	1	3	2	15	4	1	2	0	0	33
11	14	35	3	20	2	1	14	14	61	7	187	30	43	2	0	433
12	7	18	3	11	0	0	19	7	60	2	98	331	100	8	0	664
13	6	9	2	7	1	0	4	3	34	1	47	42	435	3	0	594
14	3	0	1	0	0	0	2	0	4	0	2	1	0	82	0	95
15	0	0	0	1	0	0	0	0	0	0	12	0	0	0	168	181
Total	263	195	55	237	59	12	152	77	406	40	479	449	636	109	168	3337

Number of Correctly Classified Pixels = 1950
Percentage Correctly Classified = 58.44
Kappa Value = 0.53167

[1] Procedure 1 involved independent discriminant analsis of each of the 20 spectral classes using ancillary variables.

[2] See Figure 2.

Procedure 2 involved discriminant analysis of the sample data with the ancillary variables without regard to the previous spectral classification. Table 3 indicates the results of this classification based on the three ancillary variables alone. An overall accuracy of 39.25 percent was obtained.

Third, the four bands of spectral data were used as discriminating variables with and without the ancillary data. Results of these two classifications are given in Tables 4 and 5, respectively. Use of seven spectral and ancillary discriminating variables resulted in an accuracy of 57.76 percent, while use of four spectral variables alone provided a 38.66 percent correct classification.

Fourth, subdivision of each of the four broad spectral classes in procedure 4 resulted in the error matrix shown in Table 6. Using this procedure, 43.58 percent of the pixels were correctly classified.

Results of the various discriminant analyses are summarized in Table 7. The z-statistics presented here indicate the significance of the kappa value. A z-statistic greater than 1.96 suggests that kappa is significantly greater than

Table 3. Classification Results for Discriminant Analysis Procedure 2.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	66	1	0	0	0	1	2	2	6	0	20	15	10	0	0	123
2	0	26	0	0	0	0	0	0	6	0	0	0	1	0	0	33
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	5	7	21	129	17	0	31	14	0	15	32	15	36	0	0	322
5	11	16	2	15	21	0	0	13	0	1	12	4	10	0	0	105
6	3	0	0	3	0	5	1	0	12	0	2	0	4	0	0	30
7	2	0	0	2	0	0	10	1	0	1	3	1	0	0	0	20
8	0	0	0	0	0	0	0	4	4	1	4	1	0	0	0	14
9	1	18	1	0	0	0	1	4	50	2	10	1	16	0	2	106
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	77	24	28	27	13	0	23	17	63	10	125	9	66	0	17	499
12	23	27	0	42	0	4	59	4	28	5	75	295	155	0	0	717
13	45	22	3	15	5	2	4	7	120	2	41	39	172	1	0	478
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	4	0	0	0	3	0	5	7	29	3	35	0	14	0	159	259
Total	237	141	55	233	59	12	136	73	318	40	359	380	484	1	178	2720

Number of Correctly Classified Pixels = 1062
Percentage Correctly Classified = 39.04
Kappa Value = 0.30927

---

[1] Procedure 2 involved discriminant analysis of the entire sample using ancillary variables.

[2] See Figure 2.

Table 4. Classification Results for Discriminant Analysis Procedure 3A.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	197	16	3	4	6	0	22	6	34	1	31	5	0	0	1	326
2	1	30	0	3	0	0	3	2	6	1	2	0	1	0	1	50
3	2	0	14	8	0	0	2	1	1	0	3	0	1	0	0	32
4	2	4	29	174	1	0	22	7	6	12	17	2	15	0	0	291
5	18	28	0	4	43	0	1	9	4	4	7	2	0	0	0	120
6	0	0	0	0	0	12	0	0	3	0	1	0	0	0	0	16
7	4	2	0	10	3	0	46	2	3	1	8	1	1	0	0	80
8	2	6	1	11	4	0	13	34	36	4	31	3	5	0	0	150
9	4	25	1	2	1	0	5	6	154	3	54	7	26	0	0	288
10	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
11	6	8	7	11	1	0	9	6	34	9	98	21	47	0	5	262
12	0	17	0	3	0	1	11	0	19	3	65	305	100	0	0	523
13	1	5	0	3	0	0	2	0	18	1	27	34	285	1	1	378
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	15	0	3	0	170	188
Total	237	141	55	233	59	12	136	73	318	40	359	380	484	1	178	2706

Number of Correctly Classified Pixels = 1563
Percentage Correctly Classified = 57.76
Kappa Value = 0.52697

---

[1] Procedure 3A involved discriminant analysis of the entire sample using four spectral and three ancillary variables.

[2] See Figure 2.

Table 5. Classification Results for Discriminant Analysis Proceudre 3B.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	115	45	2	16	12	0	39	12	51	5	39	6	8	2	0	352
2	5	37	4	8	0	0	2	0	4	0	28	0	10	13	0	111
3	0	0	3	0	0	0	0	0	0	0	3	0	1	0	0	7
4	0	12	24	151	1	0	13	11	13	14	14	4	12	0	0	269
5	183	14	2	1	30	0	11	0	22	2	12	5	3	0	0	285
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	0	0	2	0	0	0	2	1	0	0	0	6
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	64	117	1	43	18	6	57	29	190	16	148	75	43	18	2	827
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	18	8	6	7	0	0	5	1	22	1	25	5	13	53	0	164
12	30	94	7	12	2	6	39	31	276	11	293	683	374	5	6	1869
13	11	13	5	21	0	0	5	5	52	0	121	102	490	38	5	868
14	1	14	0	0	0	0	0	0	2	0	7	2	13	57	1	97
15	0	0	1	1	0	0	0	0	0	0	12	0	5	0	165	184
Total	428	354	55	260	63	12	173	89	632	49	704	883	972	186	179	5039

Number of Correctly Classified Pixels = 1948
Percentage Correctly Classified = 38.66
Kappa Value = 0.28973

---

[1] Procedure 3B involved discriminant analysis of the entire sample using four spectral variables.

[2] See Figure 2.

Table 6. Classification Results for Discriminant Analysis Procedure 4.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	173	29	6	14	7	1	14	16	67	6	70	23	6	0	0	432
2	1	27	0	0	0	0	0	0	6	0	0	0	1	0	0	35
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
4	2	2	19	114	5	0	21	10	0	10	31	11	24	0	0	249
5	22	30	3	17	40	0	2	17	0	5	2	1	0	0	0	139
6	4	0	0	0	0	11	3	0	15	0	5	0	3	0	0	41
7	9	0	2	14	1	0	53	4	0	1	5	9	1	0	0	99
8	2	0	0	0	0	0	2	4	1	1	4	0	0	0	0	14
9	1	21	1	0	0	0	2	4	72	2	11	4	11	0	0	129
10	0	0	0	1	1	0	1	1	1	0	2	0	3	0	0	10
11	19	6	19	9	4	0	17	7	49	7	86	5	40	0	0	268
12	0	11	0	33	0	0	17	2	9	2	59	271	151	0	0	555
13	4	15	4	30	1	0	4	8	98	6	72	56	244	1	0	543
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	237	141	55	232	59	12	136	73	318	40	347	380	484	1	0	2515

Number of Correctly Classified Pixels = 1096
Percentage Correctly Classified = 43.58
Kappa Value = 0.35451

---

[1] Procedure 4 involved independent discriminant analysis of each of the four broad spectral classes using ancillary variables.

[2] See Figure 2.

zero (and the classification is significantly better than random) at a confidence level of 0.05. Comparison of the first and second procedures indicates that the spectral classification prior to the discriminant analysis was successful in increasing the accuracy over a classification using ancillary variables alone. Procedures 3A and 3B show that the classification using seven discriminating variables proved better than the classification using only four spectral bands as discriminating variables. This was expected, since increasing the number of discriminating variables increases the number of possible discriminant functions, which equivalently increases the number of axes along which groups can be separated. It appears that classification using spectral and ancillary data at the same stage (procedure 3A) results in an accuracy similar to that obtained with a two-stage classification of spectral and ancillary data (procedure 1). The classifications using either spectral or ancillary data but not both (procedures 2 and 3B) also result in comparable, but lower, accuracies. The subclassification of the four broad spectral classes (procedure 4) gives intermediate results. Apparently this treatment does not utilize the spectral data to its fullest extent, but still provides an improvement over the use of only spectral or ancillary variables.

Table 7. Summary of Discriminant Analyses of Sample Data.

	Procedure	Percent Correct	Kappa Value	Z-Statistic
1:	Independent classification of twenty spectral classes using ancillary variables	58.44	0.53167	55.45921
2:	Classification of entire sample using ancillary variables	39.04	0.30927	29.32753
3A:	Classification of entire sample using four spectral and three ancillary variables	57.76	0.52697	50.16893
3B:	Classification of entire sample using only four spectral variables	38.66	0.28973	36.91422
4:	Independent classification of four broad spectral classes using ancillary variables.	43.58	0.35451	31.94113

Based on the discriminant analysis of the 5039 sample observations, some inferences may be made about the classification of the entire data set. It is evident that use of the ancillary and spectral data together provides a better classification than could be obtained using either one singly. It also seems that a more rigorous treatment of the spectral data (clustering into twenty rather than only four classes) prior to subclassification with ancillary data may yield better results. This is indicated by the fact that a level I classification (four broad groups) did not yield as high an accuracy as a level III classification (20 classes) prior to the discriminant analysis.

One-way analysis of variance is a suitable technique for evaluating the separability of groups based on a measured variable. The SPSS routine ONEWAY (Nie, et al. 1975) was used with the sample data in this context. Analysis of variance tables and Duncan's Multiple Range tests for elevation, peat depth, and texture are given in Figures 9, 10, and 11, respectively. Overall significant differences between vegetation groups are indicated by the F-statistic in the analysis of variance tables. Large F values imply highly significant differences. The Duncan's multiple range tests indicate where those differences exist. The vegetation groups are ordered in the Duncans figures by their mean value, and those groups that are not significantly different are connected by a line on the right of the figure.

It can be seen that each of the ancillary variables exhibits an overall significant difference between vegetation groups, but vegetation types are not always readily separable based on these variables. For example, in Figure 9, the mean elevation for the pine-deciduous mix (#2) group is different than the mean elevation in all the other groups except the evergreen-vines type (#6). As seen in Figure 10, the mean peat depth for the pine-deciduous mix group is inseperable from seven other groups. Therefore, one may expect more success in separating class 2 from other classes when using elevation as a criteria than when using peat depth. To determine which variable may be best for discriminating between any two types, the three Duncan's figures may be examined. That variable that provides the greatest distance between the vegetation types may be expected to be most potent in their discrimination. For example, by observation of Figures 9, 10, and 11, one notes that peat is the most appropriate variable to use if one desires to discriminate between the white-cedar group and the hydric deciduous group.

- ANOVA Table

Source	df	SS	MS	F	p-value
Between	14	135944.33	9710.31	37.748	<0.0001
Within	4543	1168632.38	257.24		
Total	4557	1304576.00			

- Duncan's Multiple Range Figure

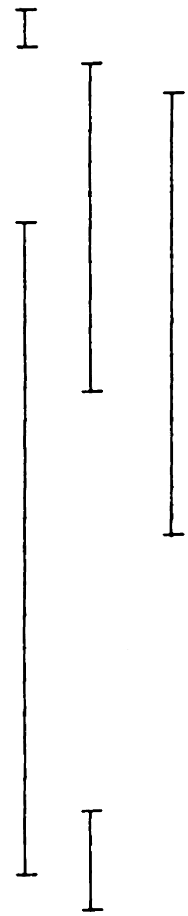
Group No.	N	Mean	Name	Significance
14	127	48.5433	Agriculture	
9	584	52.1627	Deciduous-pine mix	
15	179	52.8603	Water	
12	772	55.3666	Deciduous over deciduous	
1	367	56.6594	Pine	
8	89	56.7865	Deciduous-Broadleaved evg.	
10	49	56.9796	Deciduous-White cedar	
3	55	57.2909	White-cedar	
5	63	57.8571	Inkberry shrub	
13	897	58.5433	Deciduous-hydric	
7	173	58.8324	Broadleaved evergreen	
4	260	61.4846	White cedar-deciduous	
11	593	63.4064	Deciduous over evergreen	
6	12	67.5000	Evergreen vine	
2	338	72.6302	Pine-deciduous mix	
	4558			

Figure 9: Analysis of Variance of Elevation Data

- ANOVA Table

Source	df	SS	MS	F	p-value
Between	13	13142.59	1010.97	31.457	$p < 0.0001$
Within	2695	86611.23	32.14		
Total	2709	99768.06			

- Duncan's Multiple Range Figure

Group No.	N	Mean	Name	Significance
13	487	11.6057	Deciduous-hydric	
9	319	11.6740	Deciduous-pine mix	
6	12	11.7500	Evergreen vines	
1	237	12.2025	Pine	
11	359	13.1393	Deciduous over evergreen	
2	141	13.3333	Pine-deciduous mix	
15	178	13.9720	Water	
12	380	14.6526	Deciduous over deciduous	
8	73	14.6575	Deciduous-broadleaved evergreen mix	
10	40	15.6000	Deciduous-white-cedar mix	
3	55	15.9818	Cedar	
7	136	17.0294	Broadleaved evergreen	
5	59	17.2203	Evergreen shrub	
4	233	18.7382	Cedar-deciduous mix	
	2710			

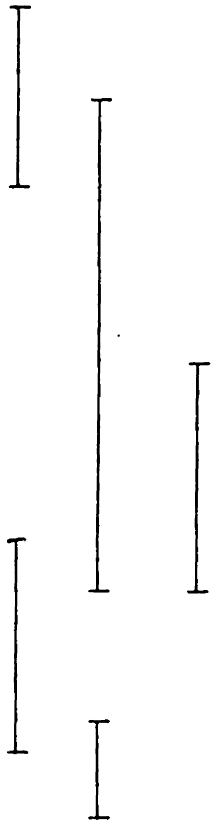


Figure 10: Analysis of Variance of Peat Depth Data

- ANOVA Table

Source	df	SS	MS	F	p-value
Between	14	557171.67	39797.97	29.077	$p < 0.0001$
Within	5024	6876340.49	1638.70		
Total	5039				

- Duncan's Multiple Range Figure

Group No.	N	Mean	Name
15	179	91.6983	Water
6	12	92.4167	Evergreen vine
12	883	98.1982	Deciduous over deciduous
8	89	99.4157	Deciduous-broad-leaved evg.
5	63	99.4762	Evergreen shrub
1	428	102.8248	Pine
13	972	107.1944	Deciduous-hydric
10	49	108.5918	Deciduous-white cedar
7	173	109.5838	Broadleaved ever-green
4	260	110.6731	White-cedar-deciduous
9	632	110.9952	Deciduous-pine mix
11	704	119.3721	Deciduous over evergreen
3	55	122.2364	White-cedar
2	354	126.6893	Pine-deciduous mix
14	186	138.8871	Agriculture
	5039		

Significance

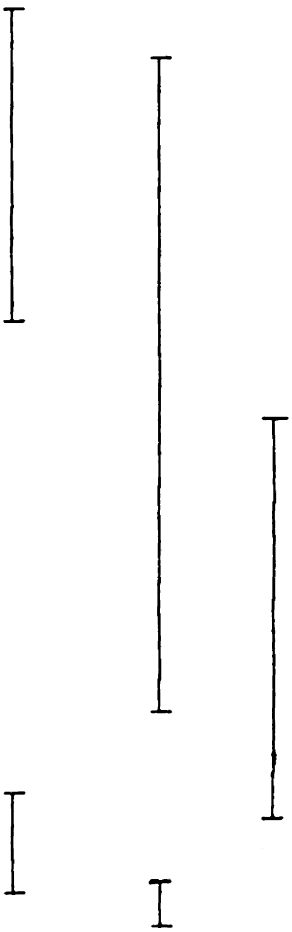


Figure 11: Analysis of Variance of Texture Data

The discriminant analyses of the sample data were addressed to five questions regarding the relationships of spectral and ancillary data types, and their utility in cover-type discrimination. The next analysis involved comparisons of the three classification schemes that were applied to the entire Dismal Swamp data.

Classifications of the Entire Data Set

The accuracies for the three test classifications of the entire Dismal Swamp are given in Table 8. The overall percentage of correctly classified pixels is given for level III and level II classes, along with the corresponding kappa values. It should be noted here that since accuracy assessment was performed on 100% of the data set, statistical comparisons of these results are not needed. The error matrices derived from all classifications may be found in the Appendix.

Since peat depth data were nonexistent for a large portion of the swamp (approximately 24,730 hectares, or about 40% of the study area), discriminant analyses could often not be performed correctly when peat depth was a required

Table 8. Summary of Classification Results for the Three Test Classifications

Classification	% Correct	Kappa
A: Single-stage Unsupervised		
Level III	37.84	0.27053
Level II	58.33	0.34668
B: Sprectral Stratification		
Level III	38.70	0.28345
Level II	58.67	0.36194
C: Layered with Ancillary Data [1/]		
Level III	19.45	0.10996
Level II	39.07	0.14956

[1/] Uses reduced data set.

discriminating variable. Therefore, the layered classification using ancillary variables produced a classification of only the portion of the swamp with a full complement of data. A comparison of this technique with the other algorithms may be biased if the accuracy for portions of the swamp lacking peat depth data differs from the accuracy of the portion with complete data. Therefore, a reassessment of all three classifications was done using a reduced data set containing only those pixels with all necessary data. A summary of these results is given in Table 9. Note that accuracies for the first two classifications did not change greatly. The 16.02% accuracy reported for the level III layered classification was obtained by classifying even those areas lacking data. In such instances, missing data were treated as if the data values were equal to zero. This allowed the classifier to produce a complete classification, although any pixels lacking data for one or more variables had a higher probability of misclassification. This process resulted in a classification of the entire data set that was 3.4% lower in accuracy than the classification of the reduced data set.

An important consideration in the reporting of accuracy figures is the quality of the reference data. Errors in the

Table 9: Classification Results for Full and Reduced Data Sets

	Full		Reduced	
	% Correct	Kappa	% Correct	Kappa
A: Single-stage Unsupervised				
Level III	37.84	0.27053	36.35	0.26981
Level II	58.33	0.34668	59.07	0.38290
B: Spectral Stratification				
Level III	38.70	0.28345	38.81	0.29705
Level II	58.67	0.36194	60.83	0.42267
C: Layered with Ancillary Data				
Level III	16.02	0.07931	19.45	0.10996
Level II	34.39	0.11113	39.07	0.14956

vegetation images may occur for a variety of reasons. From the photointerpretation to the drafting and printing of the map, and the subsequent digitization and registration of the vegetation image, errors may be expected to occur. These errors may bias the results of an accuracy assessment, and should be acknowledged in the interpretation of results. It may be logical to expect that the boundaries of vegetation types on the map and digital image may be more prone to error than the interiors of cover type polygons. This possibility led to another reassessment of accuracy using a reference data source with boundaries extracted. An edge-finding algorithm in the General Image Processing System (GIPSY) of Virginia Tech's Spatial Data Analysis Laboratory was used to locate and mark the boundaries of vegetation polygons in the digital vegetation image. All pixels within 100 meters (two pixels) of an identified vegetation boundary were then deleted from the various classified images. Accuracy assessment of these images then yielded the results given in Table 10. These results are less affected by error in the reference data set, and may represent a more meaningful evaluation of classifier performances.

Note that after eliminating "edge-effects", all classification accuracies improved, although relative to each

Table 10: Classification Results after Excluding Boundaries in Reference Data from Analysis

	Full Data Set		Reduced Data Set	
	% Correct	Kappa	% Correct	Kappa
A: Single-stage Unsupervised				
Level III	43.39	0.32837	41.58	0.31826
Level II	63.59	0.40834	63.77	0.44912
B: Two-stage Spectral Stratification				
Level III	44.23	0.34083	44.05	0.35067
Level II	63.66	0.42976	65.44	0.48626
C: Three-stage Layered				
Level III	17.39	0.09332	21.46	0.12994
Level II	36.14	0.13455	41.71	0.18306

other, the situation remained the same. The first two classifications were comparable, with the accuracy of the spectral stratification classification slightly higher than that of the single-stage unsupervised classification. Again, the layered procedure performed only half as well as the others. This situation is is not at all what was expected, and presents a task in explaining the apparent contradiction.

Explanation of Results

There are two aspects of the results that merit examination and explanation. The first is why the classifications in general produced results as low as they did, and the second is why they produced different results relative to each other.

There are several factors that may have hindered the ability of these procedures to produce accurate classifications. One possible explanation lies in the structure of the vegetation classification framework. As noted in the literature, detailed classes such as the level III groups used here are typically not classified as well as more general level II and level I classes (Fleming and Hoffer, 1979). Therefore, it is not suprising that the rigorous classification into level III categories did not yield results comparable to other studies in which level II classes were used. The accuracies reported here for the classifications aggregated to level II are not, however, unreasonably low when compared to other research efforts (Mead and Meyer, 1977).

Another factor that may have adversely influenced the results is the quality of the spectral data. The raw data (after masking of urban and agricultural areas outside the swamp) had a narrow range and little spectral variability. The Landsat multispectral scanner is capable of discretizing radiance measurements into 128 values for bands four, five, and six, and 64 values for band 7. However, the widest range of values in the Dismal Swamp data was in Band 7, and only spanned from 0 to 30. All three classifications began with unsupervised clustering of the data, and therefore relied on the ability of the clustering algorithm to find "naturally occuring" groups or clusters in the data. The low variability of the spectral data caused the algorithm to go through many iterations with successively smaller clustering criteria (minimum distances between cluster centroids and maximum standard deviations of clusters) at each iteration.

In addition, certain preprocessing algorithms such as destriping, smoothing, and registration involve replacement of a pixel value with the average of its neighbors (Sabins, 1978). This may result in further lowering of the variability of the data prior to training and classification. If such a reduction in variation does indeed interfere with clustering and classification, then such image enhancement techniques may have adversely affected classification accuracy.

The spectral stratification procedure consistently proved better than the single-stage unsupervised procedure. Rohde (1978) developed this technique under the premise that independent clustering of several first-stage groups would allow differential manipulation of clustering criteria so that more important groups can receive more detailed treatment. This is shown in the class assignments in Figures 5 and 7. When a fairly accurate water class (# 1) was defined in the first stage of spectral stratification, no further processing was done on pixels in that class, as only one water class was desired. One clustering iteration within the broad agriculture class (# 4) split it into five classes that revealed additional pine-deciduous mix pixels. An intensive clustering of the deciduous class (#2) created nine classes, including an atlantic white-cedar class and eight classes that could be assigned on the basis of vegetation in the understories. Second-stage treatment of the evergreen class (# 3) allowed recognition of five mixtures of pine and deciduous types.

Conversely, the single-stage unsupervised procedure resulted in 29 classes only after intensive clustering. Note that nine of these were small classes assigned to agriculture, a more rigorous breakdown of that cover type than was needed. Two water classes were formed, and the atlantic white-cedar was missed altogether. It appears that the spectral stratification served its purpose and resulted in an increase in accuracy over the single-stage unsupervised procedure.

The layered procedure must be examined in an attempt to explain the seemingly contradictory decrease in classifier performance. The first observation to be made is that there may be an inconsistency in the logic of the procedure used. As Robinove (1981) pointed out, supervised procedures are based on an _a priori_ land classification scheme, while unsupervised procedures allow the situation and the data to determine how the land will be classified. The layered procedure in this study started as an unsupervised process and ended in a supervised classification. This means that while originally allowing the data to determine groups in the clustering phase, these naturally formed groups were overridden and divided into a priori groups midway through the procedure. While the unsupervised and spectral stratification procedures were allowed to "miss" groups such as the inkberry shrub and evergreen vine communities, the layered procedure forced the assignment of at least some pixels to these groups at the expense of accuracy in other classes. This is illustrated in Figure 12. The column of accuracies on the left represents the results of the assignment of the entire class to a given group, based on figure 7. The accuracies listed on the right came from the discriminant analysis of the class and classification into component vegetation groups. It can be seen that in many cases, the increase in accuracy from splitting the classes was minimal.

A more logical procedure for the layered classifier might have been to use supervised classification at the spectral stages as well as the ancillary stage. This would have provided more consistency in the approach and in the vegetation class framework. Another alternative would be to alter the level III categorization used in the discriminant

Spectral Class	Number of Observations	No. of pixels correctly classified by spectral stratification		No. of pixels correctly classified by discriminant analysis	
		No.	%	No.	%
1	181	168	(92.82)	168	(92.82)
2	341	78	(22.87)	129	(37.83)
3	307	128	(41.69)	168	(54.72)
4	308	79	(25.65)	158	(51.30)
5	221	121	(54.75)	148	(66.97)
6	190	151	(79.47)	159	(83.68)
7	232	184	(79.31)	186	(80.17)
8	226	129	(57.08)	138	(61.06)
9	113	26	(23.01)	32	(28.32)
10	12	5	(41.67)	9	(75.00)
11	31	11	(35.48)	14	(45.16)
12	311	82	(26.37)	130	(41.80)
13	302	50	(16.56)	116	(38.41)
14	312	174	(55.77)	218	(69.87)
15	86	19	(22.09)	31	(36.05)
16	26	14	(53.85)	25	(96.15)
17	69	41	(55.07)	59	(85.51)
18	23	19	(82.61)	20	(86.96)
19	23	14	(60.87)	22	(95.65)
20	23	16	(69.57)	20	(86.96)
Total	3337	1509	(45.22)	1950	(58.44)

Figure 12. Comparison of Accuracy for Two Different Classifications of the Sample Data

analyses such that only those vegetation groups identified in the unsupervised spectral classification would be used in the discriminant analysis. Either of these two possibilities would eliminate the inconsistency of switching from data-defined groups (from the unsupervised clustering) to *a priori* groups (in the discriminant analysis or supervised classification).

In the classification of the whole swamp, a tradeoff is observed between identification of previously missed classes and accuracy in remaining classes. Table 11 is the error matrix from the unsupervised classification of the reduced data set (containing only pixels with a full complement of ancillary data). Table 12 is the error matrix for the layered procedure. Note that the latter procedure identified several groups that the former missed, including groups 3, 4, 5, 7, 8, and 10. However, this came at the expense of correctly classified pixels in groups 1, 9, 12, and 13, resulting in an overall decrease in classification accuracy.

Mere comparison of accuracy figures might indicate that the layered procedure does not measure up to the others. However, if identification of at least some pixels in the classes missed in the unsupervised classification is deemed

Table 11. Results for Classification A, Level III, Reduced Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	9734	839	83	166	1787	9	1376	252	874	67	592	229	41	0	3	16052
2	43	94	32	20	0	0	2	0	4	2	10	0	4	0	1	212
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	2354	3823	1253	3070	1002	269	3500	3473	9013	1136	7771	1412	1390	0	67	39533
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	201	383	185	161	78	129	445	233	710	55	719	654	626	0	32	4611
12	102	905	332	3733	33	144	1077	923	5386	819	6766	8597	3700	0	80	32597
13	133	463	408	2400	11	29	400	295	2245	497	5669	7401	16031	53	147	36182
14	24	139	114	20	7	0	18	0	9	1	23	0	34	0	0	389
15	0	0	0	26	0	0	0	1	0	0	342	0	9	0	4801	5180
Total	12591	6646	2407	9596	2918	580	6818	5177	18241	2578	21892	18293	21835	53	5131	134756

Number of Correctly Classified Pixels = 48989
Percentage Correctly Classified = 36.35
Kappa Value = 0.26481

[1] Unsupervised single-stage spectral classification of only pixels with a full complement of ancillary data.

[2] See Figure 2.

Table 12. Results for Classification C, Level III, Reduced Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	3115	223	154	76	243	0	300	589	2033	137	2962	668	2598	10	1	13109
2	63	332	14	108	2	0	17	65	1127	27	724	436	281	0	1	3197
3	94	227	247	158	26	85	273	145	507	68	430	250	288	0	29	2827
4	77	121	148	1278	12	9	603	51	397	105	771	1162	1250	1	0	5985
5	454	344	123	63	118	76	63	318	1487	69	199	192	21	0	6	3533
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	42	131	149	1029	20	0	289	163	1005	309	2197	1215	1108	0	0	7657
8	45	618	147	534	33	0	310	725	3370	229	3252	777	1121	0	14	11175
9	533	705	292	1036	194	0	266	338	2709	223	2164	2675	1921	0	60	13116
10	511	1822	392	2302	559	302	2931	1613	2614	719	2580	1389	711	0	15	18460
11	3380	1096	273	1488	491	14	439	639	800	433	3106	3907	4600	24	44	20734
12	3931	601	157	327	1101	71	1120	168	1590	38	872	3273	2419	2	30	15700
13	197	360	301	1118	99	23	170	356	584	218	2290	2337	5505	16	138	13712
14	149	66	10	53	20	0	37	6	18	3	14	12	5	0	0	393
15	0	0	0	26	0	0	0	1	0	0	331	0	7	0	4793	5158
Total	12591	6646	2407	9596	2918	580	6818	5177	18241	2578	21892	18293	21835	53	5131	134756

Number of Correctly Classified Pixels = 26209
Percentage Correctly Classified = 19.45
Kappa Value = 0.10996

---

[1] Layered classification of pixels with a full complement of ancillary data.

[2] See Figure 2.

important, the layered procedure has some merit. A weighting of costs of omission and commission errors for each class may therefore change the evaluation of the different procedures. For example, a high cost for ommitting the classes three through seven would make the layered procedure appear much better. In the decision to work with detailed level III classes, just such a weighting was made, although not quantified. The use of the Bayesian classification algorithm, which accounts for variable error costs, might be more appropriate in such a case.

Two other factors may have had a part in lowering accuracy for the layered procedure. First, the training sample may have been inadequate since many observations had missing data. This resulted in non-invertible variance-covariance matrices and eliminated some groups from consideration in the classification of some classes. A further problem with the sample data set may have been misclassifications in the reference data used for training. As noted, exclusion of boundaries in the reference vegetation image increased the evaluated accuracies. This could indicate that some boundary pixels in the "ground-truth" vegetation image were in error, or classification of boundary zones is faulty. If some of these pixels were chosen in the random sample, then the training set contained error and adversely affected the classification. Second, the quality of the ancillary data may be in question. Most pixels were encoded with artificial, interpolated values for elevation and peat depth, and may contain a great deal of error. In addition, the low sampling intensity (less than one sample per 70 to 100 hectares) of the peat depth survey does not support a great deal of confidence in the computer-generated peat depth image. Error in ancillary variables may have had significant effects in both the classification and training phases.

SUMMARY AND CONCLUSIONS

The objective of this study was to develop a procedure for Landsat classification employing ancillary data, apply it to a vegetation cover-type classification of the Great Dismal Swamp, and evaluate its effectiveness relative to two classifications of spectral data alone. An unsupervised classification and a two-stage spectral stratification classification were performed to represent the contribution of spectral data alone to a cover-type classification. The ancillary variables used were thought to be associated with vegetation distribution in a forested wetland. These included land surface elevations and peat depths, both based on field surveys, and spectral texture values indicative of stand homogeneity.

Accuracy assessments were performed for the three classifcations at level II and level III vegetation categorizations. One supplemental accuracy assessment was performed on a reduced data set including only those pixels with a full complement of spectral and ancillary data. This was done to provide a more direct comparison of the two spectral classifications (performed on the entire data set) to the layered classification, which was performed only on the reduced data set. A second supplementary accuracy assessment was conducted after eliminating suspected errors in the reference vegetation image attributed to mislocation of vegetation type boundaries.

The results indicated that fairly low accuracies (20% to 40%) were obtained for all three procedures. The layered classifier using spectral data at the first two stages and ancillary data at the third produced the lowest proportion of correctly classified pixels.

Several factors may have contributed to these unpredictably low results. First of these is the rigorous _a priori_ class structure for the layered classifier in the final stage. Complete identification of detailed classes was demanded, and was acheived only at the expense of accuracy in other classes. A weighting scheme to indicate this demand should be developed and applied equally to accuracy assessment of all three classifications. Otherwise, the groups defined in the clustering of the data should be accepted and used as _a priori_ groups for the discriminant analysis.

Second, low variance of the spectral data may have interfered with the performance of the unsupervised clustering algorithm. Not enough is known about the effects of preprocessing on classification accuracy, but the author suggests that some preprocessing techniques will lower variation in the data, and actually hinder the performance of the clustering algorithms. Two possibilities exist to overcome this effect.

First, supervised procedures are not as adversely affected by low data variation, since they do not rely on the outcome of a clustering algorithm, and thus may be used when a number of "pixel averaging" type preprocessing techniques have been performed. Secondly, clustering may be performed on original data, prior to any preprocessing. This may entail a "backwards" registration of reference data to the geometrically distorted Landsat data, in order to locate pixels in the Landsat scene that correspond to pixels in the reference data. Also, the miscalibrated lines of pixels in a badly striped image may be eliminated from consideration and left unclassified, rather than undergoing radiometric correction. Then, after neighboring pixels (whose radiance values are assumed to be correct) are classified, the unclassified pixels may be assigned to a cover type based on the cover type identity of their neighbors. As indicated, more study may be required to determine the effect of preprocessing on data variation, and the subsequent effect of data variation on clustering performance.

Next, the inconsistent logic of the layered procedure used in this study may have resulted in lower accuracies. It is recommended that consistent vegetation categorizations be used throughout the procedure, whether they are derived through clustering, or known _a priori_. Otherwise, lower classification accuracy may result from a mid-procedural change in definition of acceptable classes.

Finally, data quality will always influence study results, and Landsat classification is no exception. Data of questionable accuracy may have affected classifier performance in several ways. First, as was seen with boundaries in the vegetation cover-type image derived from a photo-interpretation map, errors in reference data may affect not only the training of the classifier, but the assessment of accuracy as well. Future studies involving the use of digitized cover-type maps in training and assessment should consider where errors may occur in the reference map, and follow procedures to correct for this. In this project, exclusion of suspect boundaries was noted to increase reported accuracies, and if done prior to sampling for classifier training, may have improved performance further. Data quality is questionable for both the peat depth and elevation images where values are the result of interpolation. Techniques for creating continuous spatial surfaces of data from a sample of points are used frequently, although little verification of their accuracy has been done. Perhaps assessments of ancillary data accuracy are needed in studies such as this one, although they may be expected to be costly and time-consuming. If ancillary data sets are used for purposes other than Landsat classification (as in a geographic information system), assessment of their quality is more important yet.

Finally, missing data in the peat depth image resulted in a smaller effective sample than was intended, and reduced the quantity of data available for training the classifier. Reducing the data set from which samples were drawn (such that it contains only those pixels with a full complement of ancillary data) would have alleviated this problem.

From this study, increases in classification accuracy as the result of incorporation of ancillary data cannot be claimed. In general, Landsat classifications of the Great

Dismal Swamp did not yield results useful for resource managers, due to unacceptably low accuracies. However, more accurate classifications of less detailed cover-type classes would be no more suitable. Some suggestions have been made for further attempts to include ancillary data in the Landsat classification of vegetation. Although results from this study do not foreshadow greatly improved accuracies, neither do they close the door for further efforts in that direction.

LITERATURE CITED

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

Anuta, P.E., 1976, Digital registration of topographic data and satellite MSS data for augmented spectral analysis, in Proc. of the 42nd Ann. Mtg. of Amer. Soc. of Photogram., Wash., D.C., p.180-187.

Anuta, P.E., Hauska, Hans, and Levandowski, D.W., 1976, Analysis of geophysical remote sensing data using multivariate pattern recognition techniques: in Proc. 1976 Symp. on Mach. Process. of Rem. Sens. Data, Purdue Univ., pp. 1B-11 - 1B-14.

Bryant, J., 1979, On the clustering of multidimensional pictorial data: Pattern Recog. 11(2):115.

Carter, V., and Gammon, P., 1976, Great Dismal Swamp vegetative cover map: U.S. Geological Survey Open-file Map 76-615.

Carter, V., Bedinger, M.S., Novitzki, R.P., and W.O. Wilen, 1978, Water resources and wetlands, in Wetland Functions and Values: The State of Our Understanding, Proc. of the National Symposium on Wetlands, Minneapolis, Minn., pp. 344-376.

Caruso, V.M., and Paschal, J.E., Jr., 1980, Computer modeling of the Great Dismal Swamp, in Coastal Zone '80: Proc. of Symp. on Coastal Zone Management, Miami, Fla., pp 2308-2321.

Clark, J., 1981, Improved land use classification from Landsat and Seasat satellite imagery registered to a common map base, in Proc. of 47th Ann. Mtg. of Amer. Soc. of Photogram., Wash., D.C., pp 591-597.

Congalton, R.G., and Mead, R.A., 1981, A quantitative method to test for similarity between photointerpreters, in Proc. of 47th Ann. Mtg. of Amer. Soc. of Photogram., Wash., D.C., pp. 263-266.

Congalton, R.G., 1981, The use of discrete multivariate analysis techniques for the assessment of Landsat classification accuracy. M.S. Thesis, Virginia Poly. Inst. and State Univ, Blacksburg, Va., 112 pp.

Daily, M.I., Farr, T., Elachi, C., and Schaber, G., 1979, Geologic interpretation from composited radar and Landsat imagery: Photogram. Eng. and Rem. Sens. 45(8):1109-1116.

Davis, J.B., and Friedman, S.Z., 1979, Assessing urbanized area expansion through the integration of Landsat and conventional data: in Proc. of 47th Ann. Mtg. of Amer. Soc. of Photogram., Wash., D.C., pp.776-791.

Ernst, C.L., and Hoffer, R.M., 1979, Using Landsat MSS data with soils information to identify wetland habitats: in Proc. of Fifth Ann. Wm. T. Pecora Symp., Sioux Falls, S.D., pp. 474-478.

Fleming, M.D., and Hoffer, R.M., 1979, Machine processing of Landsat MSS and DMA topographic data for forest cover type mapping: in Proc. of 1979 Symp. on Mach. Proc. of Rem. Sens. Data, Purdue Univ., pp. 377-390.

Gammon, P.T., Rohde, W.G., and Carter, V.C., 1979, Accuracy evaluation of Landsat digital classification of vegetation in the Great Dismal Swamp, in Fifth Ann. Wm. T. Pecora Symp., Sioux Falls, South Dakota, p. 463-473.

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

Hie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H., 1975, Statistical Package for the Social Sciences, 2nd edition: McGraw-Hill, New York, 675 pp.

Hixson, M., Scholz, D., Fuhs,N., and Akiyama,T., 1980, Evaluation of several schemes for classification of remotely sensed data: Photogram. Eng. and Rem. Sens. 46(12):1547-1553.

Hoffer, R.M. and staff., 1975, Computer-aided analysis of Skylab multispectral scanner data in mountainous terrain for land use, forestry, water resource, and geologic applications: Laboratory for Applications of Rem. Sens. Tech. Rep. 121275, Purdue Univ., 381 pp.

Hoffer, R.M., Fleming, M.D., Bartolucci, L.A., Davis, S.M., Nelson, R.F., 1979, Digital processing of Landsat MSS and topographic data to improve capabilities for computerized mapping of forest cover types: Laboratory for Applications of Remote Sensing Tech. Rep. 011579, Purdue Univ.,159 pp.

Hornung, R.J., and Smith, J.A., 1974, Incorporation of spatial information as additional features for multispectral land-use classification: Final Report (Vol. I) for NAS contract NAS 9-12972, 179 pp.

Hutchinson,C.F., 1982, Techniques for combining Landsat and ancillary data for digital classification improvement: Photogram. Eng. and Rem. Sens. 48(1):123-130.

Johnson, G.R., and Rohde, W.G., 1980, Landsat digital analysis techniques required for wildland resource classification, in Proc. of Arid Lands Resource Inventory Conference, La Paz, N.M. pp.204-213.

Kettig, R.L., and Landgrebe, D.A., 1976, Classification of multispectral image data by extraction and classification of homogeneous objects: IEEE Trans. on Geoscience Elec. 14:19-25.

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

Linden, D.S., Rohde, W.G., and Bonner, K.G., 1980, Estimating the area of vegetation types with Landsat data and ancillary data: six approaches, in Arid Land Res. Inventory Workshop, 8 pp.

Logan, T.L., Strahler, A.H., and Woodcock, C.E., 1979, Use of a standard deviation based texture channel for Landsat classification of forest strata, in Proc. of 1979 Symp. on Mach. Proc. of Rem. Sens. Data, Purdue Univ., p. 395.

Logan, T.L., and Strahler, A.H., 1980, Forest stand delineation from unsupervised classification of optimal Landsat spectral Landsat texture, and topographic channels: in Proc. of 1980 Symp. on Mach. Proc. of Rem. Sens. Data, Purdue Univ., p. 338.

Mead, R.A., and Meyer, M.P., 1977, Landsat digital data application to forest vegetation and land use classification in Minnesota, in Proc. of Symp. on Machine Proc. of Rem. Sens. Data, Purdue Univ., pp. 270-280.

Meyer, M., Brass, J., Gerbig, B., and Batson, F., 1975, ERTS data applications to surface resource surveys of potential coal production lands in southeast Montana: IARSL Res. Rep. 75-1, 24 pp.

Miller, L.D., Nualchawee, K., and Tom, C.H., 1978, Analysis of the dynamics of shifting cultivation in the tropical forests of northern Thailand using landscape modeling and classification of Landsat imagery, in Proc. of Twelfth Internat'l Symp. on Rem. Sens. of Env., pp. 1167-1186.

Miller, W.A., Bonner, W.J., Rohde, W.G., and Schwartz, L.P. 1980. Digital Landsat and terrain data applied to an arid land resource inventory, in Proc. of Arid Lands Resource Inventory Conference, LaPaz, N.M.

Price, J.D., 1981, The contribution of thermal data in Landsat multispectral classification: Photogram. Eng. and Rem. Sens. 47(2):229-236.

Reeves, R.G. (editor), 1975, Manual of Remote Sensing: American Society of Photogrammetry, Falls Church, Virginia, 2142 pp.

Robinove, C. J., 1981, The logic of multispectral classification and mapping of land: Rem. Sens. of Environ. 11:231-244.

Rohde, W.G., 1978, Improving land cover classification by image stratification of Landsat data, in Proc. of Twelfth Internat'l Symp. on Rem. Sens. of Env., pp. 729-741.

Rosenfield, G.H., 1981, Analysis of variance of thematic mapping experiment data: Photogram. Eng. and Rem. Sens. 47(12):1685-1692.

Sabins, F. F., Jr., 1978, Remote Sensing: Principles and Interpretation: W. H. Freeman and Co., San Francisco, 426 pp.

Scholz, D.K., and Weismiller, R.A., 1980, Utilization of digital ancillary data in satellite land use mapping, in Proc. of 1980 Symp. on Machine Processing of Rem. Sens. Data, Purdue Univ., p. 226.

Sharp, J.M., 1979, Demonstrating the value of Landsat dataa case for lowered expectations: Photogrammetric Eng. and Rem. Sens. 45(11):1487-1493.

Shasby, M.B., Burgan, R.E., and Johnson, G.R., 1980, Broad area forest fuels and topography mapping using digital Landsat and terrain data, in Proc. of 1980 Symp. on Machine Processing of Rem. Sens. Data, Purdue Univ., pp 142-147.

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

Strahler, A.H., 1981, Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data: Internat'l J. of Rem. Sens. 2(1):15-41.

Strahler, A.H., Logan, T.L., and Bryant, N.A., 1978, Improving forest cover classification accuracy from Landsat by incorporating topographic information, in Proc. of Twelfth Internat'l Symp. on Rem. Sens. of Env., pp. 927-942.

Swain, P.H., Wu, C.L., Landgrebe, D.A., and Hauska, H., 1975, Layered classification techniques for remote sensing applications, in Proc. of Earth Resources Survey Symp., Houston, Tex., pp.1087-1097.

Swain, P.H., and Huaska, Hans, 1977, The decision tree classifier- design and potential: IEEE Trans. on Geoscience Elec., GE-15(3):142-147.

Swain, P.H., and Davis, S.M. (eds.), 1978, Remote Sensing: the Quantitative Approach: McGraw-Hill: New York, 396 pp.

Tatsuoka, M.M., 1971, Multivariate Analysis: Techniques for Educational and Psychological Research: John Wiley and Sons, New York, 310 pp.

Tom, C.H., Miller, L.D., and Christenson, J.R., 1978, Spatial land-use inventory, modeling, and projection-Denver metropolitan area with inputs from existing maps, airphotos, and Landsat imagery: NASA Tech. Memo. 79710, 225 pp.

U.S. Department of Agriculture Forest Service, 1965, Silvics of Forest Trees of the United States, Agriculture Handbook No. 271, U.S. Government Printing Office, Washington, D.C., 762 pp.

Woodcock, C.E., Strahler, A.H., and Logan, T.L., 1980, Stratification of forest vegetation for timber inventory using Landsat and collateral data, in Proc. of Fourteenth Internat'l Symp. on Rem. Sens. of Env., 19 pp.

APPENDIX . . . . . . CLASSIFICATION ERROR MATRICES

Results for Classification A, Level III, Entire Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	13872	1699	83	207	1834	9	1756	269	2583	146	1686	575	481	198	3	25401
2	77	458	32	22	0	0	32	0	51	2	417	1	182	106	5	1385
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	3929	9028	1253	3598	1076	269	4086	3851	12822	1482	9668	6428	3345	7	69	60911
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1306	1546	185	186	122	129	718	285	3958	88	4597	3299	1752	1415	34	19620
12	701	2656	332	4195	35	144	1162	1437	7590	910	8473	20325	6392	65	81	54498
13	750	1081	408	2689	14	29	439	430	4096	552	8122	12379	28784	162	152	60087
14	279	716	114	20	11	0	137	5	718	1	1004	327	465	2766	12	6575
15	15	0	0	26	0	0	0	1	1	1	342	1	16	0	4801	5204
Tot.	20929	17184	2407	10943	3092	580	8330	6278	31819	3182	34309	43335	41417	4719	5157	233681

Number of Correctly Classified Pixels = 88425
Percentage Correctly Classified = 37.84
Kappa Value = 0.27053

[1] Unsupervised single-stage spectral classification of all pixels.

[2] See Figure 2.

Results for Classification A, Level II, Entire Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	16450	3531	3051	3342	304	8	26786
2	0	0	0	0	0	0	0
3	17808	5431	18155	19441	7	69	60911
4	16035	2792	19346	94123	1642	267	134205
5	1129	148	724	1796	2766	12	6575
6	41	0	3	359	0	4801	5204
Total	51463	12002	41279	119061	4719	5157	233681

Number of Correctly Classified Pixels = 136295
Percentage Correctly Classified = 58.33
Kappa Value = 0.34668

[1] Unsupervised single-stage spectral classification of all pixels.

[2] See Figure 2.

Results for Classification B, Level III, Entire Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	14019	1787	436	465	1839	9	1855	277	2774	171	1978	591	639	490	5	27335
2	40	472	7	16	0	0	31	6	52	0	401	3	166	57	5	1256
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	43	360	465	3630	11	0	289	302	473	311	712	155	519	0	8	7278
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	4236	9466	865	3302	1121	377	4270	4002	13593	1492	10588	9030	3809	23	76	66250
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	859	1545	213	229	65	93	621	260	3418	101	4551	2710	1493	1397	31	17586
12	848	2547	201	2676	35	88	929	1220	8080	844	9366	21035	8244	488	101	56702
13	674	679	207	598	13	13	263	210	3075	262	5988	9743	26305	230	132	48392
14	197	328	13	1	8	0	72	0	354	1	394	67	233	2034	6	3708
15	13	0	0	26	0	0	0	1	0	0	331	1	9	0	4793	5174
Tot.	20929	17184	2407	10943	3092	580	8330	6278	31819	3182	34309	43335	41417	4719	5157	233681

Number of Correctly Classified Pixels = 90432
Percentage Correctly Classified = 38.70
Kappa Value = 0.28345

[1] Two-stage spectral stratification classification of all pixels.

[2] See Figure 2.

Results for Classification B, Level II, Entire Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	21740	4034	4366	5164	547	18	35869
2	0	0	0	0	0	0	0
3	17869	5768	19087	23427	23	76	66250
4	11276	2120	17470	89435	2115	264	122680
5	539	80	355	694	2034	6	3708
6	39	0	1	341	0	4793	5174
Total	51463	12002	41279	119061	4719	5157	233681

Number of Correctly Classified Pixels = 137089
Percentage Correctly Classified = 58.67
Kappa Value = 0.36194

[1] Two-stage spectral stratification classification of all pixels.

[2] See Figure 2.

Results for Classification C, Level III, Entire Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	7149	5225	154	564	385	0	1033	1017	2266	506	5263	7572	6765	1031	4	43934
2	1079	1323	14	158	2	0	63	70	2778	68	2934	1555	1266	1072	9	12391
3	582	983	247	173	31	85	420	165	1227	80	2678	1093	1383	235	30	9412
4	99	543	148	1374	12	9	609	75	454	107	1008	1207	1538	3	1	7187
5	1419	422	123	63	118	76	63	335	1875	69	471	513	71	12	6	5636
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	139	1277	149	1363	22	0	317	682	1588	408	2519	6745	2581	15	0	17805
8	939	975	147	542	33	0	571	756	5884	234	4731	8474	2890	469	18	26663
9	844	1010	292	1313	197	0	389	347	3695	245	3160	4043	4719	754	62	21070
10	542	2048	392	2302	559	302	2963	1615	2743	719	2727	1539	950	0	15	19416
11	3483	1311	273	1501	492	14	450	662	1036	475	3748	4656	5728	61	47	23937
12	4019	1280	157	334	1116	71	1139	182	2369	47	1720	3384	4478	60	32	20388
13	245	558	301	1164	101	23	190	359	697	220	2873	2512	8859	48	138	18288
14	377	229	10	66	24	0	123	12	207	4	146	41	180	959	2	2380
15	13	0	0	26	0	0	0	1	0	0	331	1	9	0	4793	5174
Tot.	20929	17184	2407	10943	3092	580	8330	6278	31819	3182	34309	43335	41417	4719	5157	233681

Number of Correctly Classified Pixels = 37441
Percentage Correctly Classified = 16.02
Kappa Value = 0.07931

---

[1] Layered classification of all pixels.

[2] See Figure 2.

Results for Classification C, Level II, Entire Data Set.[1]

Classified Vegetation Groups	Reference Vegetation Groups[2] 1	2	3	4	5	6	Total
1	19815	2649	13813	34262	2341	44	72924
2	4955	596	4957	12900	27	6	23441
3	11346	5014	16238	33233	1223	95	67149
4	14626	3496	6047	37958	169	217	62613
5	682	147	223	367	959	2	2380
6	39	0	1	341	0	4793	5174
Total	51463	12002	41279	119061	4719	5157	233681

Number of Correctly Classified Pixels = 80359
Percentage Correctly Classified = 34.39
Kappa Value = 0.11113

[1] Layered classification of all pixels.

[2] See Figure 2.

Results for Classification A, Level III, Reduced Data Set.[1]

See Table 11, Page 78.

---

[1] Unsupervised single-stage spectral classification of only pixels with a full complement of ancillary data.

Results for Classification A, Level II, Reduced Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	11011	3174	1199	876	0	4	16464
2	0	0	0	0	0	0	0
3	10500	4771	13622	10573	0	67	39533
4	9406	2346	11163	50163	0	0	73390
5	297	25	10	57	0	0	389
6	26	0	2	351	0	4801	5180
Total	31240	10316	25996	62020	53	5131	134756

Number of Correctly Classified Pixels = 79597
Percentage Correctly Classified = 59.07
Kappa Value = 0.38290

[1] Unsupervised single-stage spectral classification of only pixels with a full complement of ancillary data.

[2] See Figure 2.

Results for Classification B, Level III, Reduced Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	9779	867	436	407	1790	9	1451	254	855	93	612	235	51	0	3	16842
2	4	93	7	13	0	0	0	6	2	0	5	2	2	0	1	135
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	207	128	465	3329	11	0	277	272	369	308	660	84	430	0	8	6368
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	2448	3980	865	2793	1043	377	3648	3609	9438	1130	8437	2131	1566	0	74	41529
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	92	349	213	206	22	93	372	216	627	71	950	570	415	1	26	4283
12	135	873	201	2334	33	88	827	678	5442	736	7243	9509	4895	0	97	33091
13	80	288	207	497	13	13	230	141	1499	239	3647	5762	14402	52	129	27199
14	26	68	13	1	6	0	13	0	9	1	7	0	7	0	0	151
15	0	0	0	26	0	0	0	1	0	0	331	0	7	0	4793	5158
Tot.	12591	6646	2407	9596	2918	580	6818	5177	18241	2578	21892	18293	21835	53	5131	134756

Number of Correctly Classified Pixels = 52293
Percentage Correctly Classified = 38.81
Kappa Value = 0.29705

[1] Two-stage spectral stratification classification of only pixels with a full complement of ancillary data.

[2] See Figure 2.

Results for Classification B, Level II, Reduced Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	15555	3538	2159	2081	0	12	23345
2	0	0	0	0	0	0	0
3	10076	5068	14177	12134	0	74	41529
4	5475	1691	9649	47453	53	252	64573
5	108	19	10	14	0	0	151
6	26	0	1	338	0	4793	5158
Total	31240	10316	25996	62020	53	5131	134756

Number of Correctly Classified Pixels = 81978
Percentage Correctly Classified = 60.83
Kappa Value = 0.42267

[1] Two-stage spectral stratification classification of only pixels with a full complement of ancillary data.

[2] See Figure 2.

Results for Classification C, Level III, Reduced Data Set.[1]

See Table 12 , Page 79.

---

[1] Unsupervised single-stage spectral classification of only pixels with a full complement of ancillary data. Pixels on boundaries between cover types were excluded from the evaluation.

Results for Classification C, Level II, Reduced Data Set.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	6435	1570	5251	11820	11	31	25118
2	2335	566	3351	4932	0	6	11190
3	8937	4595	12540	16590	0	89	42751
4	13229	3528	4826	28309	42	212	50146
5	278	57	27	31	0	0	393
6	26	0	1	338	0	4793	5158
Total	31240	10316	25996	62020	53	5131	134756

Number of Correctly Classified Pixels = 52643
Percentage Correctly Classified = 39.07
Kappa Value = 0.14956

[1] Layered classification of only pixels with a full complement of ancillary data.

[2] See Figure 2.

Results for Classification A, Level III, Entire Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	11071	1159	5	71	1542	6	1250	124	1470	68	778	221	197	161	0	18123
2	53	79	14	13	0	0	11	0	8	2	55	1	31	77	0	344
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1973	6148	642	1859	586	190	2990	2252	10314	870	6697	4391	1930	2	5	40849
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	627	875	99	94	55	96	383	169	2433	51	2813	2304	815	1020	6	11840
12	191	1432	142	2420	2	52	615	724	5993	521	6062	17159	4090	27	19	39449
13	315	511	177	1730	4	8	223	184	2651	250	5150	10033	22704	73	70	44083
14	64	343	109	1	4	0	41	0	206	1	490	154	96	2074	0	3583
15	2	0	0	4	0	0	0	1	0	0	121	0	7	0	4460	4595
Tot.	14296	10547	1188	6192	2193	352	5513	3454	23075	1763	22166	34263	29870	3434	4560	162866

Number of Correctly Classified Pixels = 70674
Percentage Correctly Classified = 43.39
Kappa Value = 0.32837

---

[1] Unsupervised single-stage spectral classification of all pixels. Pixels on boundaries between cover types were excluded from the evaluation.

[2] See Figure 2.

Results for Classification A, Level II, Entire Data Set with Boundaries Excluded.[1]

Classified Vegetation Groups	Reference Vegetation Groups[2]						
	1	2	3	4	5	6	Total
1	12465	2809	1672	1283	2388	0	18467
2	0	0	0	0	0	0	0
3	10622	3766	13436	13018	2	5	40849
4	8613	1438	12976	71130	1120	95	95372
5	517	45	207	740	2074	0	3583
6	6	0	1	128	0	4460	4595
Total	32223	8058	28292	86299	3434	4560	162866

Number of Correctly Classified Pixels = 103565
Percentage Correctly Classified = 63.59
Kappa Value = 0.40834

---

[1] Unsupervised single-stage spectral classification of all pixels. Pixels on boundaries between cover types were excluded from the evaluation.

[2] See Figure 2.

Results for Classification B, Level III, Entire Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	11152	1154	222	230	1539	6	1273	112	1514	74	913	228	207	358	0	18982
2	19	86	3	9	0	0	6	5	15	0	58	2	39	39	0	282
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	18	215	251	2467	0	0	154	102	350	190	366	77	205	0	1	4396
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	2135	6392	388	1669	616	247	3159	2348	10846	894	7375	6589	2203	16	5	44882
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	366	909	151	108	17	73	329	155	2090	58	2814	1805	626	952	0	10453
12	276	1353	77	1382	10	23	446	647	6312	439	6640	17591	5496	339	29	41060
13	283	299	85	322	6	3	121	84	1843	107	3759	7964	21036	141	72	36125
14	45	139	11	1	5	0	24	0	105	1	127	7	56	1589	0	2110
15	2	0	0	4	0	0	0	1	0	0	114	0	2	0	4453	4576
Tot.	14296	10547	1188	6192	2193	352	5513	3454	23075	1763	22116	34263	29870	3434	4560	162866

Number of Correctly Classified Pixels = 72034
Percentage Correctly Classified = 44.23
Kappa Value = 0.34083

[1] Two-stage spectral stratification classification of all pixels. Pixels on boundaries were excluded from the evaluation.

[2] See Figure 2.

Results for Classification B, Level II, Entire Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	15826	2979	2362	2095	397	1	23660
2	0	0	0	0	0	0	0
3	10584	4022	14088	16167	16	5	44882
4	5611	1028	11735	67731	1432	101	87638
5	196	29	106	190	1589	0	2110
6	6	0	1	116	0	4453	4576
Total	32223	8058	28292	86299	3434	4560	162866

Number of Correctly Classified Pixels = 103687
Percentage Correctly Classified = 63.66
Kappa Value = 0.42976

[1] Two-stage spectral stratification classification of all pixels. Pixels on boundaries were excluded from the evaluation.

[2] See Figure 2.

Results for Classification C, Level III, Entire Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
1	5099	3460	53	237	240	0	615	658	5225	267	3170	5741	5246	789	0	30800
2	628	679	4	69	0	0	25	36	2091	42	1831	1008	570	791	0	7774
3	251	604	111	82	5	70	201	91	802	43	1614	780	672	141	0	5468
4	56	357	60	921	9	3	324	30	307	59	664	956	1060	2	1	4808
5	843	341	46	39	62	57	26	206	1334	24	283	371	26	1	0	3659
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	55	629	74	820	9	0	176	361	1175	237	1689	5482	1765	0	0	12472
8	332	578	70	267	5	0	377	483	4615	105	3588	7193	1914	326	0	19853
9	480	591	164	846	165	0	137	206	2573	123	1989	2943	2983	545	35	13780
10	228	1387	210	1263	310	190	2322	846	2246	451	1991	1058	540	0	1	13043
11	2766	735	143	767	343	1	234	284	614	273	2350	3957	4378	12	22	16879
12	3207	746	120	166	993	27	941	89	1557	23	1126	2764	3267	22	17	15065
13	120	293	126	668	39	4	97	162	472	114	1720	1991	7412	17	31	13266
14	229	147	7	43	13	0	38	1	64	2	37	19	35	788	0	1423
15	2	0	0	4	0	0	0	1	0	0	114	0	2	0	4453	4576
Tot.	14296	10547	1188	6192	2193	352	5513	3454	23075	1763	22166	34263	29870	3434	4560	162866

Number of Correctly Classified Pixels = 28322
Percentage Correctly Classified = 17.39
Kappa Value = 0.09332

---

[1] Layered classification of all pixels. Pixels on boundaries were excluded from the evaluation.

[2] See Figure 2.

Results for Classification C, Level II, Entire Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	12671	1492	9651	23312	1723	1	48850
2	2847	330	3337	9616	1	0	16131
3	6416	3506	11648	24199	871	36	46676
4	9857	2679	3588	28965	51	70	45210
5	426	51	67	91	788	0	1423
6	6	0	1	116	0	4453	4576
Total	32223	8058	27292	86299	3434	4560	162866

Number of Correctly Classified Pixels = 58855
Percentage Correctly Classified = 36.14
Kappa Value = 0.13455

[1] Layered classification of all pixels. Pixels on boundaries were excluded from the evaluation.

[2] See Figure 2.

Results for Classification A, Level III, Reduced Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Tot.
1	8191	581	5	64	1537	6	1039	117	393	34	229	82	13	0	0	12291
2	37	25	14	12	0	0	0	0	2	2	0	0	0	0	0	92
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1282	2628	642	1643	546	190	2546	2008	7482	650	5659	766	792	0	5	26839
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	82	216	99	83	42	96	296	143	558	28	450	563	397	0	6	3059
12	32	534	142	2195	2	52	580	431	4413	470	5001	7077	2435	0	19	23383
13	51	209	177	1550	4	8	209	124	1630	225	3845	6004	12824	7	70	26937
14	10	80	109	1	4	0	14	0	1	1	0	0	7	0	0	227
15	0	0	0	4	0	0	0	1	0	0	121	0	2	0	4460	4588
Tot.	9685	4273	1188	5552	2135	352	4684	2824	14479	1410	15305	14492	16470	7	4560	97416

Number of Correctly Classified Pixels = 40509
Percentage Correctly Classified = 41.58
Kappa Value = 0.31826

[1] Unsupervised single-stage spectral classification of only pixels with a full complement of ancillary data. Pixels on boundaries between cover types were excluded from the evaluation.

[2] See Figure 2.

Results for Classification A, Level II, Reduced Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	8929	2582	548	324	0	0	12383
2	0	0	0	0	0	0	0
3	6195	3282	10140	7217	0	5	26839
4	5370	1289	8022	38596	7	95	53379
5	200	18	2	7	0	0	227
6	4	0	1	123	0	4460	4588
Total	20698	7171	18713	46267	7	4560	97416

Number of Correctly Classified Pixels = 62125
Percentage Correctly Classified = 63.77
Kappa Value = 0.44912

[1] Unsupervised single-stage spectral classification of only pixels with a full complement of ancillary data. Pixels on boundaries between cover types were excluded from the evaluation.

[2] See Figure 2.

Results for Classification B, Level III, Reduced Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Tot.
1	8224	577	222	215	1534	6	1059	108	370	40	214	89	12	0	0	12670
2	2	25	3	7	0	0	0	5	2	0	1	2	0	0	0	47
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	9	88	251	2265	0	0	144	91	263	190	335	30	164	0	1	3831
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1334	2719	388	1462	576	247	2698	2089	7808	662	6172	1374	910	0	5	28444
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	32	227	151	97	4	73	249	139	514	39	640	475	242	0	0	2882
12	45	505	77	1226	10	23	419	335	4486	380	5334	7757	3402	0	29	24028
13	23	96	85	275	6	3	109	56	1033	98	2495	4765	11739	7	72	20862
14	16	35	11	1	5	0	6	0	3	1	0	0	0	0	0	79
15	0	0	0	4	0	0	0	1	0	0	114	0	1	0	4453	4573
Tot.	9685	4273	1188	5552	2135	352	4684	2824	14479	1410	15305	14492	16470	7	4560	97416

Number of Correctly Classified Pixels = 42911
Percentage Correctly Classified = 44.05
Kappa Value = 0.35067

[1] Two-stage spectral stratification classification of only pixels with a full complement of ancillary data. Pixels on boundaries between cover types were excluded from the evaluation.

[2] See Figure 2.

Results for Classification B, Level II, Reduced Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	11888	2743	1069	847	0	1	16548
2	0	0	0	0	0	0	0
3	5903	3521	10559	8456	0	5	28444
4	2839	896	7080	36849	7	101	47772
5	64	11	4	0	0	0	79
6	4	0	1	115	0	4453	4573
Total	20698	7171	18713	46267	7	4560	97416

Number of Correctly Classified Pixels = 63749
Percentage Correctly Classified = 65.44
Kappa Value = 0.48626

[1] Two-stage spectral stratification classification of only pixels with a full complement of ancillary data. Pixels on boundaries between cover types were excluded from the evaluation.

[2] See Figure 2.

Results for Classification C, Level III, Reduced Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Tot.
1	2500	89	53	45	185	0	96	384	1649	58	2009	519	1968	0	0	9555
2	24	187	4	41	0	0	3	32	1018	5	575	358	145	0	0	2392
3	45	158	111	76	4	70	164	89	407	37	231	210	140	0	1	1743
4	52	83	60	883	9	3	320	15	280	57	540	932	945	0	0	4179
5	294	286	46	39	62	57	26	203	1171	24	134	122	4	0	0	2468
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	15	48	74	664	9	0	167	65	794	186	1539	888	722	0	0	5171
8	16	398	70	267	5	0	226	462	2872	100	2594	497	726	0	0	8233
9	397	475	164	668	164	0	106	206	2157	115	1505	2121	1321	0	35	9434
10	217	1261	210	1263	310	190	2306	846	2136	451	1887	987	414	0	1	12479
11	2726	695	143	759	343	1	229	276	495	243	2086	3260	414	6	22	14838
12	3195	359	120	160	992	27	936	82	1055	19	609	2706	1850	0	17	12127
13	108	185	126	646	39	4	89	162	438	114	1481	1882	4678	1	31	9984
14	96	49	7	37	13	0	16	1	7	1	1	10	2	0	0	240
15	0	0	0	4	0	0	0	1	0	0	114	0	1	0	4453	4573
Tot.	9685	4273	1188	5552	2135	352	4684	2824	14479	1410	15305	14492	16470	7	4560	97416

Number of Correctly Classified Pixels = 20903
Percentage Correctly Classified = 21.46
Kappa Value = 0.12994

[1] Layered classification of only pixels with a full complement of ancillary data. Pixels on boundaries were excluded from the evaluation.

[2] See Figure 2.

Results from Classification C, Level II, Reduced Data Set with Boundaries Excluded.[1]

Reference Vegetation Groups[2]

Classified Vegetation Groups	1	2	3	4	5	6	Total
1	4411	854	4031	8572	0	1	17869
2	1466	321	2443	3409	0	0	7639
3	5406	3307	9345	12052	0	36	29146
4	9222	2660	2884	22106	7	70	36949
5	189	29	9	13	0	0	240
6	4	0	1	115	0	4453	4573
Total	20698	7171	18713	46267	7	4560	97416

Number of Correctly Classified Pixels = 40636
Percentage Correctly Classified = 41.71
Kappa Value = 0.18306

[1] Layered classification of only pixels with a full complement of ancillary data. Pixels on boundaries were excluded from the evaluation.

[2] See Figure 2.

USE OF ANCILLARY DATA IN A LANDSAT CLASSIFICATION OF A FORESTED WETLAND

By

Stephen P. Prisley

(ABSTRACT)

Digital Landsat cover-type classifications have often proved less accurate than hoped for, or have been less detailed than needed. Recent research efforts have used additional data to supplement the four bands of Landsat MSS data in an attempt to increase the accuracies of computer classifications. The goal of this study was to evaluate the use of vegetation-related ancillary variables for improving the performance of a Landsat classification of the Great Dismal Swamp.

Ancillary data considered to be related to the distribution of vegetation types in the swamp were registered with Landsat multispectral scanner data to a 50 meter UTM grid. The ancillary variables were peat depths and elevations from field surveys, and spectral texture values from the Landsat data. Discriminant analyses of a sample of pixels were performed to investigate the ability of spectral and ancillary data, separately and in combination, to discriminate between vegetation cover types.

A layered classification procedure was developed that used discriminant analysis of ancillary data after a previous unsupervised spectral classification. This was compared to a spectral stratification classification and a straightforward unsupervised classification of spectral data alone.

The layered procedure resulted in an accuracy of 21.46% for level III classes and 41.71% for level II classes. The accuracies for level III and level II classifications using the unsupervised procedure were 41.58% and 63.77%, respectively.

Some possible explanations of the seemingly contradictory results were posed, and alternative procedures suggested.