2000

Accuracy Assessment of Remote Sensing in a Tidal Wetland

Jason S. Goldberg
College of William and Mary - Virginia Institute of Marine Science

ACCURACY ASSESSMENT OF REMOTE SENSING IN A TIDAL WETLAND

A Thesis

Presented to

The Faculty of the School of Marine Science

The College of William and Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Science

---

by

Jason S. Goldberg

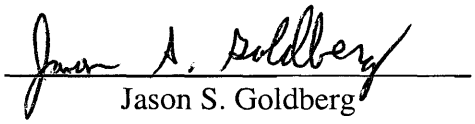
2000

APPROVAL SHEET

This thesis is submitted in partial fulfillment of

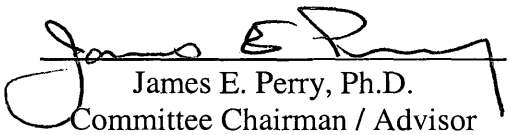
the requirements for the degree of

Master of Science

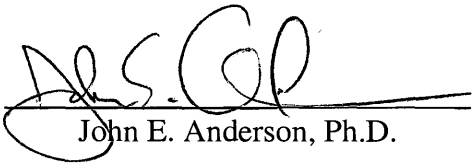


Jason S. Goldberg

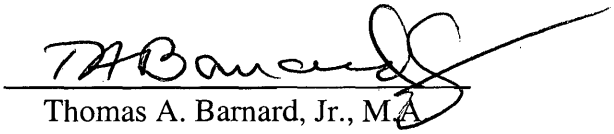
Approved, July 2000



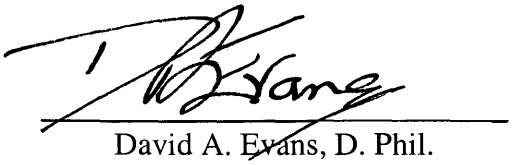
James E. Perry, Ph.D.
Committee Chairman / Advisor



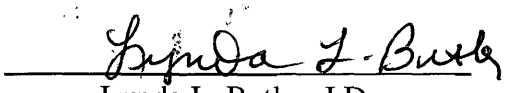
John E. Anderson, Ph.D.



Thomas A. Barnard, Jr., M.A.



David A. Evans, D. Phil.



Lynda L. Butler, J.D.
College of William and Mary
School of Law

DEDICATION

I know I promised to dedicate my first book to Dean Holden, but I am sure she will understand that this one really has to be for Brock, Amber, Mom, and Dad, without whom all of my efforts would have been possible. Throughout my research two have been far, one has been farther, and one has been even farther. Yet, somehow we managed to stay together.

G'donya and mahalo for everything.

And Carol – the *next* one, okay?

TABLE OF CONTENTS

ACKNOWLEDGMENTS

It may look as though there are only a few authors, but in reality this research required the help of many people. Some folks assisted in fieldwork, some assisted with technical work, and some assisted in helping me keep my sanity. Many sincere thanks to everyone who contributed to this final product, though, of course, any remaining errors are still solely my responsibility.

I gratefully acknowledge the Virginia Institute of Marine Science, School of Marine Science, for their Fellowship award and Assistantship as a TA for Statistics, and in particular John Milliman, Michael Newman, David Evans, and Sue Presson. Not only was my research funded, but I even had enough left over to eat without having to ever, ever, resort to Ramen Noodles. Additional thanks go to John Anderson, who paid for the DMSV imagery and plane time, and Jim Perry, who provided additional assistantship funding. Completing dual master's degrees proved quite challenging and I also would like to acknowledge John Milliman, Jim Perry, David Finifter, and Elaine McBeth for their aid in helping me meet the needs of both the M.S. and M.P.P. degrees.

Working in a tidal wetland during a heat wave proved to be an interesting challenge, yet one year later after completing fieldwork I must admit to nostalgia for Goose Creek. The absence of ticks and mosquitoes made it a great place to work, and in some ways all wetlands I shall encounter in the future just will not compare. The Virginia Department of Transportation was kind enough to allow me to conduct my research in their mitigation bank along the Goose Creek. In particular, I want to single out Steven Russell and Steven Long with VDOT for their assistance.

Tom Barnard and Walter Priest made this work possible through their own past efforts in Goose Creek. The pipes that they set out in transects made my own research feasible and saved me a great deal of work in not having to establish a similar system. I also want to thank the individuals who gave up their free time to stand in the marsh under the hot sun for several hours, making sure I didn't get lost in a *Phragmites* stand: Jill Meyer, Jim Perry, Tom Barnard, Elizabeth Mountz, Molly Mitchell, Rebecca Arenson, Helen Woods, Sylvia Cook, Thea Rittenhouse, Mary Rizza, Arminda Gensler, and Landon Ward. Patrick Richardson, as Volunteer Coordinator at VIMS, was instrumental in finding Thea, Mary, and Sylvia and deserves mention.

The folks in the Comprehensive Coastal Inventory lab in the Department of Coastal and Ocean Policy have my gratitude for all of their kind assistance and support over the years. They provided me with the technical resources (software, GPS equipment, etc.) that I needed for my research. I would like to single out Harry Berquist, Dave Weiss, Sharon Dewing, and Julie Glover for all the times they never complained when I had 'just one more question.' Harry also provided me with past data from Goose Creek that aided in the georectification of my data. As a co-conspirator and friend involved with similar research, Jill Meyer also deserves special notice here for her time and willingness to work with me in solving mutual and unique problems.

The folks in Information Technology and Networking Services also were an enormous help whenever I needed it. Steve and Kevin always had a minute to help

me solve the inevitable glitches in data storage, etc. Gary Anderson introduced me to S-Plus, which saved me many nights going crazy trying to learn SAS.

I would like to thank my committee for their efforts in helping me achieve the results before you. Their comments were very helpful in the design of this experiment and the final draft you have before you. Notably, I credit John Anderson for teaching me most of what I know about remote sensing, and Jim Perry for teaching me most of what I know about wetland ecology. All faults contained herein are entirely my own, but there wouldn't be any sound science herein, either, without the efforts of these two gentlemen. Jim deserves special recognition as a mentor, advisor, and friend over the past few years. His humor, respect, and patience were highly valued commodities to me and helped immensely through the rough spots of this research. I still don't quite know what to make of the alpacas, though Koko is still one of the sweetest llamas I have ever met. There are just some things you don't read about in the Graduate Student catalog, I suppose.

I would have never made it this far without the help of my friends, who boosted my spirits when I needed it and kept my ego in check when I deserved it. I don't have enough space to list everyone, but I should thank Aaron, Padma, Alessandra, Luke, Kevin, Mindy, Roy, Ken, Jill, Winnie, Mike, John, Kristen, Alicia, Keely, Jean, Becky, and Julie. There are many other people I know I'm forgetting here, and I apologize here in writing. If I explained how important these people, and those not listed, were, you'd be here reading until next year.

Last but not least, thanks to my family. But if you read the dedication you'd already have seen that.

Thanks again, everyone.

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF EXAMPLES

ABSTRACT

Our research explored the usefulness of Digital Multispectral Video (DMSV) as a tool for monitoring wetlands by assessing its ability to ecologically discriminate classes. The site used for this research was a ten-acre created tidal oligohaline marsh located on Goose Creek, in the City of Suffolk, Virginia. Two hundred and eighty-three sample plots were randomly selected in the wetland and sampled using a 1 m^2 quadrat. Plots were correlated using the geographic positioning system (GPS) to pixels collected by the DMSV in four bands: band 1 (770 nm - infrared), band 2 (450 nm - blue), band 3 (550 nm - green), and band 4 (680 nm - red). The data was separated into three clusters using divisive hierarchical and k-means algorithms, methods common in the 'traditional' literature for clustering data, and which formed the basis for our ecological interpretation.

The data in each cluster were divided in half. Half of the data were used for accuracy assessment, while the other half were used to test the effect of effort on classification. Classifications were tested using the kappa statistic for significance from a random classification. Classifications were also tested pairwise to assess the impact of effort and the clustering algorithm on the classification.

We were not able to answer our question conclusively due to problems with our research that highlights the need for more advanced statistics in the field of remote sensing. While remote sensing holds promise as a tool for monitoring and assessment, traditional techniques should still be used as a baseline for wetlands monitoring. We also found that it is not possible to obtain some of the most common ecological measures, such as Simpson's index, from remotely sensed data. We found that it was not possible to calculate diversity as applied to traditional techniques. However, there may be statistics that can be calculated using the imagery that can not be obtained from ground data, such as spatial diversity.

While we found no difference between unsupervised and supervised classifications, our work may suffer from Type I and Type II errors because, to our knowledge, the statistics to properly analyze our work do not exist. At best, we recommend extreme caution in using our work. We have greater trust in our ground data than in the thematic maps we produced. Further work remains to be done before we would trust remote sensing to be used as a tool, but it does hold great promise. Remote sensing has been shown to be extremely versatile in a number of other situations. We do not see anything so extraordinary about wetlands that they can not also be studied with this tool. However, the heterogeneity of wetland ecosystems and the needs of resource managers do pose unique challenges that require further statistical refinement before remote sensing can be fully utilized.

ACCURACY ASSESSMENT OF REMOTE SENSING IN A TIDAL WETLAND

INTRODUCTION

Wetlands are a vital part of our ecosystem. They exist in forms ranging from tidal saltwater marshes to prairie potholes, and can be identified by their soils, hydrology, and vegetation (Mitsch and Gosselink 1993). Wetlands act as buffers against storm erosion by absorbing wave impacts that might otherwise damage the shoreline. They can absorb water like a sponge, reducing flood damage, and can also remove harmful impurities through water filtration. They serve as a habitat for numerous species, including many fish species vital to commercial fisheries. Wetlands also provide recreation for people such as hunters, birders, and people seeking solitude (Mitsch and Gosselink 1993).

Remote sensing has been applied in recent years to purposes as wide-ranging as wetland delineations, land-use mapping, locating valuable minerals, and stress detection in vegetation (Lillesand and Kiefer 1994). It represents a potentially valuable tool for wetlands research. Collecting data for parameters such as species diversity and importance values from the ground can be a time-consuming and arduous process. Remote sensing might be able to collect the data needed to estimate these parameters more efficiently. However, its accuracy must be assessed in order to judge its effectiveness as a tool.

Many wetlands are located along or near coastlines, where they may face conflicts with encroaching development (Mitch and Gosselink 1993). Careful monitoring and development is vital if wetlands conservation is to be balanced with

the needs of the growing population. We need to be certain that the tools we use and develop are capable of meeting the demands of those needs. The purpose of this study is to consider the advantages, and disadvantages, of studying marsh vegetation communities using traditional techniques and remote sensing for wetland studies. This was accomplished by performing a land cover classification using traditional techniques of surveying and extensive site sampling and comparing that to classifications derived from remotely sensed imagery. Comparing these techniques should allow us to shed light on the usefulness of remote sensing as a tool for wetlands monitoring.

RESEARCH OBJECTIVES

This research explored the utility of using remote sensing as a tool for wetlands research by comparing it to 'traditional' ground-based techniques. Much remote sensing research has focused on the ability of remote sensing to provide thematic maps for further analysis or processing, such as their inclusion into a geographic information systems (GIS) database. Few studies, apart from its use in the study of change detection, have actually researched the issue of whether remote sensing can be used to provide information about important ecological data that can be used to describe the structure and composition of an ecosystem. While there is a great body of research on the statistical methods of testing remotely sensed imagery, there are relatively few studies that apply the mathematics to specific wetland applications. Our research tested several different hypotheses meant to explore that issue.

H_{01}: Wetland classification parameters (importance value, diversity, and evenness) derived from traditional ground-based sampling techniques are not significantly different from those derived by digital multispectral video (DMSV).

We had initially believed that it was possible to test this hypothesis with our research, but later findings showed us that this was not possible. We will describe the problem we encountered in testing this hypothesis during our discussion of our

research. However, we note and describe it in order to demonstrate its importance to the initial design of our experiment.

H_{02}: Accuracy of the Digital Multispectral Video (DMSV) classification is not affected by the number of quadrats taken during the ground truthing stage.

The amount of field effort that goes into the classification stage of analyzing remotely sensed data may impact accuracy. Supervised classification involves providing the computer with specific pixels that the analyst already knows to be of a certain class. The computer then classifies the remaining scene, leaving the analyst free from having to do such intensive delineation through fieldwork. When the computer is provided with no reference information, it produces an unsupervised classification. We predict that as the computer is provided with more pixels in a supervised classification, the accuracy of the classification improves. We also predict that supervised classifications will be more accurate than unsupervised classifications because of the additional information used to create them. We will also examine briefly the user's and producer's accuracy of the remotely sensed imagery (Congalton 1991).

H_{03}: Using only dominant species to create training signatures does not affect the accuracy of the remotely sensed image.

The entire ground-truthed dataset will be compared with a dataset consisting only of dominant species in the marsh. The data can represent a 'cloud' in n-space, where n is the number of variables measured, such as the number of species or the

number of bands. Clouds near each other may be separated in different ways depending on the multivariate analysis used to tell them apart. The amount of ground-truthed data may play a role in the results depending on how well the data describe the 'shape' of the cloud. Each band of the remotely sensed data adds to our understanding of what is happening across the marsh because of the different spectral reflectances that vegetation presents across the study site. The training data used to produce categories in a supervised classification are based upon features of interest, such as a particular species. If we are only interested in studying major trends across a wetland, such as separating the high from low marsh, nondominant species may negatively affect the classes produced through clustering by introducing unneeded variability and unnecessarily changing the shape of the clouds. To examine this hypothesis, the groundcover data were compared to a classification in which only dominant species were included.

Remotely sensed imagery may 'see' rare species with less accuracy than traditional techniques. Rare species may occur only once within a given plot. The chance of a rare species appearing in a ground plot is probably greater than the chance of it occurring in a large clump that the DMSV can observe. In order for the DMSV to observe a rare species that can be classified, it must be present in a large enough patch of vegetation to be visible to both the sensor and classification technique employed. However, a skilled observer in the field should be able to locate a rare species in a given plot if only one specimen is present. As a result, remote sensing should be better at detecting common species than rare ones.

LITERATURE REVIEW

Sampling design for ground-based research in wetlands science has been well-described in the literature (e.g. Mueller-Dombois and Ellenberg 1974). Accuracy assessment of remotely sensed data has also received much attention (e.g. Congalton 1991).

Ecological Parameters

Wetland studies, especially delineation work, focus on vegetation, soils, and hydrology (Environmental Laboratory 1987). These parameters are used to determine the current state of a plant community and may offer some predictive power on the future state of the ecosystem (Mitsch and Gosselink 1993). Quantitative field data can be collected on a variety of vegetative features. The most important in community sampling include density (number of individuals), frequency (presence/absence of a species), and cover (surface area of a species that 'covers' the ground) (Mueller-Dombois and Ellenberg 1974). Measures used to describe parameters, such as species diversity, can be calculated from the same data.

Density is a measure of abundance that describes how many individuals exist in a given area. Frequency refers to whether a species is found within a given area. Cover refers to how much surface area on the ground a given species shades from sunlight (Mueller-Dombois and Ellenberg 1974). These measures can be translated

into relative values. The sum of these relative values is known as the importance value and are defined by Mueller-Dombois and Ellenberg (1974) as:

$$\textbf{Relative density} = \frac{\textbf{number of individuals of species}}{\textbf{total number of individuals}} * 100$$

$$\textbf{Relative frequency} = \frac{\textbf{frequency of a species}}{\textbf{sum of the total frequency from all species}} * 100$$

$$\textbf{Relative dominance} = \frac{\textbf{dominance of a species (species cover)}}{\textbf{dominance (cover) of all species}} * 100$$

Importance value of a species = relative density + relative frequency + relative dominance

Cover (relative dominance) is generally regarded as being more useful than frequency because it provides a better estimate of biomass (Rice 1967, Daubenmire 1968). Absolute measurements, measurements not made in proportion to other species, may also be used to reveal information about an ecosystem. For example, a particular species may have an absolute cover of 10%, yet if no other species are present, it will have a relative dominance of 100%. Thus, both measures should be used in deriving information about a community.

The two most commonly used diversity indices are the Shannon and Simpson indices (Magurran 1988). The Shannon index, H', a measure of species richness, measures the degree of uncertainty of being able to predict the species in a community that were picked at random. It ranges from 0 (a single species – completely certain of what species will be picked) to very high values, and is defined by Zar (1984) as:

$$H' = \frac{n \log n - \sum_{i=1}^{k} f_i \log f_i}{n},$$

where
k = number of categories (species),
n = sample size, and
f_i = number of observations in category i.

Species evenness can be derived from H' (Magurran 1988):

$$J' = \frac{H'}{H_{max}},$$

where
J' = evenness,
H' = Shannon index, and
H_{max} = maximum possible value of Shannon index (all observations divided into identical proportions in each category) which is calculated as log k.

The Simpson index, SI', is based on the probability of drawing a pair of individuals of the same species at random from a community. It represents a dominance index, is weighted towards the most abundant species in a community, and is defined as (Magurran 1988):

$$SI' = \sum_{i=1}^{s} [n_i(n_i-1)] / [N(N-1)]$$

where
n_i = number of individuals per species and
N = total sample size.

Quantitative data on plant species and their numbers can be very extensive. Mathematical formulas, such as those used to calculate species cover and the Simpson index, provide a way to break the data down into a more interpretable form. Perry and Hershner (1999), for example, looked at species richness, evenness, and

diversity in tidal freshwater marsh vegetation. Their results showed that, over time, all measures except the importance values did not change. However, the species comprising those measures did change, suggesting that more salt tolerant species were taking over the marsh. They concluded that an oligohaline ecosystem was developing in the marsh. Aerial imagery from years or decades prior could have been used to study temporal changes occurring in the same marsh and, possibly, may have added extra support for their conclusions.

'Traditional' Wetland Sampling Techniques

Considerations in Wetland Sampling

Mueller-Dombois and Ellenberg (1974) discuss several features they regard as essential to selecting a proper sampling technique that can be used to measure the previously described measures. If a scientist is studying gradients, the ecosystem might be composed of different vegetative classes. A scientist could choose from a variety of classes to study, ranging from the broad to the specific, affecting the accuracy of a study utilizing remote sensing. The accuracy of remotely sensed data may improve as species are clumped together into broad classes (Rosenfield 1986). That is, if broad land cover classes are used for classification instead of classifying the map to the species level, then the map should be more accurate. The decision of how to clump different categories should be based on a goal defined by an experiment's objective. If it is not necessary to study the difference between two or more plant species, and it is scientifically valid to clump them together, doing so could improve the accuracy of the study.

Remote Sensing Techniques

Classification

The digital image collected by the DMSV is organized in the form of a matrix, where each cell of the matrix contains the brightness value of a particular pixel for a given band. Lillesand and Kiefer (1994) define a pixel as a 'discrete picture element.' When one views a digital image, one is actually looking at a large number of dots, or pixels, where each pixel represents a defined area on the ground. The amount of data depends on the number of bands and the type of sensor used. For a sensor such as the DMSV, each pixel in an image has four data associated with it; one datum for each of the four cameras on the DMSV. This results in four matrices, where the same cell in each matrix represents a different band for the same pixel.

The data can be used to classify each pixel. Classification is a process where the data is used to place a pixel into a category, or class. For example, each dot in Example 1 represents a combination of two brightness values (i.e. the blue and green wavelengths). If three brightness values had been used to characterize each pixel, the graph would be three-dimensional, with pixels appearing in 'clouds.' The four brightness values collected by the DMSV for each pixel can be represented in 4-dimensional space. The value of each pixel is determined by the reflectance characteristics of a particular object. It is assumed that the same type of object will have approximately the same reflectance values across a scene. For example, the eight pixels represented in the right side of Example 1 might be assumed to belong to the same object, or class. The number of classes depends on the type of classification used and the objective of the analyst. If we were viewing data from three bands, we

might see several clouds of varying shapes and proximity to each other, much as Example 1 shows two-dimensional data.

Two types of classification are used in remote sensing. Unsupervised classification is performed with minimum analyst interference. The computer seeks out 'natural' groupings of pixels to separate into different classes. The analyst chooses the number of classes, and the computer mathematically separates out that number of categories from the data. Pixels that are closer together are assigned to the same class. There are a variety of algorithms that the computer can use to separate out classes, which can lead to different classifications (Lillesand and Kiefer 1994). For example, different algorithms might yield a number of different categories in Example 1. Note that the graph only ranges in brightness value from zero to ten. If the full range of the DMSV had been used, these values could have been spread from 0 to 255. Pixels appearing in the range shown here could easily be interpreted as normal variation around one class, or just as easily deemed separable into multiple categories, depending on the algorithm used.

Supervised classification is more analyst-intensive. The analyst selects 'training pixels' of a known class on an image based on their knowledge of the area in question. For example, they might select pixels belonging to trees that they know to be of a certain species. The computer derives the spectral attributes of a class based on the training areas that the analyst has chosen. The training pixels are used to identify the rest of the scene. Each pixel in the scene is placed into a class based on its similarity to the training pixels the analyst selected (Lillesand and Kiefer 1994).

Supervised classification may allow for more precise results, as the computer has an additional guide to classification other than just the 'raw' spectral data.

The mathematical formula used may affect the classification. The formulas can change how different pixels are organized into different classes because there are a variety of means available from which to cluster samples together. Image processing packages, such as ERDAS Imagine, allow the analyst to select different mathematical techniques. The different algorithms may yield different classifications. Such software packages allow the training data to include either parametric or non-parametric signatures. A signature contains the training information the computer uses in assigning pixels to a given class. Non-parametric signatures are based on actual discrete objects drawn by the analyst on the original image, and define the boundaries of a given class. Parametric signatures, instead of using actual locations to classify pixels, are based on statistics, such as the mean, that describe the training data (Smith *et al.* 1994). The values of unknown pixels are compared to the parametric signatures in order to decide which class they should be assigned to. Once parametric signatures have been collected, the computer has a choice of options of how to assign a particular pixel to a certain class. Mathematical functions available in ERDAS include the minimum distance, Mahalanobis distance, and maximum likelihood decision rules.

The minimum distance decision rule "calculates the spectral distance between the measurement vector for the candidate pixel and the mean vector for each signature" (Smith *et al.* 1994). It is similar to Euclidean distance in that a given pixel is assigned to the particular category with which it has the closest similarity of

spectral values. It is calculated as follows (from Smith *et al.* 1994, referring to Swain and Davis 1978):

$SD_{xyc} = \sqrt{(\sum^{n}_{i=1} (\mu_{ci} - X_{xyi})^2)}$,

where
n = number of bands,
i = a particular band,
c = a particular class,
X_{xyi} = data file value of pixel x,y in band I,
μ_{ci} = mean of data file values in band i for the sample for class c, and
SD_{xyc} = spectral distance from pixel x,y to the mean of class c.

SD_{xyc} is computed for all available classes, as determined by the analyst. A given pixel is assigned to whichever class provided the lowest SD_{xyc}. Minimum distance assigns every pixel in the area of interest to one, and only one, class, with no unclassified pixels emerging from the classification (Smith *et al.* 1994). The likelihood of the spectral distance being equal for an unknown pixel to two or more categories is highly improbable. However, it is possible that a pixel may exist in a scene that does not truly belong to any of the available classes. The minimum distance decision rule will always result in a pixel being assigned to one of the classes identified by the analyst. As a result, it is extremely important that the classes incorporate all available possibilities that might occur in a scene. Otherwise, the analyst should perform more complex analyses in order to improve the accuracy of the supervised classification. A second problem with using the minimum distance decision rule is that it does not account for the variability of the individual classes. Variance of the different classes may not be the same. If a pixel is, therefore, an outlier of a given class, it may actually be closer to the mean spectral value for another class that has a tighter variance (Smith *et al.* 1994).

The Mahalanobis distance decision rule uses a covariance matrix, rather than the spectral distance, to assign pixels to classes. Using the covariance matrix reduces the problem that variance can cause in the minimum distance decision rule. It is calculated as (Smith *et al.* 1994):

$$D = (X-M_c)^T (Cov_c^{-1}) (X-M_c)$$

where
D = Mahalanobis distance,
c = a particular class,
X = the measurement vector of the candidate pixel,
M_c = the mean vector of the signature of class c,
Cov_c = the covariance matrix of the pixels in the signature of class c,
Cov_c^{-1} = inverse of Cov_c, and
T = transposition function.

Analogous to the minimum distance decision rule, a pixel is assigned to the class for which D is smallest. While using Mahalanobis distance as a decision rule is somewhat more accurate than using the minimum distance rule, since it takes the variability of the classes into account, it may still suffer from large inaccuracies if the signatures have large values in the covariance matrix. Another potential problem with Mahalanobis distance is that it is a true parametric rule. Data collected for each band must have a normal distribution (Smith *et al.* 1994).

Like the Mahalanobis rule, maximum likelihood also assumes that the bands follow a normal distribution. In addition, it assumes that the probability of a pixel belonging to a particular class is the same as for any other class. It is calculated (Smith *et al.* 1994) as follows:

$$D = \ln(a_c) - [0.5 \ln (|Cov_c|)] - [0.5(X-M_c)^T(Cov_c^{-1})(X-M_c)]$$

where

D = weighted distance (likelihood),
c = a particular class,
X = the measurement vector of the candidate pixel,
M_c = the mean vector of the sample of class c,
a_c = percent probability that any candidate pixel is a member of class c (1, or based upon *a priori* knowledge),
Cov_c = the covariance matrix of the pixels in the sample of class c,
$|Cov_c|$ = determinant of Cov_c,
Cov_c^{-1} = inverse of Cov_c, and
T = transposition function

As before, D is calculated for each band and the pixel is assigned to the class for which D is the smallest. The maximum distance classifier is regarded as the most accurate available to ERDAS since it takes the most variables into account (Smith *et al.* 1994). By using the covariance matrix, it avoids the variance problem that affects the minimum distance decision rule. However, it suffers the same problem that affects the Mahalanobis decision rule in that it tends to overclassify categories (attribute more pixels to a given class than what exists in reality) that have large values in the covariance matrix (Smith *et al.* 1994).

Unsupervised classification is often useful for preliminary work, but supervised classification should be used to ensure more accurate results. All remotely sensed imagery needs to be verified with field data. For example, the computer may produce a thematic map containing five classes, when the analyst, who has visited the scene and knows what to expect, may see only two. Conversely, the computer may accidentally combine classes together into one class, or misclassify pixels of one class into another. The issue of accuracy in remote sensing is extremely important, as users of the data must have some measure of confidence in their data.

Accuracy of Remote Sensing

Discrete Nature of Remotely Sensed Data

Data collected through remote sensing is discrete, not continuous. A sensor can not record a brightness value of 142.29 for a given pixel; it must record 142 or 143. The sensor's ability to discriminate shades, or levels of brightness, determines its dynamic range. The dynamic range of the DMSV is 0-255. Each pixel, depending on its brightness, is assigned a value of 0, 1, 2, 3, 4, 5...255, where 0 means that the DMSV recorded no light, and 255 means that the sensor was saturated. The discrete nature of the data is important as only discrete multivariate methods must be used to analyze the data. Statistical tests such as ANOVA, which assume a continuous distribution, should not be used.

Description of the Error Matrix

A tool commonly used to assess the accuracy of remotely sensed data is the error matrix (Congalton 1991). An error matrix is a table that shows the number of pixels that have been correctly identified within a scene and allows an estimate to be made of the effectiveness of remote sensing in classifying an image. It is derived from a sample of pixels rather than from an evaluation of every single pixel in an image. Therefore, there would be 235 pixels in Example 2. Example 2 provides a case where an imaginary scene consisting of water (W), roads (R), trees (T), and shrubs (S).

The columns refer to the reference categories – what the pixels are as observed through ground-truthing. Of the 235 pixels that were ground-truthed, 72 are

water and 112 are trees. The rows refer to the classified categories from the sampled pixels from the classified image. For example, from the 235 sample pixels, 55 appear as water and 125 appear as trees. From Example 2, one can determine which categories were correctly or mistakenly classified. For Example 2, 50 pixels were correctly classified as water, five pixels that were classified as water are actually trees, and twenty pixels that should have been classified as water were misclassified as trees. Seventy-two of the 235 total pixels in the scene are truly water; however, only 55 pixels were classified as such.

Producer's and User's Accuracy

The most commonly used measure of accuracy in remote sensing is total accuracy (Congalton 1991). It takes into account only the major diagonal of the error matrix. It is the number of pixels that are correctly identified divided by the total number of pixels sampled in the error matrix. In example 2, the total accuracy is (50+25+100+20)/235 = 82.98%. Total accuracy is regarded as a poor measure of accuracy (Congalton and Green 1999). A user may believe that the 82.98% accuracy applies to all categories equally, which may not be the case. This discrepancy has led to the development of errors of omission (producer's accuracy) and commission (user's accuracy) as additional measures of accuracy (Congalton 1991).

Producer's accuracy measures how many pixels are left out of a correct classification (Congalton 1991). Producer's accuracy measures how many pixels that truly should have been identified as one class were mislabeled as another. It is calculated by dividing the correct number of pixels by the column total. Example 2 contains 72 pixels that represent water (note that only 55 pixels are actually seen as

water on the 'real' image). Out of those 72 pixels, 22 were misclassified and 50 were classified correctly. The producer's accuracy is thus calculated as (50/72) = 69.4%. The producer's accuracy for shrubs is (20/20) = 100%. From the producer's perspective, every pixel that truly was a shrub was identified as a shrub.

The error of commission is found by dividing the correct number of pixels in a category by the row total (Congalton 1991). In Example 2, water has a user's accuracy of (50/55)=90.91% while the user's accuracy for shrubs is (20/28) = 71.43%. The producer is only interested in knowing how many pixels *are not* being included in a given class, i.e. pixels that are omitted. A user, someone working from the map that a producer provides, is more interested in how many pixels *are* included in a given class.

The calculated values for errors of omission and commission are very different for water and shrubs in Example 2. While a producer is 100% certain that a shrub on the ground will be included in the 'shrub' classification, someone using the map can only be 71.43% certain that what they are looking at is, in fact, a shrub. A tree may be misclassified as a shrub. Conversely, while a producer may not be able to capture every pixel of water correctly, the category 'water' is not very likely to include pixels from other categories. Due to the differences between producer's and user's accuracy, Story and Congalton (1986) recommended reporting both measures.

Kappa Statistic

Another method that has been proposed to measure accuracy is the KHAT statistic. It is meant to determine whether classifications improve the accuracy of the

data, or if the results are no different from a random classification. Lillesand and Kiefer (1994) describe the KHAT (k') statistic, which calculates the probability that a computed classification is better than a random assignment of classes.

Conceptually,

$$k' = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}}$$

As the chance that the matrix could have been randomly created drops to zero, k' approaches 1. It decreases as the probability that the classifications could have been randomly assigned increases. The upper limit of this statistic is 1, but there is no lower limit (Rosenfield and Fitzpatrick-Lins 1986). Negative values of k' are possible if the classification is very poor, but the magnitude of negative values is less important than the sign (Lillesand and Kiefer 1994). Kappa can be tested for significance by estimating its variance and using a Z test. A classification can be tested to see if it is significantly different from random, or different classifications can be compared to each other (Lillesand and Kiefer 1994, Congalton 1991).

Mathematically, k' is calculated as

$$k' = \frac{N * \sum^{r}_{i=1} x_{ii} - \sum^{r}_{i=1} (x_{i+} * x_{+i})}{N^2 - \sum^{r}_{i=1} (x_{i+} * x_{+i})}$$

where
r = number of rows in the error matrix,
x_{ii} = number of observations in row i and column i (on the major diagonal),
x_{i+} = total of observations in row i,
x_{+i} = total of observations in column i, and
N = total number of observations included in the matrix (Lillesand and Kiefer 1994).

For example, using Example 2,

$\sum x_{ii} = (50+25+100+20) = 195$
$\sum(x_{i+} * x_{+i}) = (55*72) + (27*31) + (125*112) + (28*20) = 19357$

$$k' = \frac{(235*195) - 19357}{235^2 - 19357} = 0.7379, \text{ or } 73.79\% \text{ better than random chance.}$$

While k' applies to the entire error matrix, Rosenfield and Fitzpatrick-Lins (1986) quote Bishop *et al.* (1975), who describes a conditional kappa for an individual category instead of the entire error matrix. They state that k' and conditional k' are the most statistically sound measures of accuracy. They base this statement on a comparison of other measures (e.g. Short 1982 and Turk 1979). Conditional Kappa is defined in Rosenfield and Fitzpatrick-Lins (1986) as:

$$\mathbf{K_i = \frac{p_{ii} - p_{i+}p_{+i}}{p_{i+} - p_{i+}p_{+i}}}$$

where
p_{ii} = proportion of pixels in a given cell,
p_{i+} = proportion of pixels in a given row, and
p_{+i} = proportion of pixels in a given column

Calculating Significance of the Kappa Statistic

A Kappa statistic can be tested to see whether it is significantly different from a randomly produced classification, and whether it is significantly different from another error matrix (Congalton and Green 1999). Congalton and Green (1999) report that the Delta method can be used to calculate the approximate large sample variance:

var. $k' = [\theta_1(1-\theta_1)/(1-\theta_2)^2 + 2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)/(1-\theta_2)^3 + (1-\theta_1)^2(\theta_4-4\theta_2^2)/(1-\theta_2)^4] / n$,

where
$\theta_1 = 1/n \sum^1_{i=1} n_{ii}$,
$\theta_2 = 1/n^2 \sum^1_{i=1} (n_{i+} * n_{+i})$,
$\theta_3 = 1/n^3 \sum^1_{i=1} n_{ii}(n_{i+} + n_{+i})$,
$\theta_4 = 1/n^3 \sum^1_{i=1} \sum^1_{j=1} n_{ij} (n_{j+} + n_{+i})^2$;

and

n_{ii} = cell in row i and column i,
n_{i+} = sum of all cells in row i,
n_{+i} = sum of all cells in column i, and
n_{j+} = sum of all cells in row j.

The estimate of Kappa can then be tested for significance by calculating a Z-score as follows: $Z = k'/\sqrt{(\text{var. } k')}$. The test assumes a two-tailed Z test and infinite degrees of freedom. Two Kappa statistics can be compared to one another using a similar Z-test: $Z = |k_1' - k_2'| / \sqrt{(\text{var. } k_1' + \text{var. } k_2')}$, where k_1' and k_2' represent the desired Kappa scores one wishes to test.

Other Accuracy Measures

Kappa and conditional Kappa are only two of several measures that have been proposed for assessing the accuracy of remotely sensed data. Aronoff (1985) developed two statistics similar to user's and producer's accuracy. Together they form the *minimum accuracy value*, which is "the lowest expected accuracy of a thematic map given an observed accuracy test result and the user selected consumer risk." It is used to determine whether a map is accurate enough to meet an individual's needs. Consumer risk measures the probability that "a map of unacceptable accuracy will pass the accuracy test" (Aranoff 1985). It is different

from k' is used to reveal whether the classification is significantly different from random. The minimum accuracy value test provides more of a confidence interval than a test for significance by providing a measure of the uncertainty surrounding an accuracy value (for example, saying that an individual is 95% certain that a classification is 80% accurate). Aranoff's measures, which date back to the early stages of remote sensing accuracy analysis, do not seem to have been adopted by the remote sensing community since they were proposed, based on a literature search by the authors.

Another measure of accuracy is the τ statistic (Ma and Redmond 1995). The tau statistic resembles the Kappa statistic:

$$\tau = \frac{P_c - \sum^{q}_{k=1} \beta_k p_{+k}}{1 - \sum^{q}_{k=1} \beta_k p_{+k}}$$

where
P_c = overall accuracy,
β_k = the *a priori* probability, set by the user, that a given pixel belongs in class *k*, and
q = total number of classes.

Kappa uses proportions derived from the map to assess accuracy, while tau uses probabilities set by the analyst. They both attempt to determine whether a classified scene is different from one which was randomly produced.

Stehman (1997) analyzed the usefulness of the kappa and tau statistics, and stated that they are not good measures because a user is only interested in knowing whether a given pixel is correctly identified. A randomly classified pixel classified correctly is in the same class as if the computer had performed an accurate classification based on a mathematical technique. "If the overall map accuracy is

80%, the user holds a map for which a randomly selected area has an 80% chance of being correctly classified" (Stehman 1997). However, he ignores the value that kappa provides to the producer of the map. An analyst is striving to produce the best product possible. A measure such as kappa, which reveals how much better the image is over a random classification, can serve as a guide in evaluating classification techniques. For the user however, Stehman has a point when he states that overall, user's, and producer's accuracy "are more relevant accuracy parameters because of their direct interpretation as probabilities characterizing data quality[.]"

Sampling and Analysis Effects on Accuracy

Several studies have considered the effects of sampling collection and different data analyses on accuracy. Ginnevan (1979) suggested that a sampling procedure should have a low probability of generating a poor (low accuracy) map, a high probability of generating a good (high accuracy) map, and require some minimum number of ground samples.

Sampling Scheme

Hay (1979) recommended that at least 50 samples should be collected for each category for which accuracy information is desired. Simple random sampling may be inefficient in areas where some categories have small coverages, because pixels are randomly chosen until all categories have at least 50 pixels sampled. Areas with large coverages will probably have many more than the 50 pixels needed to efficiently test the accuracy of that category. Hay (1979) recommends two alternatives to this method. The overall data is stratified into known (or suspected)

categories and 50 pixels are selected from each class, or random samples are drawn until one category has 50 pixels, when the overall sample is treated as complete. Sampling continues, but selected samples that fall into completed categories are rejected. The proportions of the initial sample may not necessarily match the proportions of the ground coverage. It is likely that Hay intended the initial set of data (where pixels are selected until one category has 50 pixels) to be used in computing the overall accuracy, while the subsampling allows categorical accuracy to be determined. Hay (1979) used these separate error matrices to determine whether categorical accuracy was over- or under-estimated.

Hay (1979) also used this sampling scheme to determine whether errors found within a category were randomly distributed or not, using the binomial and Poisson distributions. The binomial distribution is used under the assumption that errors are distributed evenly among all cells. With this assumption, the probability that a given cell contained y errors could be calculated. The likelihood that y errors is within the realm of random chance can be approximated using the Poisson distribution. The mean number of errors in a given category is then calculated. Hay then consulted Poisson tables for the distribution given this mean. The distribution gives the probability that a given cell will contain a certain number of errors. The likelihood that a cell will contain y *or more* errors can be calculated from the Poisson distribution. A subjective measure of probability can be used to evaluate whether the number of errors in a certain cell is random or not. This can be used to determine whether a misclassification between categories is random or if one category is repetitively classified as another.

Congalton (1988) calculated error matrices based on simple random sampling, stratified random sampling, cluster sampling, systematic sampling, and stratified systematic unaligned sampling for three sets of data of varying complexity. He reported that simple and stratified random sampling are the best procedures for accuracy assessment. Stratified random sampling is important when the user is interested in small areal coverages, such as diversity studies of heterogeneous areas. Congalton (1988) recommended that stratified systematic unaligned sampling and systematic sampling should not be used for error analysis because of spatial autocorrelation effects. Congalton (1991) recommends stratified random sampling in areas where a minimum number of samples are collected. He notes that the Kappa statistic assumes a multinomial sampling model, which only simple random sampling satisfies completely. The effect of using a different sampling effort on the Kappa statistic is unknown (Congalton 1991).

Sample Size

Early research focused on the use of the binomial distribution to compute the pixel size required to obtain a desired accuracy (i.e. Hay 1979). For purposes of time and efficiency, one may sample only a few pixels. However, it is possible for all sampled pixels to be accurate when the accuracy of the entire image is not 100% (Hay 1979). For example, if a given map has an accuracy of 85%, and 10 pixels are sampled, the probability that 9 or 10 pixels will be 100% accurate is 55%. Hay (1979) provided equations for determining the appropriate sample size without giving a formal explanation for their use. He reported that the overall sample size required should be calculated as

$$N \approx \frac{100 * n}{S_{max}}$$

where
N = overall sample size,
n = minimum sample size for a category, and
S_{max} = the percent of land in the largest category.

Hay (1979) reported that S_{max} is usually inversely related to the number of categories. As the number of categories rise, so will N, which might lead to the overall accuracy being greater than the accuracy of individual categories (Hay 1979). Small errors within each category would be masked by the overall dataset in calculating the overall accuracy.

Hay (1979) reported that some authors had attempted to use the standard error equation for binomial data to calculate a 95% confidence interval:

$$SE\% = \sqrt{((p\% * q\%) / n)}$$

where
SE = standard error,
p = percentage correct,
q = 1 – p, and
n = sample size.

The equation can be used as long as the assumption that the errors are distributed normally is not violated, and that p and q are large. Hay (1979) does not explain what happens when p is large, since this would imply by definition that q is small. Based on this equation, and without explanation, he states that the sample size for each category must be greater than 50 if this equation is to be used with any certainty.

Rosenfield *et al.* (1982) used the cumulative binomial distribution and a preliminary estimate of the accuracy of the map to derive the number of sample points required to achieve a desired accuracy with a 95% confidence interval. As the true accuracy of the map increases, the number of points needed to verify that accuracy declines (Rosenfield *et al.* 1982). Rosenfield *et al.* (1982) also provide a description of a hypothesis test that allows one to test whether, given a specified number of sample pixels, a category's accuracy meets or exceeds some desired accuracy level. Their hypothesis test is performed for a desired accuracy level of 85% ($H_o : p \geq p_o = 0.85$). They also provide a list of critical values, i.e. points that must be correctly classified given a specified number of sample points, for a second hypothesis test ($H_o : p < 0.85$). Their discussion is not developed enough to allow the reader to determine the critical values for other potentially desirable accuracy, such as 0.95.

Rosenfield *et al.* (1982) regard the hypothesis test for category accuracy as different from overall accuracy. Testing for overall accuracy is more than simply combining hypothesis tests for different categories because the boundary of a classification is important to the overall task of classifying a pixel into one class or another. They state that stratified systematic unaligned sampling is an effective procedure to use for determining the total accuracy (as defined earlier) of the entire map because it is area-weighted. In a sense, total accuracy is area-weighted because it can be represented not only by **(the total number of pixels correctly classified / total number of pixels in the image)**, but also by **(the total area correctly classified / total area in the image)**. Sampling pixels (points) is simply an allowable

approximation to sampling areas. They describe mathematical procedures for computing the variance and mean of this overall accuracy measure. It was not clear to us how the procedures they describe could be used to determine the number of pixels needed to sample an area, or how to test those pixels for accuracy after they had been collected.

Congalton (1991) found several methods for determining the required sample size to be flawed because they did not consider the confusion in an error matrix. Misclassified pixels fall into the off-diagonals of an error matrix. Collecting too few pixels may result in many cells of an error matrix with values of 0, which may not itself be accurate of the entire classification. Rosenfield *et al.* (1982) did not consider this possibility. Congalton and Green (1999) recommend that 50 pixels be ground-truthed for every class present in a thematic map, an approach adopted by other authors (such as Lillesand and Kiefer 1994). Congalton and Green (1999) note that if the region or number of categories being studied are especially large, then 75-100 samples per category should be collected. He also notes that the sampling number can be rearranged depending on the needs of the user and the variability of the data. Fewer pixels may need to be ground- truthed for less variable regions.

Congalton (1988) recommended that 1% of the total pixels in a scene be sampled to develop an error matrix. One would assume that, given a choice of sampling 50 pixels per category or 1% of the total pixels, the larger of the two values would be preferable for accuracy analysis. Dicks and Lo (1990, quoting Fitzpatrick-Lins 1981) recommended that the following equation be used to determine the number of pixels needed:

$$N=(Z^2*p*q) / (E^2)$$

where

N = the sample size,
p = the expected percent accuracy,
q = 100-p,
E = the allowable error, and
Z = 1.96 ("the standard normal deviate for the 95 percent two-sided confidence interval").

However, this equation would appear to be questionable because it does not take into account the number of pixels in the image under study. As the image size increased, it seems that more pixels would need to be collected to generate the same given accuracy. Fitzpatrick-Lins (1981, quoting Snedecor and Cochran 1967), report the 95% lower confidence limit of the true accuracy value as:

$$p_L = p' - \{1.645 * \sqrt{(p'q'/n) + (50/n)}\},$$

where
p_L = the lower limit of the accuracy (expressed as a percent),
p' = total accuracy as defined previously,
q' = 100 – p', and
n = sample size.

Fitzpatrick-Lins (1981) did not report the upper boundary surrounding the confidence limit as she did not consider errors of omission. It is not made clear why such errors would not affect the lower limit if they had affected the upper limit.

The size of the sample used will also affect the power of the statistical test. Stehman (1997) discussed other studies where extremely large sample sizes (n=62727) resulted in finding kappa values as low as 0.077 significant. Conversely, small sample sizes may not be enough to detect whether kappa is significant, even when it is large. It should also be noted that authors who have expounded on the

notion of at least 50 pixels per classified category have violated this minimum requirement on occasion. Congalton and Green (1993), for example, provided accuracy measurements for three categories (big old growth, little old growth, and other) in an old growth forest using only 40 pixels instead of at least 150 as Congalton (1991) might have suggested. They did note the small sample size and its statistical effect on the error matrix, and suggested that collecting the recommended sample size would have defeated the purpose for initially using remotely sensed imagery, that of saving time and effort. Congalton and Green (1999), based on studies such as these, recommended as a 'rule-of-thumb' that fifty pixels be acquired for each desired class in testing an error matrix.

Classification Scheme

Several studies have addressed the issue of whether different classification schemes, or different analysts classifying an image, might affect the final result. Congalton (1991) compared supervised versus unsupervised classifications using the Kappa statistic and found no difference. However, a modified approach that combined elements of supervised and unsupervised classifications resulted in a higher accuracy.

Stehman (1997) provided other examples. One such paper, Congalton *et al.* (1983) calculated overall accuracy and the kappa statistic for four different classification algorithms. In each case, the z-statistic computed to compare algorithms did not change. However, the study by Congalton *et al.* (1983) did compare algorithms pairwise. While not affecting the conclusion of whether a

particular algorithm was significant, it could have led to significance being found between algorithms, much for the same reasons that an ANOVA is preferred to multiple t-tests. The study found that, if one ranked each algorithm by its accuracy, there was a slight difference depending on whether overall accuracy or kappa was used. Another report quoted by Stehman (1997), Jakubauskas *et al.* (1992), reported the same results. As in the paper by Congalton *et al.* (1983), the order of most to least accurate changed slightly, but only occurred when the differences between classification algorithms were initially slight. Stehman (1997) did report that differences between overall accuracy and kappa were possible between images if the scenes being classified had very different land-cover structure. However, as he pointed out, "[W]hat objective motivates a comparison of accuracy for...very different regions?"

METHODS

Site Description

The site used for this research is a ten-acre created tidal oligohaline marsh located on Goose Creek, in the City of Suffolk,Virginia (Figure 1). The predominant vegetation on the site is comprised of *Phragmites australis*, *Amaranthus cannibina, Scirpus robustus*, *Spartina alterniflora*, *Spartina cynosuroides, Aster* spp., and *Typha spp.*, though heterogeneous patches of other vegetation also exist on the site. The property is divided approximately '50-50' into high and low marsh (Barnard, pers. comm.).

Model for Determining the Experimental Procedure from a Remote Sensing Perspective

I. Identifying the object of interest

The sensors we used were selected based on our knowledge of the area, as some sensors work better in some environments than others. The part of the electromagnetic spectrum (EMS) to be studied will also be determined by the environment in question. For example, the 400-2500 nm range of the EMS includes most of the incident radiation of the solar spectrum, and is therefore most widely used for remote sensing of vegetation (Carter 1993). Other characteristics, such as fog or water turbidity, may also restrict the sensors available for use (Lillesand and Kiefer 1994).

II. Considering the scientific goal

Different sensors provide different accuracy and precision. Accuracy refers to the ability of the sensor to detect the true value of the reflectance or other parameter being measured, while precision refers to the ability of the sensor to obtain the same value over repeated samples of the same area (Anderson, pers. comm.). An individual interested in a preliminary survey for coastline changes due to erosion or development may be able to accomplish their goals using aerial photographs (Hardisky *et al.* 1986, Environmental Laboratory 1987). Someone interested in detecting stress over a wide area might want to consider using a tool such as the DMSV, which covers four bands in the EMS. To determine the effects of stress on a previously unstudied species, one might consider using the spectroradiometer, which collects data from a much larger portion of the EMS than the DMSV. The spectroradiometer collects data from a fixed point, essentially one pixel, across a broad range of the EMS. The DMSV, on the other hand, collects data representing many pixels from only a few selected points in the EMS.

III. Acquisition of data

There are different ways to acquire data. The energy required for detection, whether to use a space or aerial platform, and the sensor orientation are all issues that need to be addressed when carrying out a study. Digital imagery, for example, often requires more time to process than film. Film, while it may be more sensitive to particular wavelengths, may not reflect the exact mix of wavelengths present at a site. Film emulsion, age, exposure, and processing technique also need to be considered

when examining photographic data (Murtha 1992). Spaceborne platforms are typically used to acquire data from larger areas than aerial platforms (Lillesand and Kiefer 1994).

There are many types of digital sensors. Examples include Landsat, Thermal Infrared Multispectral Sensor (TIMS), SPOT, and the DMSV (Lillesand and Kiefer 1994). The DMSV is comprised of four cameras, each of which can record data from a desired portion of the EMS. The DMSV is sensitive from 350-950 nm (Anderson, pers. comm.). The entire desired range of vegetative study in the EMS is 400-2500 nm (Lillesand and Kiefer 1994). Much of the higher part of the range is used to study vegetative stress. For example, the EMS from 1300-2500 nm can be used to study water loss from plant leaves (Carter 1991). Data from the blue, green, red, and near infrared wavelengths were considered sufficient for our study. While the spectroradiometer can collect more data across the EMS than the DMSV, it is very time-consuming. The information from four bands that the DMSV examines is sufficient for many studies. At present, sensors are not capable of collecting data from across the entire range of the EMS for all the pixels in a scene because of the time it takes to measure spectral data and the computer memory required. The spectroradiometer works well when data is only needed for a few points. Sensors such as the DMSV are more desirable when information needs to be gathered from a wide area. (Anderson 1997).

IV. Classification and error analysis of digital data

There are several techniques that can be used to classify and assess the accuracy of digital data. Someone interested in just the overall picture of a thematic map might use an overall accuracy technique, such as taking the number of correctly identified pixels in a scene, and dividing by the total number of pixels in a scene (Congalton 1991). Someone interested in categorical data might need to consider errors of omission or commission, errors that describe changes in a specific category rather than the overall scene (Congalton 1991).

Collection of Ground Data

Field data were collected using a\ simple random design. The target goal was to collect fifty samples for every class to be analyzed. All sample points were located in the marsh to precisely correspond with remotely sensed imagery using transects placed by T. Barnard and W. Priest. Five parallel transects were located within the marsh. Each transect was 30 meters apart and laid out along an angle of approximately 42 degrees. Pipes were placed at ten meter intervals along each transect (Figure 2). Each pipe marked the center of a rectangle five by fifteen meters long. The goal of sampling was to collect 500 data plots, which were enough to analyze five classes. Half of the points were to be used for error matrix calculations, while the other half were to be used for supervised classification. Microsoft Excel was used to generate the random points. Pipes were randomly assigned a number and randomly selected. Next, a number between zero and five, and another number between zero and fifteen were selected. Positive and negative (0 or 1) numbers were

generated for the latter two to indicate direction away from the selected pipe. The direction of plots from the pipe were converted to a compass heading and distance using trigonometry. Plots were then located in the field.

Vegetation cover data was collected in the field. The imagery collected was to provide data at ¼ and 1-meter^2 resolution, so two quadrats were used to collect data corresponding to these sizes. When a sample point was located, the center of the quadrat was placed over the end of the transect tape, and the quadrats were aligned parallel to the transect. Vegetation cover was measured by including only the area within the quadrat, and by counting plants that would have received sunlight when the sun was directly overhead. For example, an area of a particular plant shadowed by a taller plant was not included in the cover estimate. The same researcher collected all ground cover to ensure consistency. Due to aerial problems, only one-meter resolution was obtained from the DMSV. As a result, we ignored the data collected using the ¼ meter^2 quadrat in our study.

Field collection of the data started on July 17, 1999 and continued until August 22, 1999, while imagery was acquired on July 27, 1999. If data had been collected at the beginning of the summer, and the imagery was taken at the end, the imagery would not have accurately reflected the data being collected from the ground as ground cover could have changed. The heat wave that occurred during this time made data collection difficult. There were several days when the excessive heat made it too dangerous for fieldwork. Goose Creek could also only be sampled during low tide so sampling during hours when the heat was not oppressive was not always possible. As a result, sampling proved more difficult than preliminary work had

suggested. As a result, we reduced the sampling size and the number of classes. The outermost transects of Goose Creek (the 'A' and 'E' transects) were dropped, and effort was focused on the three inner transects, which provided a continuous sampling area. We dropped the number of classes to three, which meant that approximately 300 points had to be collected. In all, 283 sample points were acquired for use in our study. The distribution of points between classes was not equal and a breakdown of fifty data points per category was not achieved.

The location of each sample point had to be precisely determined. Locations of several pipes were obtained from past research at Goose Creek (Berquist, unpub. data). The geographic coordinates of all of the pipes were not available, so differential geographic positioning system (GPS) equipment was used in the field to collect additional information. The collection of such data is time-consuming, so GPS data from each pipe were not collected. Instead, eleven pipes were selected in the field, according to ease of access, along the B, C, and D transects. Their coordinates were compared, when available, with those provided by Mr. Berquist. This allowed for verification that sample points were located near the correctly labeled pipe. The accuracy of the differential GPS data is within one meter (Berquist, pers. comm.). The software program Geocalc (Blue Marble Geographics) was used to calculate the coordinates of the other pipes and the sample points. All coordinates were computed to UTM geographic coordinates, using the NAD83 datum, so as to be compatible with the remotely sensed imagery correction.

Collection of Remotely Sensed Data

Imagery was collected using the Digital Multi-Spectral Video (DMSV) sensor. The DMSV collects information from 578*740 (427,720) pixels, which is the size of its digital image. Data were collected from four bands: band 1 (770 nm - infrared), band 2 (450 nm - blue), band 3 (550 nm - green), and band 4 (680 nm - red). These four bands are useful for vegetation classification (Anderson, pers. comm.).

The DMSV was supposed to be flown at two heights to provide imagery resolution of 1-meter^2 and ¼-meter^2. However, on the day the DMSV was flown, thermal currents hindered the airplane's efforts to collect data at the ½ meter resolution. As a result, the data collected on the ground using the ¼-meter^2 quadrat could not be used. Goose Creek was flown on July 27, 1999 during low tide. Ground targets had been placed on the ground prior to image collection. The targets were 1-meter^2 pieces of styrofoam painted with various shades of gray, except for one white target. The targets allow for radiometric correction, as they have a known reflectance that can be compared to what the DMSV returns as a reflectance value. Five targets were initially placed, however only three were visible on the resulting imagery: white (reflectance of 83%), drover gray (reflectance of 45%), and universal gray (reflectance of 62%). GPS coordinates had been obtained for the seal gray target yet it was not observable on the imagery and had to be ignored. The last target, dark secret gray, was not visible on the imagery and also had to be ignored.

Geometric and Radiometric Correction of the Remotely Sensed Data

The scene was georectified before undergoing radiometric correction. All image processing was performed using Imagine (ERDAS 1994). The scene was subsampled, or cropped, to remove areas surrounding Goose Creek because georectification is more effective when the pixels used for the process are spread out across the scene. As Goose Creek was the only area of interest, we selected points from only that part of the original imagery. The image was georectified using an older image of Goose Creek that was accurate to within one meter (Berquist, unpub. data). While the vegetation characteristics of interest in this study had probably changed since the older data were acquired, features such as trees and markings on a nearby road were still visible. Seven ground correction points around the image were used to georectify the image. An affine geometric model for georectification was attempted, but for some unknown reason Imagine would not allow this model to be used. Instead, a polynomial geometric model with the following features was employed: UTM Projection, GRS 1980 Spheroid, Zone # 18, and NAD83 Datum. The total root mean square (RMS) value, which provides a measure of the precision for the georectification, equaled 0.2757.

The image was resampled using the nearest neighbor technique. In georectification, the computer fills a new grid, where each cell has a known geographic coordinate, with brightness values from the original image. In using the nearest neighbor technique, the computer assigns a brightness value to the new cells by choosing the digital number of the 'old' cell closest to the 'new' cell, if the two images were overlaid on top of one another. In order to produce an image that looks

'normal,' one must click on the option marked 'Ignore 0 in stats' in the Imagine software for georectification. It is unknown why this is so. It does not pose a problem for this research as no zero brightness values were recorded by the DMSV.

The georectification was performed before the radiometric correction because the radiometric correction could alter brightness values. We wanted to ensure that the radiometric correction produced results that would not be altered. We had placed five targets in the field but were only able to see white, drover, and universal gray on the actual imagery. Each target was visible as several pixels in the image, rather than only one as was expected, presumably because the target was captured as a component of a group of pixels centered around the location of the actual target. The aerial platform may also have not exactly achieved 1-meter^2 resolution accuracy. Unless the DMSV sampled one pixel exactly at the target's location, the target would be a part of several pixels. As a result, digital numbers were collected for the brightest part of the target visible, which we assumed to be unaltered by surrounding vegetation. We also assumed, in collecting ground cover data that the measurements for each sample plot corresponded to one and only one pixel, and was not affected by surrounding vegetation. It is possible that each sample point may actually be a component of several pixels depending on the actual space that each pixel represents.

The radiometric correction was performed by regressing reflectance onto the digital number. The radiometric correction converts all of the digital numbers, which range from 0 to 255, to reflectance values ranging from 0 to 100. Our inital correction produced results with extreme bias and gain, suggesting that a dark target was needed (Anderson, pers. comm.). A pixel from the Elizabeth River, which

contains many dark areas, was selected. The reflectance in the infrared band was assumed to be 5% for the river (Anderson, pers. comm.). Rather than assume a particular reflectance for the other bands for the dark pixel, the reflectance was computed using the ratio ($DN_1/DN_2 = Ref_1/Ref_2$), where DN_1 refers to the digital number of water in each band, DN_2 is the digital number of a given target, and Ref_1 and Ref_2 refer to the reflectance of the water and target, respectively. Ref_1 was calculated using each target and the average was then computed to find the reflectance for each band. The regressions were then recomputed to produce the following results (Figure 3):

Band 1 (770 nm): y=0.4414x – 4.0393
Band 2 (450 nm): y=0.2675x – 3.241
Band 3 (550 nm): y=0.3223x – 8.5268
Band 4 (680 nm): y=0.306x – 4.325

x=Brightness Value (0-255), y=Reflectance Value from (0-100)

Imagine separates the bands when it performs a radiometric correction into four separate images. The bands were recombined for further processing. Once a corrected image was produced, the reflectance values for each sample point were recorded. UTM coordinates of each sample point, as calculated using Geocalc, were input into a text file using Microsoft Notepad. This file was sent using File Transfer Protocol (FTP) to the Unix workstation where Imagine was being used. ArcInfo was used to convert the text file to a series of points that could be opened with Imagine. The sample points were opened with Imagine once they were converted, and the reflectance value for each sampled pixel was recorded individually for later possible

use. This task was made easier by separating the data points into ten separate files, which allowed sample points to be located on the image with greater ease.

Processing the Ground Cover Data

The vegetation cover data were processed to yield the three categories to be used for comparison with the remotely sensed imagery. All data were entered into Excel and the sum of the cover within each quadrat was calculated to ensure that it equaled 100%. Plots that did not have 100% cover, due to errors during field collection, were eliminated from the data pool leaving 283 sample points.

The data were analyzed to yield the three classes that would be used using the software program S-Plus. Hierarchical and nonhierarchical clustering techniques were used for this study. Nonhierarchical clustering was accomplished using k-means clustering. S-Plus did not offer options to affect how this calculation was performed, apart from choosing the number of clusters to be derived from the data. Hierarchical clustering was performed using a divisive clustering technique. Divisive clustering was chosen over agglomerative clustering because agglomerative clustering tends to focus on random errors in data, while divisive clustering focuses more on trends (Gauche 1995). The divisive hierarchical cluster technique was performed using Euclidean dissimilarity to distinguish clusters from one another. We also attempted to cluster the data using Manhattan dissimilarity. The resulting clusters were compared to those derived using Euclidean dissimilarity and found to match. This resulted in three distinct classes using nonhierarchical and divisive hierarchical cluster algorithms.

The data within each class needed to be divided into two sets: one set to produce the error matrix, and another set to be used as training data for the supervised classification. The classification dataset can not be used in the error matrix because training pixels would automatically be classified correctly, biasing the results from the error matrix. The data points within each class were divided in half. Matlab was used to divide points into training data and accuracy assessment data. The command 'randperm(x) > x/2,' where x is equal to the number of datapoints in each category, randomly assigned each datum to either training or accuracy data. The command generates a list of x numbers of either 0 or 1. All numbers assigned a value of '0' were used for accuracy assessment data, while data assigned a '1' were used for training data. The command ensures that each set contains the same number of data, or only one more value than the other set, for the case of odd numbered datasets. Within the training data, another series of random numbers of 0 and 1 determined whether to use a datum for signatures incorporating half of the available data, or the entire data. Data randomly assigned '0' were assigned to signatures using half of the available data as well as the signature comprised of all data, while data randomly assigned '1' were used only when the entire available training dataset was used.

Classifying the Remotely Sensed Imagery and Accuracy Assessment

All image classification was performed using ERDAS Imagine. An Area-of-Interest (AOI) was generated prior to classification. As we described, the imagery was cropped for georectification. However, part of the remaining image shows how Goose Creek fits into the surrounding area, and was not needed for the classification.

An AOI allows specific areas to be analyzed by the computer. The actual sample area is a little larger than the polygon that 'connects the dots' represented by each pipe in the field because each pipe marked the middle of a given sampling box. The extended area around each transect endpoint was calculated using Geocalc. Using Arcview, the endpoints were transformed into a vector that showed the corners of a polygon representing the study area. Connecting each point of the vector file created the AOI. The AOI was condensed slightly to reflect the actual area that could feasibly be sampled. For example, areas in the southern end of the marsh would have been included had they not been composed of forest vegetation which bordered the marsh, which was not included in this study. On the northern side of the marsh, the creek which fed into the Elizabeth River could not be crossed, therefore some areas otherwise considered available to the 'D' transect could not be sampled. As a result, the region that was classified is an irregular polygon. All data points were converted to vector format and opened in Imagine to verify that they all fell into the overall AOI.

AOI's were used to create the signatures used for supervised classification. Data points were initially entered into Excel, then exported into ArcView using FTP and converted into a vector file that could be opened with Imagine. The points for one class were grouped together and saved as an AOI. The AOI could then be entered into the Signature utility available in Imagine. Three classes were saved together in one signature file for analysis during the supervised classification.

The unsupervised classification was performed on the AOI of Goose Creek. As no pixels in our scene contained any zeros, we selected the option for Imagine not

to classify zeros. We instructed Imagine to find 3 categories, with a maximum number of 50 iterations and an agreement of 0.98. We used the minimum distance parametric rule for the supervised classification, for reasons noted previously, and the Goose Creek AOI. As with the unsupervised classification, zeros were not classified.

Overall accuracy, user's accuracy, and producer's accuracy were computed as follows:

$$\textbf{Overall accuracy} = \frac{\textbf{Correctly classified sample pixels}}{\textbf{Total number of classified pixels}}$$

$$\textbf{User's Accuracy} = \frac{\textbf{Correctly classified sample points for a particular category}}{\textbf{Number of pixels classified into a particular category}}$$

$$\textbf{Producer's Accuracy} = \frac{\textbf{Correctly classified sample points for a particular category}}{\textbf{Number of reference points available for a particular category}}$$

As we described earlier, we were interested in comparing the effect of signatures created by the entire ground-truthed dataset and one condensed to only include dominant species. To do this, the average cover was calculated for each species measured in the field. All species with an average cover of less than 15% in Goose Creek were removed. No sample points were lost due to the species reduction. The new dataset was analyzed according to the same procedures described above for the raw ground data. For example, three cover classes were generated using the spectral data and the accuracy assessment performed using these classes.

The statistical analysis for accuracy was performed by calculating total, user's, and producer's accuracy for comparisons of the different error matrices. Kappa analysis was utilized to statistically test whether the different matrices are significantly different from one another. The Kappa statistic is defined as:

$$k' = \frac{N*\sum^{r}_{i=1}x_{ii} - \sum^{r}_{i=1}(x_{i+} * x_{+i})}{N^2 - \sum^{r}_{i=1}(x_{i+} * x_{+i})},$$

where:
r = number of rows in the error matrix,
x_{ii} = number of observations in row i and column i (on the major diagonal),
x_{i+} = total of observations in row i...,
x_{+i} = total of observations in column i..., and
N = total number of observations included in matrix" (Lillesand and Kiefer 1994).

One potential problem we had with our statistical analysis was that only two error matrices can be compared at once, creating a problem analogous to performing multiple t-tests when an ANOVA would be appropriate (Zar 1984). As a result, the probability of Type I errors occurring increases in our research. An exhaustive literature search did not reveal a solution to this problem. Its effect on our research will be discussed below. As we are performing numerous statistical tests, we limited the number of tests we performed by restricting our analysis to only kappa accuracy assessment.

Diversity

During the initial design of our study, we believed it was possible to obtain diversity measures from the remotely sensed data. Consideration of the issue showed that this was not the case. We will explore this in more detail below.

RESULTS

Sample Point Precision

It was not possible to determine precisely how close the sample points actually were to their calculated UTM coordinates. We have previously discussed the issues surrounding the problem, such as the accuracy of the data provided by Mr. Berquist and the use of GPS data collected from transect pipes in the marsh. Despite not knowing this precision, we do feel that we controlled it enough to permit the use of the data in our research.

Ground Cover Data

S-Plus was used to divide the entire ground dataset into three categories using the k-means and divisive hierarchical clustering techniques. The results of this categorization are listed in Table 1.

Sixteen species were observed in the sampled area of Goose Creek. We should note that other species were observed in the A transect, which was not sampled. Goose Creek thus has a higher proportion of species than our data suggests. Table 1 shows the classes that resulted from clustering the data using the k-means and divisive hierarchical algorithms. All data represent percent cover, but as we are presenting averages the sum does not exactly equal 100%. Large standard deviations

were observed in the data for several species. The mean percent cover and the standard deviation for each observed species are listed in Table 2.

As we described previously, the dataset was condensed so that only species with an average cover of more than 15% across the entire marsh were included. The following species remained: No Cover (which is being treated as a 'species' for the purposes of our work), *Spartina alterniflora, S. cynosuroides, Phragmites australis, Scirpus* spp., *Amaranthus cannibina, Typha* spp., and *Aster* spp. The condensed dataset was created in an attempt to remove variability that we believed less-important species would introduce into the clustering algorithms. The k-mean and divisive hierarchical clustering tools were run on the condensed data. The results of removing the less dominant species are listed below, next to their counterparts that used the entire dataset. There is very little difference in the percent cover of k-mean clustering between the entire dataset and the one that was condensed (Table 3), and none at all for the divisive hierarchical clustering algorithm. A side-by-side comparison of the sample points showed that only one sample point moved from group 1 to group 2 when the condensed dataset was processed using the k-mean algorithm. No points moved out of or into any group when the divisive hierarchical algorithm was used on the condensed data.

It is somewhat difficult to actually interpret what the classes represent by looking at the entire dataset, but since there is virtually no difference between the entire dataset and the condensed one in terms of how the sample points were distributed, we can focus on the eight species we have found. Class 2 in k-means clustered data is almost entirely dominated by *Phragmites*. Due to the low percent

cover of this species in the other classes, and the relatively low cover of other species in class 2 (all almost an order of magnitude less), we can regard class 2 to consist almost exclusively of *Phragmites* vegetation. The low standard deviation for *Phragmites*, relatively speaking, supports the concept that class 2 is essentially only *Phragmites*. This result agrees with our field observations where *Phragmites* appeared to be a largely homogenous class. Sample points fitting into class 2 would have been easily distinguishable in the field as large patches of *Phragmites* were visible to the naked eye. Separating class 1 from class 3 was more difficult. In a sense, every sample point represents a distinct class as no two samples were precisely the same. Combining sample points into classes facilitates possible identification of trends across the marsh. We were restricted to the number of classes we could use by the requirements of the accuracy assessment of remotely sensed data. As a result, class 1 and 3 may actually represent an amalgam of smaller, more distinct, classes that had to be placed into one or the other class because they were more similar to each other than they were to class 2 and *Phragmites*. As a result, it appears that class 1 is comprised of a mixture of *S. alterniflora* and *A. cannibina* while class 3 is largely dominated by *Scirpus* spp. The large standard deviations, relative to average ground cover, may be important in discerning how distinguishable these classes are. The variation of class 2 appears to be larger than that of class 1. The breakdown agrees somewhat with subjective field observations. *S. alterniflora* and *Scirpus* were frequently observed together, as the standard deviations above suggest. *A. cannibina* was noticed more frequently with *S. alterniflora* than with *Scirpus*. Thus, we subjectively agree with these results, though there are differences across the marsh

that we did notice that the clustering did not discern due to constraints on the number of classes allowed. For example, *Typha* did not appear on the northern side of the marsh while it was present in the southern side as a relatively frequent species. While we may have felt that there was a pattern present as one moved from north to south across the marsh, the data we have collected can not reflect this due to limitations created by the sampling needs.

Clustering using the divisive hierarchical clustering algorithm produced slightly different results (Table 4). As before, *Phragmites* appeared to separate out as a separate class by itself. In this case, S-Plus assigned it to class 3. The standard deviation and value appear to be comparable to that of class 2 calculated using the k-mean algorithm. Class 1 and 2, as determined by the divisive hierarchical classification, appear to be different from that calculated using the k-means classification scheme, however they still seem to be comprised of an amalgam of species. Class 1 appears to be most heavily dominated by *Scirpus spp.*, while class 2 has the greatest cover from *S. alterniflora*. Again, subjective field observations suggested this difference; patches of *S. alterniflora* were distinguishable from *Scirpus* spp. in Goose Creek. We should note that these coverages do have large standard deviations, as seen in Table 2. The presence of 'no cover' appeared to also play a role in discerning class 1 from class 2 in that its 'presence' is more than double in class 2. *S. cynosuroides*, *A. cannibina*, and *Typha* spp. are more common in class 1 than class 2. Class 1 could be regarded as being more of a collection of species than class 2, which is largely dominated by No Cover, *S. alterniflora*, and to a smaller degree, *A. cannibina*.

We would like to note that subjective field observations might have led to a different breakdown of classes based upon distance into Goose Creek. *Phragmites* was clearly a dominant species and would have been entered into its own class. The other two classes would have been divided into low marsh and high marsh.

The large standard deviations of our data suggest that the k-means and divisive hierarchical classification tools are imperfect classifications. As clustering tools, they can be used to separate the data, but they can not describe what is causing the trends that led to the clustering. Principal components analysis (PCA) was used for this purpose. Figure 4 shows the results of the PCA as calculated using CANOCO. Figure 5 shows the results of the PCA performed using S-Plus. If we assume that the PCA was calculated using the same parameters (it was not possible to set precise terms on the S-Plus calculation), the two outputs reveal the same results using a slightly different perspective. Figure 4 shows the output by species, while Figure 5 actually breaks apart the components. The variation appears to be largely driven by three factors: *Phragmites*, *Scirpus* spp., and *S. alterniflora*. In addition, *S. cynosuroides* and No Cover also appeared to explain variation within the data. Figure 5 confirms this. The first five components explain over 95% of the variation in our data. The first component is driven almost entirely by *Phragmites*. Combined with the results of figure 4, which shows *Phragmites* separate from all the other species, it suggests that *Phragmites* is a stand alone species. The divisive hierarchical and k-means classification demonstrate this result, as *Phragmites* strongly appeared as its own class using both clustering tools. The second component, explaining an additional 15% of the variation, is caused by *Scirpus* spp. and *S. alterniflora*, which

run in opposite directions. Figure 5 also demonstrates these results. The components attributed to *Scirpus* spp. and *S. alterniflora* are not as long as that of *P. australis*, but they are still the next two longest lines on the PCA diagram. They also are separated, suggesting that they will not be found together within a sample point. The results agree with the results produced by the k-means and divisive hierarchical clustering algorithms. The other trends appear less important but still reveal patterns in the data. The species we had separated out to form the condensed data set are the species that S-Plus found explained the greatest trends in the data. No Cover and *S. alterniflora* are slightly distinct from one another, as are *A. cannibina* and *Scirpus* spp. In addition, *Typha* spp. and *A. cannibina* are opposed to one another, suggesting that they are not frequently found together in the wetland. These results agree with our subjective field observations.

The k-mean and divisive hierarchical clustering techniques were also applied to the entire dataset to produce five distinct clusters, instead of the three described previously. While we did not perform classification and accuracy assessment on the expanded clusters, we did want to compare them with the data clustered in three categories.

The result of categorizing the data using k-means into five categories is provided in Table 5. If we focus on the species comprised of major cover, it appears that Class 1 is heavily dominated by *P. australis*. All cover in Class 1 is at least an order of magnitude below *P. australis*. *Phragmites* also appears to be important in Class 2. While it is not as dominant as it is in Class 1, it still has the largest percent cover of any plant species. No cover and *S. alterniflora* also appear to be significant

components of class 2. No cover has its highest cover percentage in Class 2. Class 2 could possibly be interpreted to be comprised largely of No Cover, *S. alterniflora*, and *P. australis*, with a little *D. spicata* as well. Class 3 is heavily dominated by *S. alterniflora*, while *A. cannibina* also has a high percentage of cover represented. No Cover also is an important component of class 3. *S. cynosuroides* and *Scirpus* spp. also appear to have higher cover in Class 3. Class 4 is almost entirely dominated by *Typha* spp. No Cover is also present, as it is in every other class. *P. australis, S. alterniflora*, and *Scirpus* spp. are also present. Class 5 appears to be heavily dominated by *Scirpus* spp. with smaller cover from No Cover, *S. alterniflora*, and *S. cynosuroides* all present.

The result of categorizing the data into five clusters using the divisive hierarchical clustering algorithm is given in Table 6. The composition of the clusters produced using the divisive hierarchical clustering algorithm appears to have produced different results from that of k-means clustering. Class 1 is largely dominated by *Scirpus* spp., but No Cover, *S. alterniflora*, *S. cynosuroides*, *A. cannibina*, and *D. spicata* all appear to also be a part of this class. Class 2 is largely dominated by *S. alterniflora*, with No Cover and *A. cannibina* also present to a lesser extent. Class 3 is dominated almost entirely by *P. australis*. Class 4 appears to be comprised mostly of patchy *S. alterniflora*, as the presence of No Cover is the dominating feature of this class. *S. cynosuroides, P. australis,* and *Scirpus* spp. are present as much smaller features. Class 5 is comprised largely of *Typha* spp. with some bare patches as demonstrated by the presence of No Cover.

Importance Values

The data was analyzed according to ground cover because it is the only ecological measure actually assessed by remote sensing. Unless one is working at a very high resolution, it is impossible to count individual plants and determine relative density. Measuring frequency is only possible insofar as a given plant's brightness value within a given pixel is strong enough that the pixel is classified to that given class containing the species in question. As a result, any maps produced, including supervised classifications, rely on cover. However, calculating the importance value, which neglects any influence from lack of cover, may allow additional information to be obtained about the ecology of the area. We had to use a modified importance value as we did not have time to collect data used for relative density. We note that these values may be slightly biased because not all plants were classified to the species level, such as the *Scirpus, Aster,* and *Typha* genuses.

Table 7 shows the results of calculating the importance value for the k-means and divisive hierarchical clustered data in three categories. The results appear similar to our dominant species determined based on cover. For k-means clustered data, *S. alterniflora* is the clear dominant in class 1, followed by *A. cannibina*. Other species that appear important to the composition of this class include *Scirpus* spp., *P. australis*, *Aster* spp., *Typha* spp., and *S. cynosuroides*. Class 2 is almost entirely dominated by *P. australis*, though *Scirpus* spp., *S. alterniflora*, and *A. cannibina* also appear to play a role. Class 3 is largely dominated by *Scirpus* spp., though *S. alterniflora*, *S. cynosuroides,* and *A. cannibina* also appear to be important to this

class. The importance values we measured do appear to suggest the same dominant species for each class that we have calculated before.

We obtained similar results for data clustered using the divisive hierarchical technique (Table 7). Class 1 is again dominated by *Scirpus* spp., with *S. cynosuroides, S. alterniflora*, *A. cannibina*, and *Typha* spp. also common members of this class. Class 2 is largely dominated by *S. alterniflora,* with *A. cannibina* as noted before. *Aster* spp., *S. cynosuroides*, *P. australis*, and *Scirpus* spp. also appear to play a role in the composition of this class. Class 3 is dominated almost entirely by *P. australis*, with *S. alterniflora, Scirpus* spp., *A. cannibina,* and *Typha* spp. also present, though their importance values are much lower than that of *P. australis*. These findings appear to support our previous results based only on cover estimates, even with lack of ground cover not under consideration.

Table 8 shows the results of calculating the importance value for the k-means and divisive hierarchical clustered data in five categories. Like the data in three classes, there appear to be strong similarities between these results and the analysis done by examining ground cover alone (Tables 5 and 6). In the k-means clustered data, *P. australis* is the most dominant species, though *Scirpus* spp. and *A. cannibina* also appear in this class. Class 2 is an amalgam of species. *P. australis* and *S. alterniflora* are the most dominant. However, *D. spicata, A. cannibina, Typha* spp., *Scirpus* spp., and *S. cynosuroides* also appear to be less dominant, but still important, species in this class. Class 3 is dominated by *S. alterniflora*, but *A. cannibina* is also important. *Scirpus* spp., *S. cynosuroides*, and *Aster* spp. also appear to be important to the composition of this class. Class 4 is most heavily dominated by *Typha* spp.,

but *S. alterniflora, P. australis*, and *Scirpus* spp. are also present. We should note that class 4 was comprised of only 4 sample points. *Scirpus* spp. is the most dominant species in class 5, but *S. alterniflora*, *S. cynosuroides*, and *A. cannibina* also contribute to this class. As before, these class distinctions are similar to those found from just examining the ground cover data.

The classes derived from the importance values calculated for the divisive hierarchical clustering to five classes also resemble those obtained from the examining only the ground cover data. Each class appears to be composed of one major dominant species and several minor dominant species. Class 1 is dominated by *Scirpus* spp., with the minor dominants including *S. alterniflora, A. cannibina*, and *S. cynosuroides*. *S. alterniflora* is the dominant species of Class 2, and lesser dominants include *A. cannibina, Scirpus* spp., *Aster* spp., *P. australis*, and *S. cynosuroides*. Class 3 is heavily dominated by *P. australis*, though *Scirpus* spp., *A. cannibina, S. alterniflora,* and *Typha* spp. appear to be minor dominant species in the class. Class 4 is an amalgam of many species. *S. alterniflora* is the major dominant species, but minor dominant species include *A. cannibina*, *P. australis, Scirpus* spp., *Typha* spp., *S. cynosuroides*, and *Aster* spp. Class 5 only contains four points but is heavily dominated by *Typha* spp., with *S. alterniflora*, *P. australis*, and *Scirpus* spp. present as minor dominants in the class.

Classification and Accuracy Assessment

The clusters describe the groups that were input into Imagine for classification and accuracy assessment. As described previously, half of the data was used to

produce supervised classifications and half was used to test the accuracy of the remotely sensed data. The data used to produce the supervised classifications was again split in half to test the hypothesis that effort affects accuracy.

The classifications produced by ERDAS are presented in Figures 6-16. Each image was analyzed using the Kappa statistic to determine whether it was significantly different from a random classification. An error matrix was produced for each accuracy assessment. We have listed one in Table 9; the others produced during our study are provided in the Appendix. Kappa was calculated according to the formula described previously. Theta One, Two, Three, and Four were determined using Excel to allow the variance of the Kappa statistic to be measured, which was then used to calculate the Z-score. A Z-score greater than 1.96 suggests that Kappa value is significant at an alpha level of 0.05. In this case, the classification was significantly different from one produced randomly.

Only one unsupervised classification needed to be performed, as we were only interested in one image with three classes. The unsupervised image is based upon what the computer interprets as the natural separation of the brightness values for each pixel based upon some mathematical formula. The error matrix allows us to test that against what we perceive to be the 'correct' result. However, in this case we have four 'correct' accuracy datasets: divisive hierarchical clustering of the complete and condensed data, and k-means clustering of the complete and condensed data. Each could potentially lead to a different Kappa value and accuracy assessment. Another factor to consider is that the classes produced by an unsupervised classification may not correspond to the classes produced for a supervised

classification. The class dominated by *Phragmites australis* in our data may not appear as a class in the unsupervised classification. However, we are still interested in assessing whether the unsupervised classification is more accurate than one produced by random, without any mathematical consideration given to the separation of pixels. If we assume that the clusters produced from the k-mean and divisive hierarchical clustering will correspond to three classes produced in the unsupervised classification, we can use them to produce error matrices for the unsupervised classification. To simplify matters, we only used the complete dataset for the unsupervised accuracy assessment. Since class 1 of the unsupervised classification may not correspond to class 1 of the clustered data, we also assumed that the 'shuffling' of the data that led to the best Kappa value was the correct one. In other words, class 1 from the k-means clustering corresponds to class 2 of the unsupervised classification if it helps achieve a higher Kappa value for the unsupervised classification.

The results of the Z-test for each of the error matrices computed are presented in Table 10. We described earlier how the complete and condensed datasets were separated into almost exactly the same clusters when run through the divisive hierarchical and k-means algorithms. However, Table 10 shows that the Kappa values and Z-scores are different for the error matrices produced for the complete and condensed datasets. The error matrices for classifications produced on the complete or condensed datasets do not yield precisely the same Kappa values or Z-scores because different subsamples of the data were used to produce them. As we described previously, we assigned random numbers to the data to determine whether

each sample point was to be used for training or accuracy assessment. The random numbers were not assigned to the same sample points in the complete and condensed datasets. As a result, the error matrices were not evaluated according to the same values.

The kappa statistic produced for the complete dataset, k-means clustering, using only half of the available data, is the only insignificant result from all of the individual error matrices we analyzed. We found this result highly unusual. Verification of our methods failed to find an error in the production of the error matrix. We repeated the experiment by randomly selecting samples again and dividing them into accuracy/training data as well as randomly selecting half of the training data for use in signatures developed using half or all of the available training data. The retest was performed on the complete dataset clustered using the k-means algorithm, where the training signatures were created using half or all of the available training data. The results from this work are listed in Table 11. Table 11 shows that, for the retest, a significant result was obtained. That is, the classification produced from the complete dataset, clustered using the k-means algorithm, and using only half of the available data to create a training signature, is significantly different from a classification produced at random. However, an extensive verification showed that the original result, which suggested that the classification was not significant, is also valid. The effect of this on our work will be discussed in the next section.

We compared error matrices against one another (Table 12) to explore whether there were differences in how well they classified Goose Creek. The statistics do not allow us to determine what differences in the thematic maps may

exist, only whether one map produced a significantly better-than-random result. We have used both kappa values (0.123 and 0.338) for tests involving 'comkall' and listed those values separately. The results in Table 12 show that the unsupervised classifications are not significantly different from classifications produced using training data. In fact, with the exception of only one test, all images are not significantly different from one another. The one test to produce a significant result was that of comkhalf and comdvhrhalf. The only difference between these two classifications is that one used k-mean clustering, and the other used divisive hierarchical clustering. We should note that there is no significant difference when the test is performed using the alternate kappa value for the k-means classification, obtained during the verification of results.

User's and Producer's Accuracy

While the error matrix as a whole can be tested for significance, it is also possible to examine the features of each class within a given classification separately as well. Due to the number of statistical tests that we have already performed, we are not going to test the user's and producer's accuracy from each classification we have produced, as such a description would be unwieldly. However, we will briefly examine two error matrices for discussion purposes. We have selected the classifications produced using k-means and divisive hierarchical clustering, with all available data used to create the training signatures.

The results from the k-mean clustering (Table 13) show an overall accuracy of approximately 50%, which means that about half the time the computer was able to

correctly classify a given pixel. Since there were only three classes, if the computer were randomly assigning pixels to classes we would expect it to make the correct assignment about 30% of the time. There are notable differences between the user's and producer's accuracy. The user's accuracy for class 1 is 70% but falls to 30% for the producer's accuracy. Class 3 shows an opposite trend. Its user's accuracy is 28%, much less than its producer's accuracy of 65%. The kappa values are all different from the overall kappa value for this error matrix, 0.29. Class 1 appears to be much more significant, while class 3 would appear to be much less.

The results from the divisive hierarchical clustering (Table 14) also show an overall accuracy of approximately 50%. The differences between user's and producer's accuracy do not appear to be as prominent but are still present. The user's accuracy for class 1, 43%, is less than its producer's accuracy of 70%. The producer's accuracy for class 2, however, is less than its user's accuracy. The difference between user's and producer's accuracy for class appears to be closer. The overall kappa value for this error matrix was 0.34. Class 1 appears to have the lowest kappa value while class 2 has the highest. As we not performing statistical tests on each class, it is unknown precisely whether these differences are significant.

DISCUSSION

The objective of our research was to explore the usefulness of remote sensing as a tool for wetlands research by comparing it to ground-based, 'traditional,' techniques. Our results suggest that more work remains in determining protocols for collecting remotely sensed data. There are numerous mathematical steps required to process and analyze remotely sensed data. It appears that a small change at one of those steps can have a large effect on the final outcome on the classification produced using remotely sensed data. In addition, the statistics available for analyzing remotely sensed data are still developing. While much progress has been made in this field (e.g. Congalton and Green 1999), our work has highlighted problems that still exist. We thus suggest that remote sensing be used as a tool concurrently with, rather than instead of, traditional methods. Remote sensing can highlight trends not immediately obvious with the 'traditional' methods we employed in our study. However, to obtain additional information that remote sensing can provide, traditional methods must be employed first, even if only as a preliminary survey, to collect data in the field.

Clustering Analysis

The data we collected in the field was analyzed by using the k-means and divisive hierarchical clustering algorithms. Clustering allows individual sample points to be collected in such a way that data within one class are more similar to

other points in the same class. A variety of mathematical procedures exist that can be used to determine how similarity between samples is to be measured (e.g. SAS Institute, Inc. 1994). Our experiment used two clustering techniques, k-means and divisive hierarchical, that have been used in the traditional literature for ecological analysis (Gauche 1995). K-means clustering is a non-hierarchical technique. Clusters computed using k-means are exclusive of one another. As calculated by S-Plus, the software used in our experiment, the centroid for each group is determined and each datum is assigned to the nearest centroid (MathSoft, Inc. 1999). Divisive hierarchical clustering works by dividing the data set into successively smaller clusters until the desired number of clusters is achieved. Divisive hierarchical clustering is preferred to agglomerative hierarchical clustering because the latter procedure tends to focus on differences that could be due to random errors in individual data points.

We did not explore other options, such as fuzzy clustering, where sample points can be assigned fractionally to more than one group. While such techniques have been applied to remote sensing, we wanted to test our hypotheses as simply as possible. Using just the k-means and divisive hierarchical techniques showed that even simple tests produced complex results. From an ecological perspective, fuzzy clustering might have been useful in our experiment. Combining different species into mutually exclusive categories can be a questionable procedure. By their very nature, wetlands represent a transitional zone between uplands and coastal waters. As a result, species may not fall easily into one and only one class.

The clusters produced from our data appear to support these ideas. We opted to divide the data into only three clusters because of the needs of the remotely sensed data. As such, we expected the clusters to reveal only very general trends across Goose Creek. It was interesting to note that the ground cover data alone appeared to provide similar results to the modified importance value measure, which included relative frequency and relative dominance. Results from both techniques appeared to highlight the most dominant species of each class. The importance value does appear to provide a better way of separating out the dominant species in the clusters by utilizing more information, frequency, which ground cover data alone can not. In addition, the importance value ignores ground cover and compares vegetative species relative to one another. Remotely sensed data can not ignore lack of cover, as it may exist as a component of a class visible to the sensor, as it did in our experiment. This is why we used ground cover alone in creating our condensed dataset. However, the similarity between classes obtained from ground cover data alone and the importance value suggests that studies using remotely sensed data to derive classes could use either measure. However, we do not know what effect adding relative density to our modified importance value would have had on our research. While it was not tested in our experiment, it would also be interesting to see what effect, if any, exists in using frequency alone to determine classes. If only frequency information was needed, substantial time might be saved in the field since frequency is one the easiest parameters to measure. One minor problem with using the importance value is that it does not provide information on class variability, which we were able to determine using the ground cover data.

Summarizing our results, the clusters appeared to reveal the following dominant trends in Goose Creek:

K-Means Clustering – Three Classes

Class 1: *S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*
Class 2: *P. australis*,
Class 3: *Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina*

As we noted in our results, separating class 1 from class 3 was difficult due to the variability of those classes. It is possible that class 1 and 3 actually represent an amalgam of smaller, more distinct, classes that had to be placed into one or the other class because they were more similar to each other than they were to class 2. The importance value calculations showed the same species in Class 1 and 3, except that the ranking of species appears to be much different. Lack of cover was more prominent in class 1 than it was in class 2 or 3. The breakdown agrees somewhat with subjective field observation. There were differences across the marsh we noticed that the clustering did not discern due to constraints on the number of classes allowed. For example, *Typha* did not appear on the northerly side of the marsh while it was present in the southern side as a relatively frequent species. Compiled into only one of three classes, the 'signal' from *Typha* was lost in the clustering. While we may have felt that there was a pattern present as one moved from north to south across the marsh, the data we have collected can not reflect this due to sampling limitations. While clustering is useful because it provides an objective way of

looking at the data, we need to be aware of its limitations. Three categories may not be the ideal number of classes to separate the Goose Creek wetland. In addition, the analyst must still decide how each class is distinguished from the other classes. Tools like PCA can add further objective input to this process.

Applying divisive hierarchical clustering revealed the following dominant trends in Goose Creek:

Divisive Hierarchical – Three classes
Class 1: *Scirpus* spp., *S. cynosuroides*, *A. cannibina*, *Typha* spp.
Class 2: *S. alterniflora*, No Cover more prominent, *A. cannibina*
Class 3: *P. australis*

As before, *Phragmites* appeared to separate out strongly as a separate class. Classes 1 and 2 still appear to be comprised of an amalgam of species. Class 1 appears to be most heavily dominated by *Scirpus spp.*, while class 2 has the greatest cover from *S. alterniflora*. Again, subjective field observations suggested this difference; patches of *S. alterniflora* were distinguishable from *Scirpus* spp. in Goose Creek. Lack of vegetation is once again important; the 'presence' of No Cover is more than twice that in class 1. *S. cynosuroides*, *A. cannibina*, and *Typha* spp. are more common in class 1 than class 2. Class 1 could be regarded as being more of a collection of species than class 2, which is largely dominated by No Cover, *S. alterniflora*, and to a smaller degree, *A. cannibina*.

The differences between the k-mean and divisive hierarchical clustering reflect the mathematical differences in how they are calculated, but ecological interpretation of these results is possible. The dominant species indicate features

about the Goose Creek wetland. Goose Creek is a created tidal oligohaline marsh. The species we have found represent both saltwater and freshwater tolerant species. *Scirpus* spp., for example, is found in brackish or freshwater wetlands. *Scirpus americanus*, one of the *Scirpus* species we found, grows well in tidal wetlands (Silberhorn 1982). *A. cannibina* is a species able to grow well in brackish waters and is usually a less dominant species in marshes than what we observed (Barnard, pers. comm.). *S. cynosuroides* is another species that grows well in salt or freshwater wetlands (Silberhorn 1982). *S. alterniflora* is a species also able to survive in freshwater and saltwater wetlands. However, it is adapted for handling salt conditions that other plants can not (Silberhorn 1982). Our data show *S. alterniflora* clustered with other species. If the marsh were experiencing heavy saltwater inundation, we would have expected *S. alterniflora* to be found largely in its own cluster, with much lower dominance by other species. As we have not found this, we can surmise that *S. alterniflora* must compete with the other plants for available resources in the marsh. The rarer species we found also indicate features of the Goose Creek marsh. For example, *D. spicata* and *S. patens*, which were observed as less dominant, are frequently found in brackish wetlands or saltmarshes (Silberhorn 1982). It is also possible that the topography of the Goose Creek marsh may be affecting the type of vegetation through changes on tide level and sediment accumulation in the marsh (Barnard, pers. comm.).

Our results show that *P. australis* is able to outcompete other plants. Even with the other class differences found between the k-means and divisive hierarchical clustering, the signal from *P. australis* was strong enough to stand largely alone as the

only dominant in its class, even with only three classes produced during clustering. *P. australis* was found in the other two classes when another species was dominant, but the converse was not true. This suggests that once *P. australis* establishes itself, other species are not able to compete for space. Whether *P. australis* is beginning to dominate where it is found with other species, or is unable to outcompete due to environmental conditions in those parts of the marsh is not immediately clear. *P. australis* tends to grow in freshwater marshes and in areas where wave action is low (Marks *et al.* 1994). The salinity range for *P. australis* has been reported as 12 ppt to 29 ppt (Marks *et* al. 1994). This might explain why it is has not yet dominated Goose Creek. *S. alterniflora* and other species tolerant of brackish water might be better able to withstand the tidal action and salt water influx coming into Goose Creek. As a result, these species may persist only where *P. australis* is unable to outcompete them. As the marsh is oligohaline, which tends to range in lower salinities, there might be other unknown factors explaining the movement of *P. australis* across Goose Creek. The maps generated from the remotely sensed data will shed additional light on this issue, described below.

The results from clustering indicate that our data contain some variability. There is no reason for preferring either the k-means or the divisive hierarchical clustering technique to the other. They both show that *P. australis* is a strongly dominant species in Goose Creek. It was not one of the original species planted in the mitigation bank, suggesting that it has been able to invade with remarkable success (Barnard, Jr. *et al.* 1997). The other feature in common with both clustering techniques is that *S. alterniflora* was associated with a higher proportion of No Cover

than the other classes were. The reason for this is unclear. *S. alterniflora* may not be as successful at colonizing as other species. We have noted that the species grows where other species can not. The converse of this is that it is probably outcompeted in other areas. There may be other reasons as well. Young plants may have trouble establishing because of the conditions *S. alterniflora* inhabits.

Tables 5, 6, and 8 show the result from producing five clusters using the k-means and divisive hierarchical techniques. Our description from the results is summarized below:

K-Means – Five Classes
Class 1: *P. australis*
Class 2: *P. australis* (much less than Class 1 but still notable), No Cover, *S. alterniflora*
Class 3: *S. alterniflora*, *S. cynosuroides*, *Scirpus* spp. (the importance value calculations also would suggest *A. cannibina* is a dominant).
Class 4: *Typha* (main dominant), *P. australis*, *S. alterniflora*, *Scirpus* spp.
Class 5: *Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides*, *A. cannibina,* No Cover

Divisive Hierarchical Clustering – Five Classes
Class 1: *Scirpus* spp. (main dominant), No Cover, *S. alterniflora*, *S, cynosuroides*, *A. cannibina*, and *D. spicata* also common
Class 2: *S. alterniflora* (main dominant), No Cover and *A. cannibina*
Class 3: *P. australis*
Class 4: No Cover dominates, *S. alterniflora* (main dominant species), *A. cannibina*
Class 5: *Typha*

We can infer many of the same conclusions from these results that we have just derived about Goose Creek. The inclusion of *Typha* into its own class is an interesting addition. However, it should be noted that only four sample plots from the entire dataset were included in this class. We did not classify the remotely sensed

imagery into five classes because of limits in testing the accuracy and do not want to dwell long on these results. However, they highlight that adding classes to cluster analysis can possibly change the interpretation of trends in a marsh. Previously *Typha* was included into other classes. We now find that, by slightly expanding the scope of investigation, *Typha* is prominent enough to fall into its own class. One must therefore be very careful in deciding what classes to use for remote sensing.

Diversity

We had initially wanted to test whether remotely sensed data produces diversity index values similar to those obtained from data collected on the ground. While it is possible to use the class or brightness values for the diversity equations, we realized late in the experiment that it would be irrational to do so. Diversity measures the richness, evenness, or both, of an ecosystem (Magurran 1998). Richness refers to the number of species found in the ecosystem of interest, while evenness refers to the balance of the number of individuals among the species found (Magurran 1998). A variety of measures exist that allow richness and evenness to be estimated separately, or that combine the two into a given mathematical formula (Magurran 1988).

We initially believed that diversity could be calculated from remotely sensed data. The classes from a thematic map produced by a classification, or the raw brightness values themselves, could be used. We decided that diversity should not be calculated from either. While using the classes may not appear to be an improper way to calculate diversity, examining the composition of the classes reveals why they

should not be used. It is impossible for us to break apart the composition of the classes simply by looking at them. They represent conditions found on the ground. In our study, each class represents components created by applying k-means or divisive hierarchical clustering to the data, which is not a new concept. Mueller-Dombois and Ellenberg (1974) discuss how classification can answer general questions about features such as ecosystem habitat or function. Similar species combinations often recur, and classification allows us to categorize those changes, which can help to explain them. However, in calculating diversity one needs to be able to get to the actual 'numbers' describing the scene. It is impossible to get that information from a thematic map unless the scene was classified to the level for which diversity was desired. For example, we wanted diversity information at the species level for our experiment. As we described previously, it was impossible to gather enough ground data to allow such precision. In addition, the heterogeneity of Goose Creek would have caused misclassification of many pixels in the final thematic map. We can look within each cluster to see the individual species from each sample point. This is not possible with the remotely sensed data.

One might argue, as we initially did, that there is a possible relationship in diversity measurements between ground data and the brightness values obtained through remote sensing. However, the definition of richness and evenness, as applied to ecological monitoring, make their application to remotely sensed data very difficult. The problem with richness is that the number of bands is set before the experiment begins. One might perform an experiment to see how many species are present in an area. The exact number is undetermined before sampling actually

begins. With remote sensing, the number of bands is already set. Any attempt to correlate richness with remotely sensed data is therefore impossible because the analyst 'interferes.' We would argue that an area with ten species is more diverse than an area with two. We do not believe that it is really analogous, and therefore legitimate, to say that five bands are more diverse than two. Depending on the scene of interest and the equipment available, it might be better to only use two bands in a given experiment.

Evenness is also invalid because there is no equivalent analogy that would permit us to use it. We tend to think of areas with ten individuals in each of four species as more diverse than an area where three species have one individual and the other species has 36. In remotely sensed imagery the digital number does not signal diversity. A brightness value of 100 is not necessarily stronger than a brightness value of 2, except insofar as it affects the actual sensor. The combination of brightness values across bands is used to distinguish and classify features in a scene. A high ecological diversity could be obtained from many scenes and yet, depending on the nature of the area in question, the brightness values across bands could be substantially different.

While richness and evenness are not plausible measures that can be calculated from remotely sensed data, one should discount diversity altogether. Other measures of diversity, such as spatial diversity are possible (Robinove 1986). These measures examine the edge of classes and how many other different classes they border. Areas with high spatial diversity, for example, border many classes while areas with lower spatial diversity are more homogeneous. While we did not calculate such diversity

with our data, they do show that features can be calculated from remotely sensed data that can not be estimated using ground-truthed data collected as we did in our study.

Accuracy

We must examine the accuracy of our data before addressing what ecological information we can obtain from the remotely sensed imagery. Our statistical results are unclear and almost contradictory. A substantial problem with our research was that the significance of each error matrix had to be tested separately. Each desired comparison of error matrices also had to be performed individually. This is analogous to performing multiple t-tests when using an ANOVA is preferable. The probability of committing a Type I error, rejecting the null hypothesis when it is in fact true, increases as the number of means being compared increases (Zar 1984). Our research tested the significance of 10 kappa values, not including the two that were resampled and recalculated, and we performed 16 tests to compare various error matrices with each other. Zar (1984) notes that given an alpha of 0.05, the probability of committing a Type I error by using multiple t-tests to calculate the significance between 10 means is 63%. For 20 means, the probability increases to 92%. Based on these findings, it is more than likely that our work suffers from Type I errors. When we obtained the kappa of 0.123 for the error matrix produced using the complete dataset, k-means clustering, and only half of the available data for training, that may have been reflecting the true result. The results from the verification rejected the null hypothesis. This was the only error matrix to produce a nonsignificant result, and the two different kappa values for the same classification lead to highly different

interpretations of the effectiveness of remote sensing. We will examine both interpretations and attempt to reconcile them.

Interpretation 1

The first interpretation is based upon the kappa values we obtained from the verification test we performed, which is simpler to interpret than that using the nonsignificant result. The results showed that every error matrix was significantly different from one produced by randomly assigning each pixel to a given class. Each pair of classifications we tested also yielded insignificant results, suggesting that no one classification was better than the others. The results do not imply that the interpretations of the images would be the same, only that one classification is not better than the other in terms of being able to separate the pixels into one of the three classes we set.

These results suggest that the mathematical formula used to classify the data and the effort spent ground-truthing are not important. The unsupervised classification yielded equally significant results when its accuracy was tested against the data produced using both the k-means and divisive hierarchical clustering techniques. We also obtained insignificant results when we compared classifications produced using the same amount of training data, but the accuracy standard was based on either the k-means or divisive hierarchical clustering technique. This suggests that the clustering techniques are similar enough to each other that Imagine can produce a thematic map that is interpretable using either procedure. The results do not suggest that the interpretation of each thematic map would be the same, as

there are slight differences in the composition of each class depending upon whether the k-means or divisive hierarchical clustering technique was used.

The unsupervised classifications were not significantly different from the supervised classifications, which used two levels of training data (half and all available data) to form their signatures. We had expected that accuracy would improve as the amount of data used to create spectral signatures was increased. It is possible that we obtained this result because the training signatures consisted of a small number of pixels. If we had used 500 pixels instead of only 142 to create the training signatures, we might have found accuracy improves as training sample increases. However, obtaining extra data imposes extra costs and time in the field. Whether such effort is necessary is a decision left to the user of the data who must decide how comfortable they feel with the accuracy of the data. One should not assume that these results imply that ground-truthing is not necessary. At the least, field collection of data allows the user to know what classes produced by the classification represent. Wetlands such as Goose Creek are comprised of plants that can change dominance over a season due to the growth patterns of annuals and biannuals. Applying our results to the next growing season may produce a faulty imagery assessment. One should be very careful in interpreting imagery that has not been ground-truthed. However, our results do suggest that a minimum of ground-truthing can be performed to identify classes, which can lead to savings in time and money.

As we described before, the condensed dataset and the complete dataset were comprised of almost exactly the same sample points. It is therefore not surprising

that we obtained similar results for analogous tests between them (comparing complete and half training signatures, for example). These results suggest that one could save time during analysis by focusing only on dominant species. In the field, all species would still need to be recorded for each plot, as the data could be analyzed to reflect species trends across the marsh. Such results could demonstrate characteristics such as diversity and functional changes across the marsh. If one is interested in rare species, such information is also critical.

Interpretation 2

The first interpretation of the data suggested that there was no difference between classifying the data using either the k-means or divisive hierarchical technique, and that the amount of effort used in classification was unimportant. Changing the kappa value from 0.123 to 0.382 for the one error matrix we described, however, changes the interpretation of the data. If we assume that the results for the other error matrices are correct (e.g. no other Type I errors were committed), the interpretation changes. Instead of our results being 'simple,' we now appear to have interactive effects between the cluster technique and the amount of data used to produce the training signatures. The only error matrix that is not significantly different from one produced randomly is the classification based on the complete dataset, k-means clustering, and using half of all available training data. Another classification performed with the same features, except for using the condensed dataset, is still significant. We randomly selected different samples from our dataset for the complete and condensed dataset accuracy assessments. The results suggest

that the subsample of our data that we selected is itself important to the outcome of the results. It is impossible for us to tell which result is truly the correct one.

Comparing individual error matrices yielded even more unusual results. One possible interpretation of the data is that there is an interaction between clustering technique and effort. All classifications produced using divisive hierarchical clustering were significantly different from randomly produced classifications. The unsupervised classification, compared to the results provided by the k-means clustering technique, was also significantly different from a randomly produced classification. However, when half of the available data were used to produce a training signature, the accuracy actually declined, producing an insignificant result. Increasing the training pixels appeared to improve the accuracy of the classification. The results showed that the classifications produced using half of the available data for training signatures (k-means and divisive hierarchical clustering) are significantly different from each other. This result was expected. If the classification produced using training signatures based on k-means clustering was insignificant, and the classification produced using training signatures based on divisive hierarchical clustering was significant, then we expect these two classifications, when tested against each other, to be significantly different. We found this to be the case.

However, the unsupervised classification and supervised classification, which used training signatures based on the k-means clustering, were found not to be significantly different from the classification produced using half of the available data for a training signature. In other words, two classifications, by themselves significantly different from random, are not significantly different from another

classification that is itself not significantly different from random. This result is completely unexpected. Based on our prior discussion, one explanation is that we are witnessing a complication arising from a Type I error. It is possible that the classifications are not significant and yet we have rejected that null hypothesis for both individual classifications. Our kappa values ranged around 0.3. For most remote sensing applications, a kappa value of at least 0.6 is desired (Slocum, pers. comm.). As kappa increases, the thematic map produced using the classification becomes more reliable. Our kappa values appear to be relatively low. It is possible that the classifications are not significant even though we have found otherwise. It is also possible that we have committed a Type II error. We found that the classification produced using k-means clustering of the complete data set, with only half of available training data, was not significant. However, our condensed dataset, and our work in verifying the kappa value of 0.123, both yielded significant results. It is possible that we are not rejecting the null hypothesis when, in fact, we should be.

Reconciling the Two Interpretations

We have presented two very different interpretations of the statistical meaning of our data without actually considering their actual consequences on our research. While we recognize the statistical risk we took in performing the equivalent of multiple t-tests, a more suitable statistical option did not present itself. Remote sensing is a relatively new field; accuracy assessment is even newer. The true statistics needed to answer questions such as those we have posed in this research are not yet available, and developing them is outside our ability at the moment. Our

work highlights the need to exercise *extreme caution* when working with the DMSV, and perhaps such caution may need to be applied to other sensors used for remote sensing. The accuracy of remotely sensed imagery is supposed to increase as categories are combined into others (Congalton and Green 1999). We only worked with three classes in our research. Resource managers who use remote sensing to describe many categories need to be aware that accuracy falls off as the number of categories increases. Our experiment would have benefited by collecting more data in the field. This would have allowed us to provide stronger training signatures and more data for the accuracy assessment. Another potential problem we encountered involved our radiometric calibration. As Figure 3 shows, and as we discussed previously, we had to use the Elizabeth River as a dark target to correct for bias and gain in our data. This affected the correlation in our data. In addition, we might have benefited from not using the blue band, since interaction with atmospheric particles causes the blue band of the EMS to be scattered more than other visible or IR frequencies (Lillesand and Kiefer 1994). This interaction could have possibly affected the accuracy of our classifications by distorting the data, or expanding its variability, collected by the DMSV in the blue band.

As we described before, we have an uncovered an inconsistency in our data that can only be explained by assuming that either a Type I or Type II error exists in our data. Based on our subjective experience in the field, we are inclined to believe that a Type II error was committed, and that the classification is indeed significant. We may have failed to reject the null hypothesis when it is in fact false. This is based upon our observations that the k-means and divisive hierarchical clustering

procedures did yield classes that appear to be correct in an ecological sense. In addition, as we shall discuss below, the thematic maps produced from the remotely sensed imagery appear to have a structure that is different from one we would expect to be produced at random. In addition, there are differences in the user's and producer's accuracies that exist that may highlight where the confusion in the error matrices is being derived.

Interpretation of User's and Producer's Accuracy

Before we address how the remotely sensed imagery can contribute to the understanding of the ecology in Goose Creek, we need to be aware of the problems we will have in reading the thematic maps produced by the classification. We have already addressed the issues of significance affecting each classification. While kappa reveals whether or not a classification is significant and how it compares to other classifications, it does not reveal information about the classes themselves. User's and producer's accuracy serve to fill this void. The producer's accuracy measures how many pixels are left out of a correct classification (Congalton 1991). The class as shown on a thematic map may be much larger because other pixels from other classes are being incorrectly identified. From the perspective of the producer, high accuracy is achieved if the representation of that class still includes all true members of the class. User's accuracy approaches the thematic map from a different perspective. A user, working from the map provided by a producer, is more interested in how many pixels are included in a given class. If they point to a

particular pixel on the map, a user needs to know how certain they can be that it truly is a member of the class the map says it is.

While we did not perform a detailed statistical analysis of the user's and producer's accuracy, we believe our two examples (Tables 13 and 14) show that the user's and producer's accuracy can vary greatly. For example, in examining the k-means clustered data we see that class 1, which we identified as the *S. alterniflora / A. cannibina* composite, has a producer's accuracy of 30% but a user's accuracy of 70%. From the perspective of an analyst making the map, the accuracy is very low because points belonging to this class are being misclassified as other classes. However, from the perspective of a manager using the map, the accuracy is very high because they can be reasonably certain that if they use the map in the field, 70% of the areas identified as the *S. alterniflora / A. cannibina* composite are truly those areas.

If one class has a high user's accuracy but low producer's accuracy, it seems reasonable that another class must compensate. Points misclassified into other categories will appear as other classes on a thematic map, affecting the user's accuracy. The exact nature of the effect depends on how points were classified for each class and the number of classes. In our case, class 3, the *Scirpus* spp. dominated class, we see that the user's accuracy is much lower, 28%, while the producer's accuracy is 65%. Class 2, which is dominated by *P. australis*, has user's and producer's accuracies which are very similar, 63% and 73% respectively. Examining the error matrix suggests that the sample data for class 1 is being evenly misclassified into class 2 and 3. In fact, more pixels that truly belong to class 1 are being misclassified into other categories than are correctly classified. Out of the 63 sample

points collected that were included in class 1, only 19 were correctly classified as class 1. Twenty-four were classified as class 2, and 20 were classified as class 3. This might suggest that we made an error and that we should really be referring to the pixels in class 1 as class 2, where most of the pixels were correctly identified. The reason we have not done so is because the vast majority of sample points in class 2 were correctly identified as class 2. In actually creating the thematic map, we sought to maximize the producer's accuracy. Class 1 would still have had a low producer's accuracy if we had made this change, while serving to also dramatically lower the classification of *P. australis*. After creating the map, we use user's accuracy to understand how the map should be interpreted. In this case, we can be reasonably certain of class 1 and 2, but must be careful of what class 3 shows on the map.

Understanding why the classes are misclassified depends less on the actual ground cover values we collected as it does on the distribution of the brightness values and how the software performs the classification. Table 2 shows the mean values and standard deviation for the ground cover data we collected. Class 1 and 3 appear subjectively to resemble each other more closely than class 1 and class 2. Yet, while class 1 has a low user's accuracy and misclassified almost a third of its pixels into class 3, fewer pixels were misclassified into class 1 from class 3. It is possible that the shape of the cloud described by the brightness values for class 1 is much broader than that of class 2 and 3, which might explain the variation we are observing.

This same discussion also applies to the classification obtained from using divisive hierarchical clustering, except that the differences between the user's and

producer's accuracy appear to be slimmer than they were for the classification derived from k-means clustering. Class 1, the composite of *Scirpus* spp., *S. cynosuroides*, *A. cannibina*, and *Typha* spp has the lowest user's accuracy. It appears that class 2, largely comprised of *S. alterniflora* and No Cover, is misclassified into the other two classes relatively evenly. Class 3, *P. australis*, is heavily misclassified into class 1. Again, the distribution of the brightness values across the four bands and how the computer performed the classification account for the accuracy. While we did not observe it, we still suspect that adding extra data to the training signatures may improve accuracy. However, the cost required to do so may be prohibitive, based upon our findings that doubling the number of sample points appeared to have no effect.

Ecological Interpretation of Remotely Sensed Data

Despite the problems we have described with our data, we can still cautiously attempt to interpret our results. Remote sensing has the potential to identify the distribution of vegetation within a given scene. One could accomplish the same task by examining our groundcover data for specific areas within the scene. However, compiling data increases the risk that dissimilar areas will be pooled together. In other words, we may be putting together samples that are not related to each other. Remote sensing separates each pixel into a given class so that the risk is lessened, as long as one is certain that the classes being used are distinct. As we have described, the use of only three categories to separate Goose Creek vegetation may not have been the most appropriate choice. However, we believe that we can still draw some

general conclusions about patterns in the wetland, even with the inaccuracies we have in our data.

The most obvious trend is that *P. australis*, which appeared to separate out as a separate class using both the k-means and divisive hierarchical clustering tools, dominates the center of the marsh. The other classes are all composites of other species. It would be very difficult to discern that *P. australis* has had such an influence on Goose Creek with traditional methods, such as those used in our study, other than outright surveying. For one species to dominate a marsh in this way, where conditions might otherwise suggest a more diverse habitat, is unusual. It supports the view of *P. australis* as an invasive species that is highly adapted to outcompeting other species (Marks and Randall 1994). The shape of the distribution appears to be approximately the same throughout each of the classifications we have produced. The results from the imagery do not explain why *P. australis* has the observed distribution. *P. australis* was not one of the original species planted in the Goose Creek wetland when it was created (Barnard *et al.* 1997). It is unknown what conditions have emerged over the twenty years since the wetland was created that have allowed *P. australis* to invade and thrive.

Our work only shows a rough outline of the distribution of *P. australis* when it is the dominant vegetation in a given area of the marsh. The clear signal that we observed with our research, however, suggests that remote sensing might be useful for monitoring *P. australis* invasions over time. If one assumes that the brightness values describing *P. australis* will not change over time (i.e. the signal this year is the same as the signal that would be obtained next year), one could use our data to train

imagery collected during subsequent years, with all other vegetation falling into the 'Not *P. australis*' class. *P. australis* is a hardy species. If the data is collected at approximately the same time and under similar conditions, it is unlikely that another species would have replaced *P. australis* in Goose Creek. One must still exercise caution. The marsh is changing over time. We do not know if our observations show *P. australis* invading into further areas of Goose Creek, or declining from some greater coverage it may have had in past years. Collecting time series data, or examining past vegetation studies, could help resolve this question. It is possible that conditions of water salinity or soil stability are restricting the movement of *P. australis* throughout Goose Creek, or aiding its invasion. Further research is required in order to answer these questions. Nonetheless, the imagery is useful in showing the extent of invasion by *P. australis* in the Goose Creek wetland. As the only clear dominant in any given class, it is apparent that this species is diminishing the value of Goose Creek as a mitigation bank. While *P. australis* is considered a native part of European wetlands, it is regarded along the eastern shore of the United States as an exotic that decreases the value of wetlands it invades (Marks and Randall 1994). As such, its presence in Goose Creek lowers the value of the system and may make it unusable as a mitigation bank.

As more classes were created, the signal from *P. australis* appeared to get lost because it is incorporated into other classes as a component. In other words, when classes are combined, the signal due to *P. australis* appeared to become clearer. For a manager who is only interested in following the spread of *Phragmites*, our work appears to suggest that fewer classes might be more successful at monitoring the

spread of *P. australis* because less work has to be performed for accuracy assessment. The ecology of *P. australis* as an invasive species indicates that it is going to be found in homogenous patches such as those found in our thematic maps. As a result, a classification consisting of the classes '*P. australis*' and 'not *P. australis*' might produce needed results for monitoring this species. One potential problem for monitoring *P. australis* using only remote sensing is that we do not know the 'critical mass' at which the plant becomes visible to the sensor. For someone interested in monitoring for the first sign of the species, remote sensing may not be an effective tool. However, it could be useful for monitoring widespread populations to see if eradication measures are effective over time.

The other classes, due to their composite nature, are more difficult to interpret. We noted previously that the accuracy of the unsupervised classifications was not significantly different from the supervised classifications when tested against the data clustered using both the k-means and divisive hierarchical clustering. The unsupervised image will therefore yield two slightly different interpretations, as the classes derived from k-means and divisive hierarchical clustering are different, even though the thematic map produced are the same.

K-Means

While the accuracy of the unsupervised classification was not significantly different from the supervised classifications we produced, there does appear to be a difference in the interpretation of the images. The unsupervised image (Figure 6) appears to suggest that the *S. alterniflora* / *A. cannibina* mix (Class 1) surrounds *P.*

australis (Class 2) on the northern side of Goose Creek. The *Scirpus* spp. dominated class (Class 3), on the other hand, appears to surround *P. australis* on the southern side of the marsh. This trend is not apparent in the supervised classifications. Figures 7-10 show the thematic maps produced using the k-means clustering technique. Class 1 and 3 appear to be more closely intertwined, though they still appear in discrete patches. Based on our previous discussion of the user's and producer's accuracy, Class 3 is more often than not, not what we think it is. In other words, in most cases (more than 50%) Class 3 has been misclassified because other pixels from classes 1 and 2 are being included into it. Based on the distribution of *P. australis*, we hypothesize that the error is not random and that pixels are being misclassified based on the actual distributions of the classes in the field. In other words, pixels at the interface between Class 1 and 3 on the image are probably Class 1 and not Class 2. Pixels at the interface between class 2 and 3 that are classified as Class 3 are probably *P. australis.*

This assumption is based on what we know about the ecology of the system. *P. australis* clearly appears clumped in the middle of the marsh. Goose Creek contains species adapted for conditions ranging from saltwater to brackish. It would appear that *P. australis* is not able to compete with other species in other parts of the marsh, or it has not yet had time to invade the rest of the marsh. While it is possible and does occur in patches of Goose Creek, we believe that a pixel reflecting Class 3 on the northerly and southerly edges of Goose Creek is more likely going to be represent *S. alterniflora* / *A. cannibina* than *P. australis*, because the ecology of these different classes seems to suggest that it is more likely. Salinity and other

measurements from the ground could be used to confirm this new hypothesis. In either case, it is more reasonable to assume that the low user's accuracy of Class 3, if only for the one k-means clustered error matrix that we assessed (the others are available in the Appendix), is caused by variability in brightness values between the interface of two classes as one class became another, than it is to assume that a stray patch of another class was present.

As presented in Figures 7-10, the thematic maps produced from the k-mean clustered training signatures, do not appear to immediately reveal obvious features about the Goose Creek marsh. Pixels belonging to the same class are clumped together, which is not unexpected. The classifications produced using half or all available training data are similar but do have some differences (Figures 7-14), as do classifications performed as verification of the original results when the low kappa value was first produced from the data (Figures 15 and 16). Imagery produced using only half of all available training data appears to show Class 1 (*S. alterniflora / A. cannibina*) running throughout more of Goose Creek in small patches. It is possible that spatial diversity tests, such as those suggested by Robinove (1986), might reveal a pattern to our subjective observations and suggest further ways in which remotely sensed data can be analyzed. The most obvious ecological feature is that Classes 1 and 3 are not as homogeneous as Class 2. The boundaries between *S. alterniflora, A. cannibina,* and *Scirpus* are not as precisely defined as the *Phragmites* boundary. We noted previously the high variability we observed for classes 1 and 3. This may be expressing itself through the amalgam of pixels we are observing in the imagery. In our experience on the ground, for example, *A. cannibina* was found most prominently

in the middle of the marsh at the interface between *P. australis, S. alterniflora,* and *Scirpus*. We did not believe it to be a dominant species in the northerly or southerly region of Goose Creek. However, as we only had three classes available *A. cannibina* had to be included into a larger cluster . Our cluster data noted that *A. cannibina* was most commonly found with *S. alterniflora*, but it still appeared in Class 3 with *Scirpus* spp. As a result, we can not draw definitive conclusions about this species from the thematic map. It would be equally difficult to draw conclusions about other species

The classifications produced for verification purposes relied on a different subset of the original data for training. As a result, the thematic maps do not exactly resemble those of their counterparts produced the first time. However, they are very similar. The most notable difference appears to be that a patch of *P. australis* in the northerly side of Goose Creek has been reclassified as *S. alterniflora / A. cannibina* (Class 1). Other differences appear to be minor. Despite this, our work showed the first image to be not significantly different from random, while the image produced for the verification is significant. This serves as visual demonstration of our belief that we have either Type I or Type II errors in our data, unless our maps border at the significant/not significant level. Given our low Kappa values, such a conclusion is not out of the question.

Divisive Hierarchical

The classifications produced using the divisive hierarchical clustering algorithm are similar to those produced using the k-means but there are some

differences suggesting a slightly different ecological interpretation. The differences center not on the differences between individual thematic maps produced from the divisive hierarchical clustered dataset. Instead, the focus is on the differences in thematic maps treated the same except for how the data were clustered. *P. australis*, class 3, appears as a distinct, very solid, class located principally in the center of the marsh in each thematic map (Figures 11-14). This feature is the same as that observed in the k-mean clustered data thematic maps. In addition, there appear to be only minor differences between thematic maps produced using signatures incorporating half or all of all available data. The maps produced using the complete dataset, and the condensed data, also appear very similar. The minor differences noted most likely can be attributed to random variations derived from which subset of the data was used.

The general feature that appears to run through each of the divisive hierarchical clustered data thematic maps is that Class 1 (*Scirpus* spp., *S. cynosuroides*, *A. cannibina*, and *Typha*) is separated from *P. australis* on the northerly side of Goose Creek by a buffer zone created by Class 2 (*S. alterniflora*, No Cover, *A. cannibina* less prominent), but the reverse is true on the southerly side of the marsh. In the k-means clustered data, the buffer zone is less obvious. Again, we must be wary of the user's accuracy, which is low for Class 1. It appears that pixels from class 2 and class 3 are being included into class 1 with roughly equal frequency. As we stated before for the k-means data, we believe that misclassifications result from the edge effect created at the interface of classes, rather than a random misclassification.

The reason for the difference in the images produced using the divisive hierarchical and k-means clustering technique is difficult to discern. The central difference between k-means and divisive hierarchical clustered data is that the *Scirpus* spp. dominated class in the divisive hierarchical clustered data appears to be more inclusive, also being dominated by *S. cynosuroides*, *A. cannibina* (more dominantly than in k-means Class 3), and *Typha* spp. The *P. australis* class appears to have the same general shape throughout all of the thematic maps, so one might hypothesize that the distribution of rare species throughout the marsh is enough to cause the change. However, the condensed dataset thematic maps we produced have very similar results to the thematic maps produced using the complete dataset. It is also possible that the subset of the data used for accuracy assessment and signature training affected the classifications, but the reverification tests we performed on the k-mean clustered data produced similar results to the original data (kappa values notwithstanding).

Based on the elimination of what it is not, the only option we see is that subtle differences in the clusters led to the changes. It is difficult to decipher our data simply because Goose Creek is very heterogenous. Aside from *P. australis*, the clusters are composed of a large number of species with similar attributes. The variance within each cluster for the species that comprise it is large (Table 2). We believe it likely that we have uncovered, in a sense, different homogeneous patches of heterogeneous species. Further work within each of these patches could yield answers as to whether there is an actual difference in vegetation, and if so what is the ecological basis for the difference. Based on this, and also on the kappa values we

obtained from our data, we are extremely hesitant to draw definitive conclusions from our data regarding the distribution of species within the marsh. If one is interested in simply monitoring the invasion of *P. australis*, our data should be classified to only two classes: '*P. australis*' and 'not *P. australis*.' We believe our data would be suitable for this purpose. For others interested in ecological surveys, our data need to be broken into further classes, which would require additional fieldwork if it was to be analyzed using remote sensing techniques.

CONCLUSIONS

We have attempted in our research to explore the usefulness of the DMSV as a tool for wetlands research by comparing it to ground-based, 'traditional,' techniques. As the demand for resources grows along the coastline, the need to monitor, study, and manage those natural resources becomes more important. Remote sensing has shown promise as an additional tool in the scientist's and manager's arsenal. We had hoped to determine whether one technique was better than the other for assessing the ecology of wetlands, and in particular Goose Creek, the mitigation bank that we studied. We were not able to answer this question conclusively. Instead, we found that remote sensing holds promise as a tool for monitoring and assessment. While it can provide additional information, the traditional techniques should still be used as a baseline for any wetland monitoring. While some studies (e.g. Hinson *et al.* 1994) have shown remote sensing to be a good tool for classifying wetlands, our work suggests that extreme caution is still needed and that there are many statistical questions which need to be answered before remote sensing should truly become a common tool.

One important result we obtained is that it is not possible to obtain some of the most common ecological measures from remotely sensed data. We had initially believed that it was possible to calculate diversity, such as Simpson's index, from remotely sensed data. We were interested in knowing whether a correlation existed between the diversity calculated from data collected on the ground and the remotely

sensed imagery. We found that it was not possible to calculate diversity as applied to traditional techniques. However, there may be statistics that can be calculated using the imagery that can not be obtained from ground data, such as spatial diversity. Such measures can provide additional information about an area of interest, perhaps leading to further ecological understanding by providing information on how different features of the environment interact. For example, calculating spatial diversity could have yielded information about the interface between *P. australis* and *S. alterniflora* patches in Goose Creek. We did not explore other correlations that might exist. For example, PCA breaks apart the variance in the ecological data. It is possible that this variance could be correlated to the variance in the brightness values obtained from the remotely sensed imagery. We hypothesize that if classes are composed of a given number of species, then there will be a relationship between the variation of that given class and the variation in the brightness values for that class. Such a hypothesis, if supported by future research, could be used to minimize ground-truthing in the future.

The composition of our classes appears to be highly variable. With only three classes available for analysis, assessing a heterogeneous marsh such as Goose Creek proved to be very difficult. The choice of clustering technique does not appear to affect the accuracy of the classified maps. However, differences did exist, depending on whether k-means or divisive hierarchical clustering was applied. The differences appeared to be minor and did not change our view of the processes occurring in Goose Creek. Attempting to condense the dataset by focusing on dominant species had no effect on the final outcome, suggesting that in the future we might be able to

ignore rare species in remote sensing research. Of course, the corollary is that remote sensing may not be an effective tool for studying rare species.

It was interesting that *P. australis* emerged as a lone dominant in both clustering techniques. The *P. australis* species appeared to demonstrate its traits as an exotic very well in our marsh. Remote sensing appears to offer potential for monitoring future invasions of *P. australis* by taking advantage of the species' ability to outcompete others. The large variability we encountered in the other classes was not surprising, but does make ecological assessment difficult beyond broad generalizations about Goose Creek's structure.

Our research focused largely on the accuracy of remote sensing. A key finding was that accuracy of the thematic maps we produced did not appear to change depending on how much effort was put into obtaining data for training signatures. Unsupervised classifications were not significantly different from classifications produced with a season's worth of data. It is important to note that we would not have known what the classes were if the fieldwork component had not been completed. However, if the composition of the classes in Goose Creek remains constant, the ecosystem could be monitored using our data in the future. The temporal aspect of accuracy was not explored in our research but remains open as a future question for exploration. While our research showed supervised classification is not necessary, we do not suggest that this finding should be applied to other studies. Our classifications all produced low kappa values. While the maps we produced were significant, we believe higher kappa scores could have been achieved had we obtained enough data to allow more classes to be obtained with our data.

While accuracy is supposed to increase as the number of classes decreases (Congalton and Green 1998), we believe the variability caused by combining classes together may have actually caused a decline in the overall accuracy. However, if classes had been combined to provide the *P. australis* / not *P. australis* distinction, the homogeneity of that species might have improved overall accuracy.

While it appears that there was no difference between unsupervised and supervised classifications, our work probably suffers from Type I and Type II errors because, to our knowledge, the statistics to properly analyze our work do not exist. At best, we recommend extreme caution in using our work. We have greater trust in our ground data than in the thematic maps we produced. We recommend using the maps only as an introduction to Goose Creek. Further work remains to be done before we would trust remote sensing to be used as a tool for monitoring the mitigation bank, but it does hold great promise. Remote sensing has been shown to be extremely versatile in a number of other situations (Lillesand and Kiefer 1994). We do not see anything so extraordinary about wetlands that they can not also be studied with this tool. However, the heterogeneity of wetland ecosystems and the needs of resource managers do pose unique challenges that require further statistical refinement before remote sensing can be fully utilized.

Table 1. Mean of All Species Vegetative Cover for K-Means and Divisive Hierarchical Clustering

Species	KM - Mean 1	KM - Mean - 2	KM - Mean - 3	DH - Mean 1	DH - Mean 2	DH - Mean 3
No Cover	30.13	10.53	18.43	16.45	35.68	11.52
Spartina alterniflora	32.52	1.50	11.83	12.22	40.89	1.96
Spartina cynosuroides	4.10	0.06	5.83	7.18	2.96	0.08
Spartina patens	0.00	0.04	0.00	0.00	0.00	0.04
Phragmites australis	5.90	81.32	1.13	2.09	3.76	77.98
Scirpus spp.	4.92	1.87	49.88	34.00	3.68	1.79
Amaranthus cannibina	11.60	1.66	5.83	14.55	7.69	1.75
Atriplex patula	1.13	0.23	0.30	1.25	0.76	0.24
Pluchea purperescens	0.77	0.28	0.13	0.13	0.61	0.56
Typha spp.	3.73	1.19	2.48	6.03	1.26	1.52
Distichlis spicata	3.09	1.00	2.88	4.78	1.08	1.62
Samolus parviflorus	0.52	0.02	0.00	0.03	0.14	0.42
Lythrum linaere	0.32	0.04	0.50	0.45	0.09	0.21
Juncus spp.	0.00	0.00	0.45	0.27	0.00	0.00
Solidagospp.	0.00	0.02	0.00	0.00	0.00	0.02
Aster spp.	1.13	0.25	0.38	0.55	1.27	0.28
Iva frutescens	0.16	0.00	0.00	0.03	0.17	0.02

KM = K-Means clustered data
DH = Divisive Hierarchical clustered data

Table 2. Mean and Standard Deviation for Vegetative Cover for K-Means and Divisive Hierarchical Clustering

k-means data	KM Mean 1	Std. Dev. 1	KM Mean 2	Std. Dev. 2	KM Mean 3	Std. Dev. 3
No Cover	30.13	21.85	10.53	9.03	18.43	9.39
Spartina alterniflora	32.52	24.52	1.50	3.85	11.83	12.38
Spartina cynosuroides	4.10	11.80	0.06	0.33	5.83	10.21
Spartina patens	0.00	0.00	0.04	0.46	0.00	0.00
Phragmites australis	5.90	10.95	81.32	16.14	1.13	4.31
Scirpus spp.	4.92	7.68	1.87	3.15	49.88	20.98
Amaranthus cannibina	11.60	16.14	1.66	3.85	5.83	10.10
Atriplex patula	1.13	4.12	0.23	1.25	0.30	1.34
Pluchea purperescens	0.77	4.16	0.28	0.94	0.13	0.40
Typha spp.	3.73	14.12	1.19	3.33	2.48	6.95
Distichlis spicata	3.09	12.32	1.00	4.54	2.88	12.02
Samolus parviflorus	0.52	4.51	0.02	0.18	0.00	0.00
Lythrum linaere	0.32	2.01	0.04	0.33	0.50	3.16
Juncus spp.	0.00	0.00	0.00	0.00	0.45	2.85
Solidago spp.	0.00	0.00	0.02	0.18	0.00	0.00
Aster spp.	1.13	2.55	0.25	0.86	0.38	2.37
Iva frutescens	0.16	1.37	0.00	0.00	0.00	0.00

Divisive Hierarchical	DH Mean 1	Std. Dev. 1	DH Mean 2	Std. Dev. 2	DH Mean 3	Std. Dev. 3
No Cover	16.45	8.85	35.68	22.87	11.52	9.96
Spartina alterniflora	12.22	11.07	40.89	23.72	1.96	4.77
Spartina cynosuroides	7.18	13.70	2.96	9.67	0.08	0.41
Spartina patens	0.00	0.00	0.00	0.00	0.04	0.45
Phragmites australis	2.09	5.65	3.76	7.71	77.98	19.79
Scirpus spp.	34.00	25.83	3.68	7.02	1.79	3.06
Amaranthus cannibina	14.55	19.48	7.69	10.61	1.75	3.91
Atriplex patula	1.25	4.82	0.76	2.72	0.24	1.23
Pluchea purperescens	0.13	0.46	0.61	3.30	0.56	3.23
Typha spp.	6.03	18.98	1.26	4.31	1.52	3.96
Distichlis spicata	4.78	16.79	1.08	5.88	1.62	6.33
Samolus parviflorus	0.03	0.24	0.14	1.37	0.42	4.37
Lythrum linaere	0.45	2.57	0.09	0.84	0.21	1.81
Juncus spp.	0.27	2.20	0.00	0.00	0.00	0.00
Solidago spp.	0.00	0.00	0.00	0.00	0.02	0.18
Aster spp.	0.55	2.33	1.27	2.74	0.28	0.85
Iva frutescens	0.03	0.24	0.17	1.58	0.02	0.27

Table 3. Complete and Condensed Cover Comparison for K-Means Data

K-mean Clustering-Percent Cover	Mean 1 Entire	Mean 1 Condensed	Mean 2 Entire	Mean 2 Condensed	Mean 3 Entire	Mean 3 Condensed
No Cover	30.13	30.08	10.53	10.75	18.43	18.43
Spartina alterniflora	32.52	32.78	1.50	1.49	11.83	11.83
Spartina cynosuroides	4.10	4.14	0.06	0.06	5.83	5.83
Phragmites australis	5.90	5.63	81.32	80.97	1.13	1.13
Scirpus spp.	4.92	4.94	1.87	1.87	49.88	49.88
Amaranthus cannibina	11.60	11.66	1.66	1.68	5.83	5.83
Typha spp.	3.73	3.73	1.19	1.21	2.48	2.48
Aster spp.	1.13	1.13	0.25	0.25	0.38	0.38

Table 4. Complete and Condensed Cover Comparison for Divisive Hierarchical Data

Divisive Hierarchical – Percent Cover	Mean 1 Entire	Mean 1 Condensed	Mean 2 Entire	Mean 2 Condensed	Mean 3 Entire	Mean 3 Condensed
No Cover	16.45	16.45	35.68	35.68	11.52	11.52
Spartina alterniflora	12.22	12.22	40.89	40.89	1.96	1.96
Spartina cynosuroides	7.18	7.18	2.96	2.96	0.08	0.08
Phragmites australis	2.09	2.09	3.76	3.76	77.98	77.98
Scirpus spp.	34.00	34.00	3.68	3.68	1.79	1.79
Amaranthus cannibina	14.55	14.55	7.69	7.69	1.75	1.75
Typha spp.	6.03	6.03	1.26	1.26	1.52	1.52
Aster spp.	0.55	0.55	1.27	1.27	0.28	0.28

Table 5. Producing Five Clusters with K-means

K-mean – Percent Cover	Mean - 1	Mean - 2	Mean - 3	Mean - 4	Mean - 5
No Cover	9.41	44.25	21.26	13.75	18.43
Spartina alterniflora	0.95	14.71	44.00	3.75	11.83
Spartina cynosuroides	0.04	2.62	4.99	0.00	5.83
Spartina patens	0.05	0.00	0.00	0.00	0.00
Phragmites australis	83.55	17.21	2.13	2.50	1.13
Scirpus spp.	1.87	2.42	6.48	2.00	49.88
Amaranthus cannibina	1.50	2.88	17.36	0.75	5.83
Atriplex patula	0.14	0.83	1.45	0.00	0.30
Pluchea purperescens	0.26	1.60	0.23	0.00	0.13
Typha spp.	1.22	2.88	0.26	76.25	2.48
Distichlis spicata	0.73	7.81	0.26	0.00	2.88
Samolus parviflorus	0.02	1.27	0.00	0.00	0.00
Lythrum linaere	0.05	0.77	0.00	0.00	0.50
Juncus spp.	0.00	0.00	0.00	0.00	0.45
Solidago spp.	0.02	0.00	0.00	0.00	0.00
Aster spp.	0.21	0.75	1.39	0.50	0.38
Iva frutescens	0.00	0.06	0.19	0.50	0.00

Table 6. Producing 5 Clusters with Divisive Hierarchical Clustering

Divisive Hierarchical – Percent Cover	Mean - 1	Mean - 2	Mean - 3	Mean - 4	Mean - 5
No Cover	16.62	20.87	11.52	57.89	13.75
Spartina alterniflora	12.76	52.69	1.96	23.19	3.75
Spartina cynosuroides	7.63	3.28	0.08	2.47	0.00
Spartina patens	0.00	0.00	0.04	0.00	0.00
Phragmites australis	2.06	3.56	77.98	4.06	2.50
Scirpus spp.	36.03	4.19	1.79	2.92	2.00
Amaranthus cannibina	15.43	10.44	1.75	3.56	0.75
Atriplex patula	1.33	1.20	0.24	0.08	0.00
Pluchea purperescens	0.14	0.89	0.56	0.19	0.00
Typha spp.	1.57	0.65	1.52	2.17	76.25
Distichlis spicata	5.08	0.37	1.62	2.14	0.00
Samolus parviflorus	0.03	0.00	0.42	0.36	0.00
Lythrum linaere	0.48	0.00	0.21	0.22	0.00
Juncus spp.	0.29	0.00	0.00	0.00	0.00
Solidago spp.	0.00	0.00	0.02	0.00	0.00
Aster spp.	0.56	1.61	0.28	0.75	0.50
Iva frutescens	0.00	0.28	0.02	0.00	0.50

Table 7. Importance Value Results for K-Means and Divisive Hierarchical Data – Three Categories

K-means	Cluster 1	Cluster 2	Cluster 3
Spartina alterniflora	71.47	10.26	33.70
Spartina cynosuroides	13.42	1.25	24.74
Spartina patens	0.00	0.34	0.00
Phragmites australis	17.83	125.49	4.58
Scirpus spp.	20.08	23.99	93.14
Amaranthus cannibina	34.21	13.69	19.94
Atriplex patula	6.66	1.74	1.97
Pluchea purperescens	4.76	5.05	3.35
Typha spp.	9.91	9.32	7.83
Distichlis spicata	7.39	3.78	6.72
Samolus parviflorus	1.66	0.31	0.00
Lythrum linaere	1.60	0.64	1.41
Juncus spp.	0.00	0.00	1.35
Solidago spp.	0.00	0.31	0.00
Aster spp.	10.08	3.83	1.26
Iva frutescens	0.91	0.00	0.00

Divisive Hierarchical	Cluster 1	Cluster 2	Cluster 3
Spartina alterniflora	33.86	93.45	14.90
Spartina cynosuroides	24.40	10.64	1.07
Spartina patens	0.00	0.00	0.04
Phragmites australis	7.20	14.73	120.81
Scirpus spp.	66.75	17.10	19.58
Amaranthus cannibina	33.23	29.75	11.74
Atriplex patula	5.35	6.16	2.22
Pluchea purperescens	2.72	5.22	6.49
Typha spp.	11.49	5.51	10.49
Distichlis spicata	8.71	3.81	3.78
Samolus parviflorus	0.46	0.58	0.48
Lythrum linaere	1.82	0.49	0.24
Juncus spp.	0.75	0.00	0.00
Solidago spp.	0.00	0.00	0.51
Aster spp.	2.80	11.93	7.14
Iva frutescens	0.46	0.61	0.51

Table 8. Importance Value Results for K-Means and Divisive Hierarchical Data – Five Categories

K-means	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Spartina alterniflora	8.87	43.86	86.07	19.73	33.70
Spartina cynosuroides	0.69	10.69	15.35	0.00	24.74
Spartina patens	0.38	0.00	0.00	0.00	0.00
Phragmites australis	128.05	47.34	7.80	18.28	4.58
Scirpus spp.	25.19	12.84	24.31	17.70	93.14
Amaranthus cannibina	13.06	19.17	42.83	8.56	19.94
Atriplex patula	1.13	5.98	7.73	0.00	1.97
Pluchea purperescens	5.18	7.86	3.04	0.00	3.35
Typha spp.	9.49	13.67	0.72	119.18	7.83
Distichlis spicata	3.08	20.49	1.11	0.00	6.72
Samolus parviflorus	0.35	4.27	0.00	0.00	0.00
Lythrum linaere	0.70	3.88	0.00	0.00	1.41
Juncus spp.	0.00	0.00	0.00	0.00	1.35
Solidago spp.	0.35	0.00	0.00	0.00	0.00
Aster spp.	3.49	9.34	10.39	8.27	1.26
Iva frutescens	0.00	0.60	0.64	8.27	0.00

Divisive Hierarchical	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Spartina alterniflora	34.76	97.25	10.79	83.65	19.73
Spartina cynosuroides	25.90	10.96	1.39	10.63	0.00
Spartina patens	0.00	0.00	0.30	0.00	0.00
Phragmites australis	6.55	12.45	120.85	20.11	18.28
Scirpus spp.	69.90	16.65	22.28	18.35	17.70
Amaranthus cannibina	34.79	31.95	13.93	24.63	8.56
Atriplex patula	5.67	7.77	1.83	3.06	0.00
Pluchea purperescens	2.89	5.67	5.31	4.27	0.00
Typha spp.	4.60	1.96	10.28	12.76	119.18
Distichlis spicata	9.26	1.60	5.21	8.89	0.00
Samolus parviflorus	0.49	0.00	1.25	1.81	0.00
Lythrum linaere	1.93	0.00	1.28	1.48	0.00
Juncus spp.	0.80	0.00	0.00	0.00	0.00
Solidago spp.	0.00	0.00	0.28	0.00	0.00
Aster spp.	2.48	12.83	4.73	10.35	8.27
Iva frutescens	0.00	0.92	0.29	0.00	8.27

Table 9. Accuracy Assessment- Divisive Hierarchical Classification, All Available Training Data, Complete Dataset

Divisive Hierarchical Classification, All Available Training Data, Complete Dataset				
	Class 1	**Class 2**	**Class 3**	**Row Totals**
Class 1	23	12	19	54
Class 2	5	19	7	31
Class 3	5	14	37	56
Column Totals	33	45	63	141
Kappa	0.3365209			
Theta One	0.5602837			
Theta Two	0.3372567			
Theta Three	0.3947488			
Theta Four	0.5173484			
Kappa Var.	0.0040608			
Z (one matrix)	5.2808497			

Table 10. Results of the Z-test for Each of the Error Matrices Computed

	Kappa	Kappa Variance	Z
Unsupervised, k-means, complete	0.273	0.005	4.063
Unsupervised, Div. Hier., complete	0.294	0.004	4.449
Comkall	0.294	0.004	4.532
Comkhalf	0.123	0.004	1.897
Comdvhrall	0.337	0.004	5.281
Comdvhrhalf	0.312	0.004	4.926
Ecodivallsup	0.291	0.004	4.552
Ecodivhalfsup	0.270	0.004	4.136
Ecokallsup	0.303	0.004	4.701
Ecokhalfsup	0.278	0.004	4.315

Com = Dataset clustered using all species
Eco = Dataset clustered using only dominant species
K = k-means clustered data
Dvhr = divisive hierarchical clustered data
All = All available training data used to create signatures
Half = Half of available training data used to create signatures

Table 11. Verification of Results

VERIFICATION	Kappa	Kappa Var	Z
Comkall	0.338	0.004	5.321
Comkhalf	0.382	0.004	6.135

Table 12. Error Matrices Compared Against Each Other

	Z-score	Alternate Z-Score
Unsup., k-means v. Unsup. Dvhr	0.228	
Unsup., k-mean v. Comkall	0.227	
Unsup., k-mean v. Comkhalf	1.608	1.195
Unsup., dvhr v. Comdvhrall	0.461	
Unsup., dvhr v. Comdvhrhalf	0.194	
Comkall v. Comkhalf	1.867	0.981
Comdvhrhall v. Comdvhrhalf	0.274	
Ecodivallsup v. Ecodivhalfsup	0.228	
Ecokallsup v. Ecokhalfsup	0.272	
Comkall v. Comdvhrall	0.470	
Comkhalf v. Comdvhrhalf	2.089	0.789
Ecodivallsup v. Ecokallsup	0.125	
Ecodivhalfsup v. Ecokhalfsup	0.082	

Com = Dataset clustered using all species
Eco = Dataset clustered using only dominant species
K = k-means clustered data
Dvhr = divisive hierarchical clustered data
All = All available training data used to create signatures
Half = Half of available training data used to create signatures

Table 13. User's and Producer's Accuracy Estimates – K-Means

K-Means	Class 1	Class 2	Class 3	Row Totals
Class 1	19	3	5	27
Class 2	24	43	2	69
Class 3	20	13	13	46
Column Totals	63	59	20	142
Overall Accuracy		0.53		
User's Accuracy			**Producer's Accuracy**	
Class 1	0.70		Class 1	0.30
Class 2	0.62		Class 2	0.73
Class 3	0.28		Class 3	0.65
Conditional Kappa Values				
Class 1	0.47			
Class 2	0.36			
Class 3	0.17			

Table 14. User's and Producer's Accuracy Estimates – Divisive Hierarchical

Divisive Hierarchical	**Class 1**	**Class 2**	**Class 3**	**Row Totals**
Class 1	23	12	19	54
Class 2	5	19	7	31
Class 3	5	14	37	56
Column Totals	33	45	63	141
Overall Accuracy		0.56		
User's Accuracy			**Producer's Accuracy**	
Class 1	0.43		Class 1	0.70
Class 2	0.61		Class 2	0.42
Class 3	0.66		Class 3	0.59
Conditional Kappa Values				
Class 1	0.25			
Class 2	0.43			
Class 3	0.39			

Figure 1. Location of Study Site in Suffolk County, Virginia (from Barnard, Jr. *et al.* 1997).

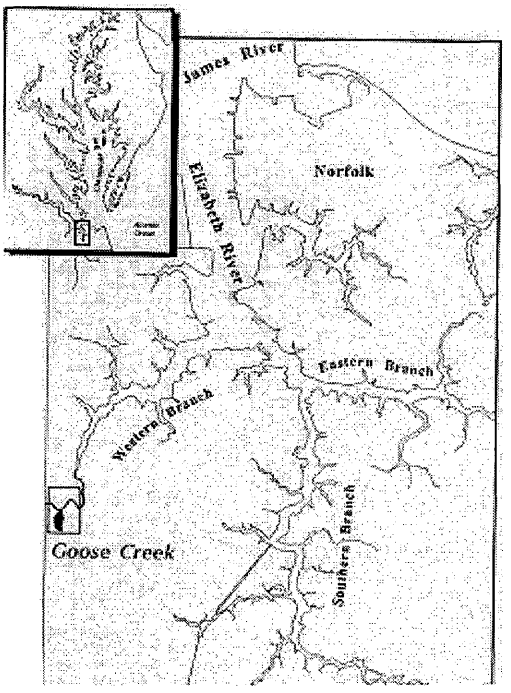


Figure 2. Outline of Study Site Highlighting Locations of Transect Pipes Used to Locate Sample Plots.

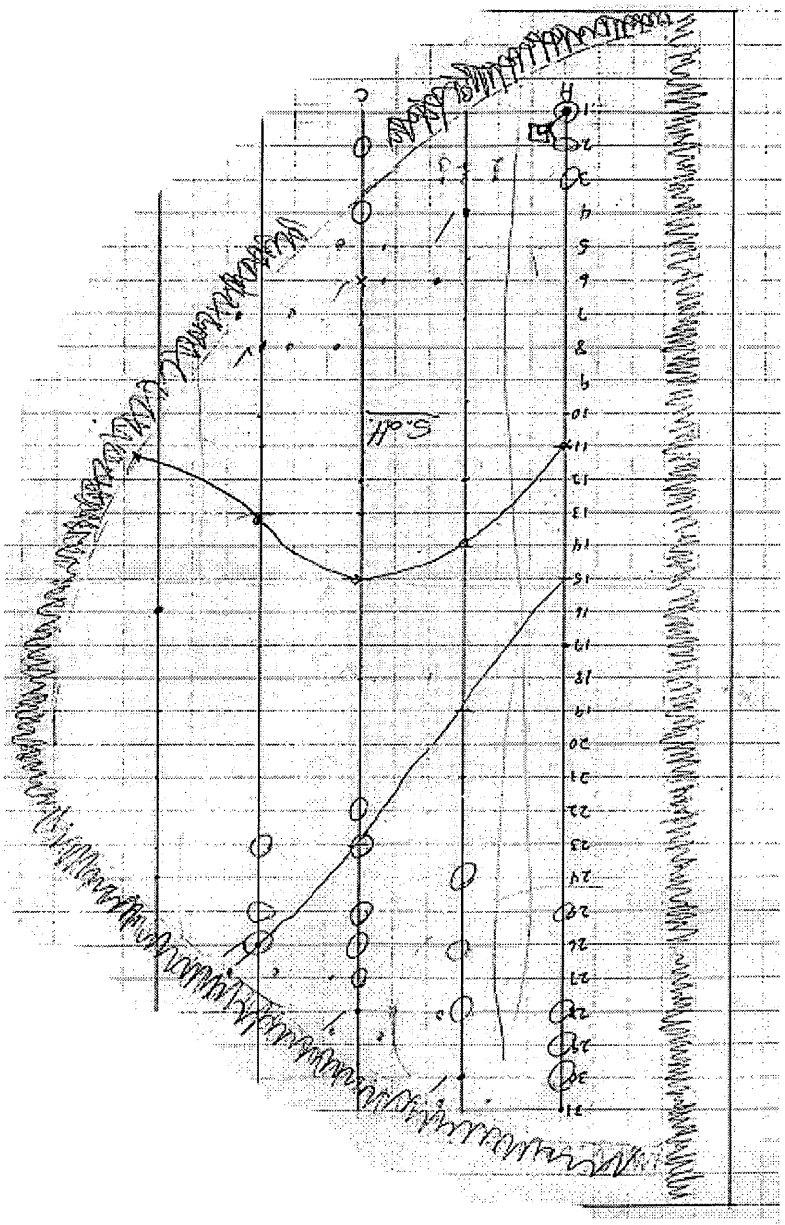
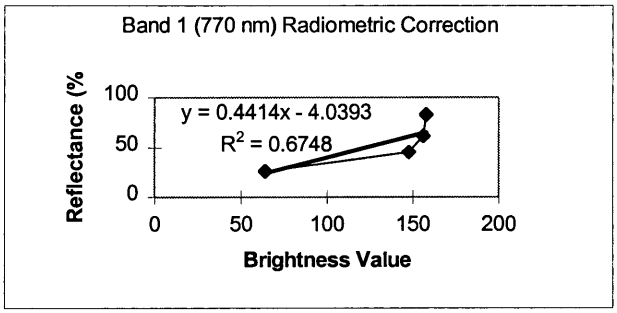
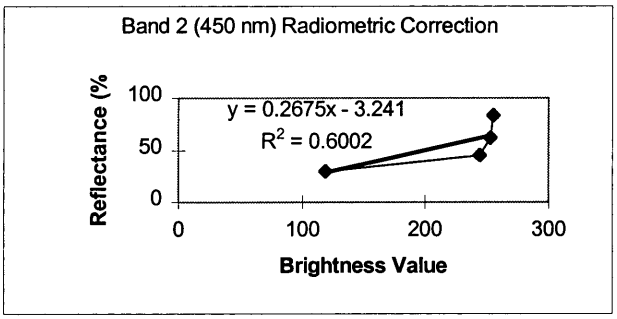
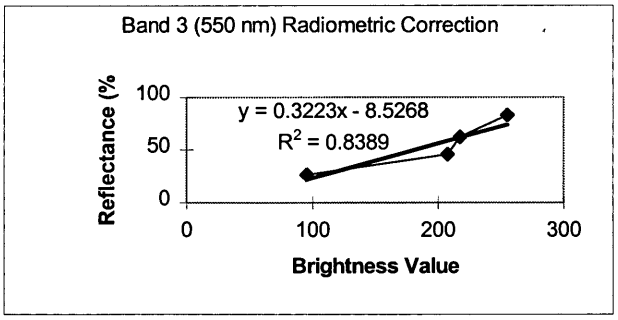


Figure 3. Radiometric Correction Results.







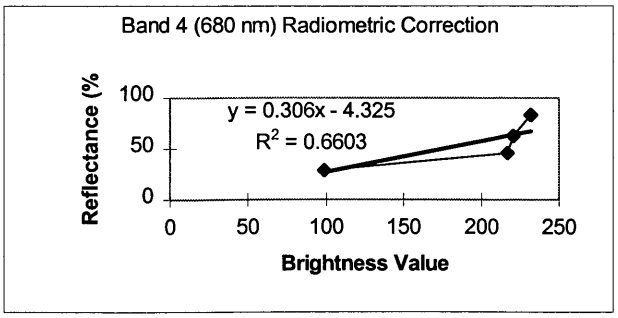


Figure 4. Principal Components Analysis (PCA) as Calculated Using CANOCO Software.

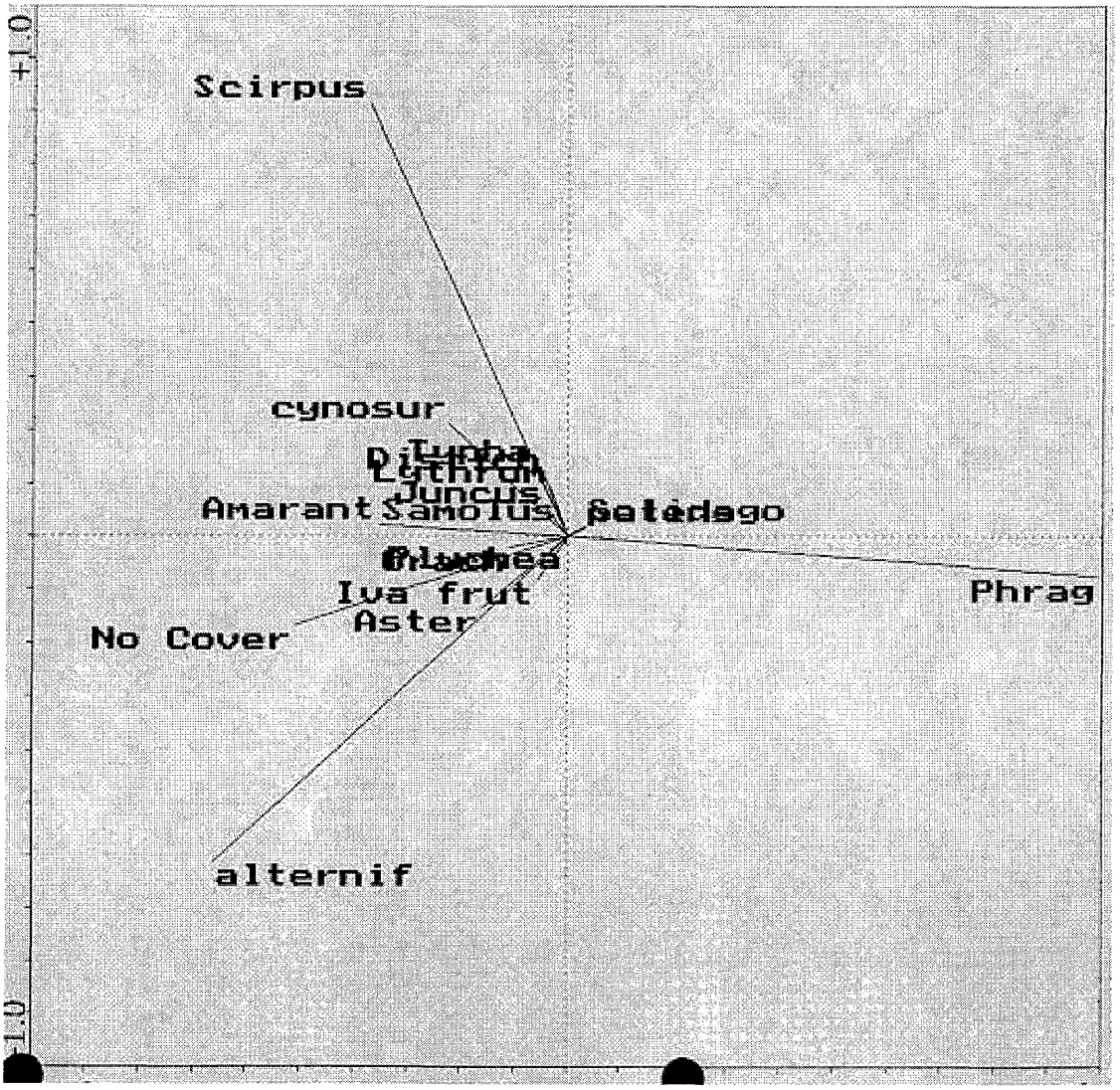
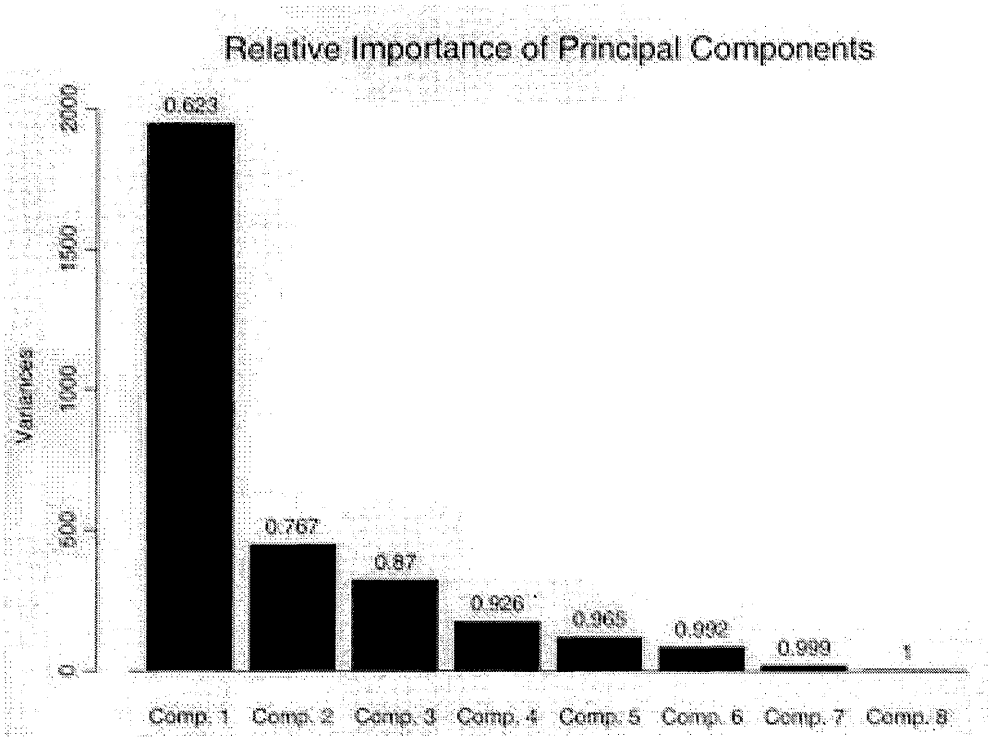


Figure 5. Principal Components Analysis (PCA) as Calculated Using S-Plus.



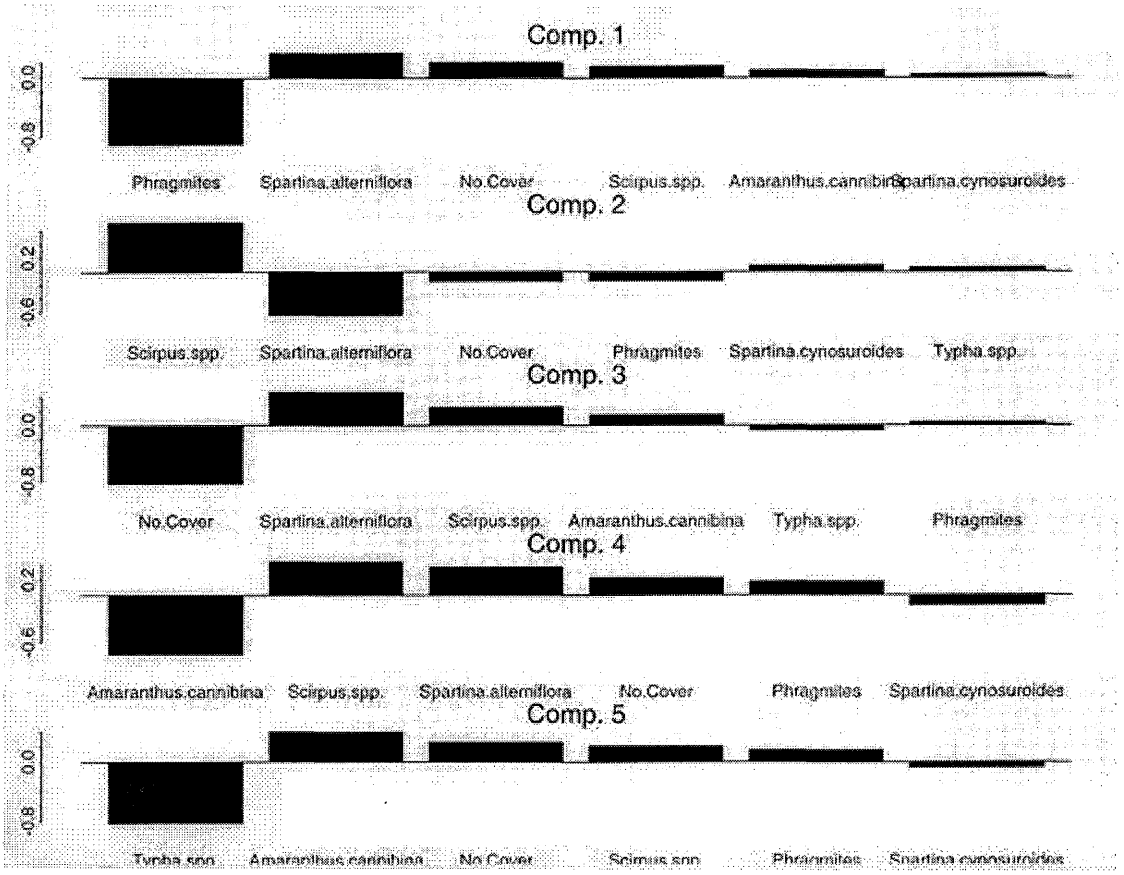

Comp. 1
0.0
-0.8
Phragmites
Spartina.alterniflora
No.Cover
Scirpus.spp.
Amaranthus.cannibina
Spartina.cynosuroides
Comp. 2
0.2
-0.6
Scirpus.spp.
Spartina.alterniflora
No.Cover
Phragmites
Spartina.cynosuroides
Typha.spp.
Comp. 3
0.0
-0.8
No.Cover
Spartina.alterniflora
Scirpus.spp.
Amaranthus.cannibina
Typha.spp.
Phragmites
Comp. 4
0.2
-0.6
Amaranthus.cannibina
Scirpus.spp.
Spartina.alterniflora
No.Cover
Phragmites
Spartina.cynosuroides
Comp. 5
0.0
-0.8
Typha.spp.
Amaranthus.cannibina
No.Cover
Scirpus.spp.
Phragmites
Spartina.cynosuroides

Figure 6. Unsupervised Classification.

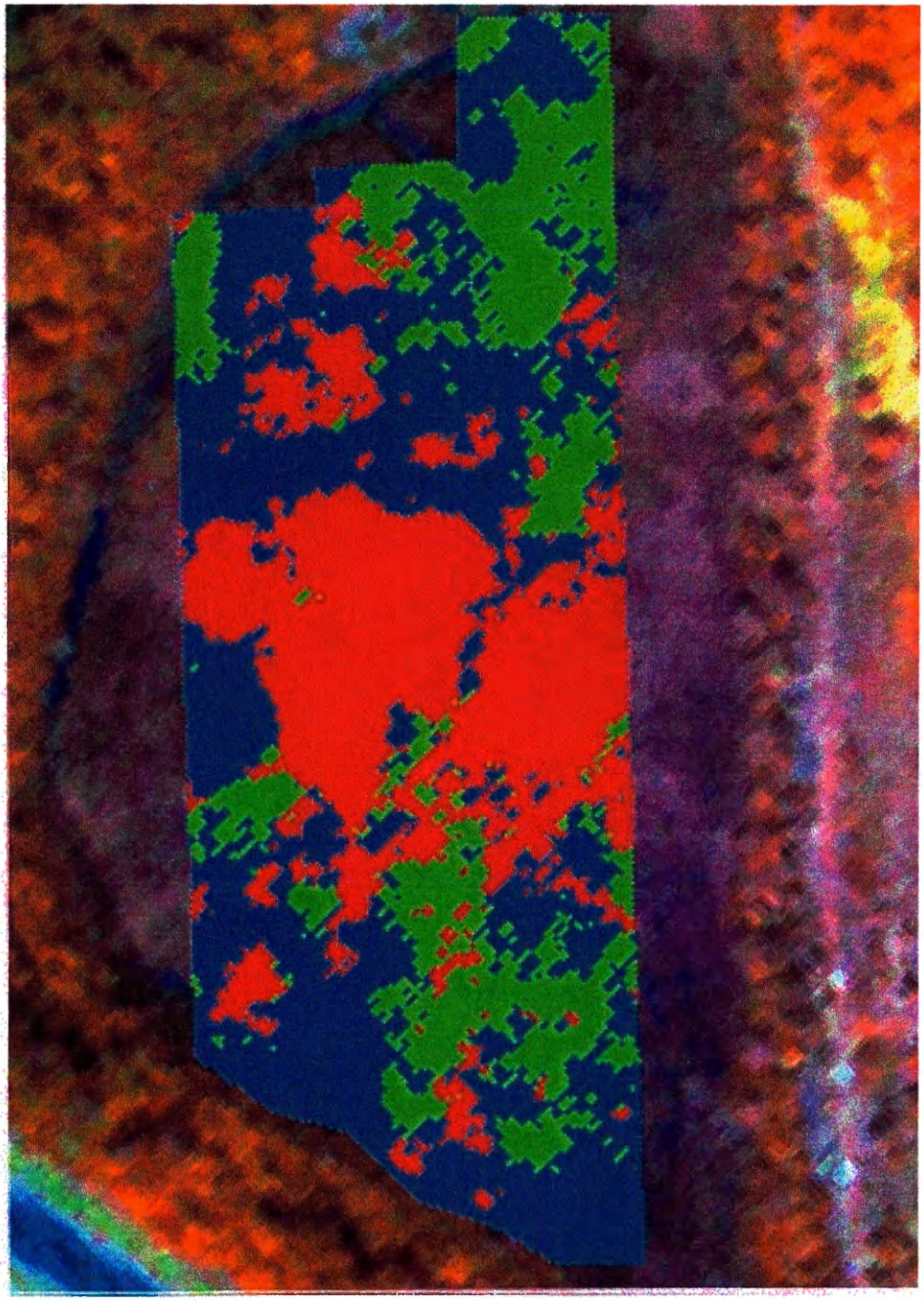


Figure 7. Supervised Classification Produced Using K-Means Clustered Data, Half of Available Training Data, Complete Dataset.
Blue = Class 1 (*S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*)
Green = Class 2 (*P. australis*)
Red = Class 3 (*Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina)*

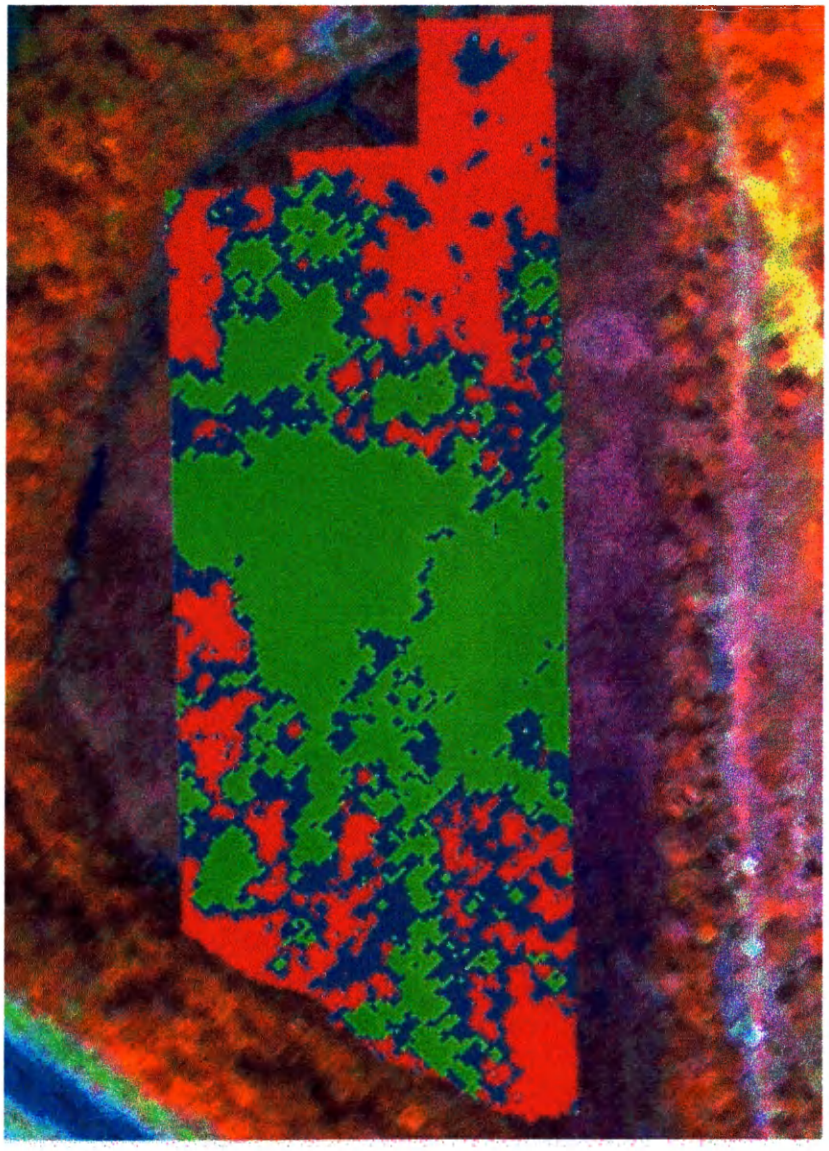


Figure 8. Supervised Classification Produced Using K-Means Clustered Data, Half of Available Training Data, Condensed Dataset.
Blue = Class 1 (*S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*)
Green = Class 2 (*P. australis*)
Red = Class 3 (*Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina)*

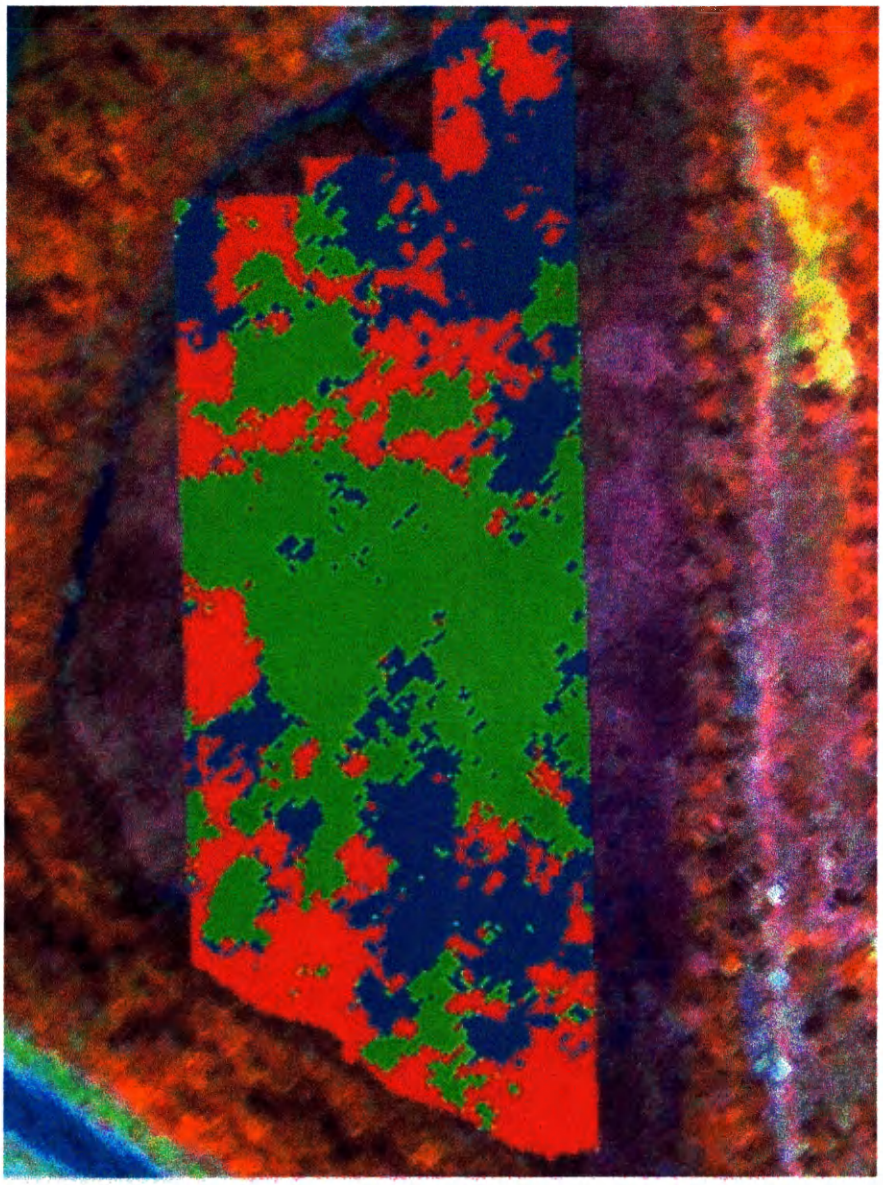


Figure 9. Supervised Classification Produced Using K-Means Clustered Data, All of Available Training Data, Complete Dataset.
Blue = Class 1 (*S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*)
Green = Class 2 (*P. australis*)
Red = Class 3 (*Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina)*

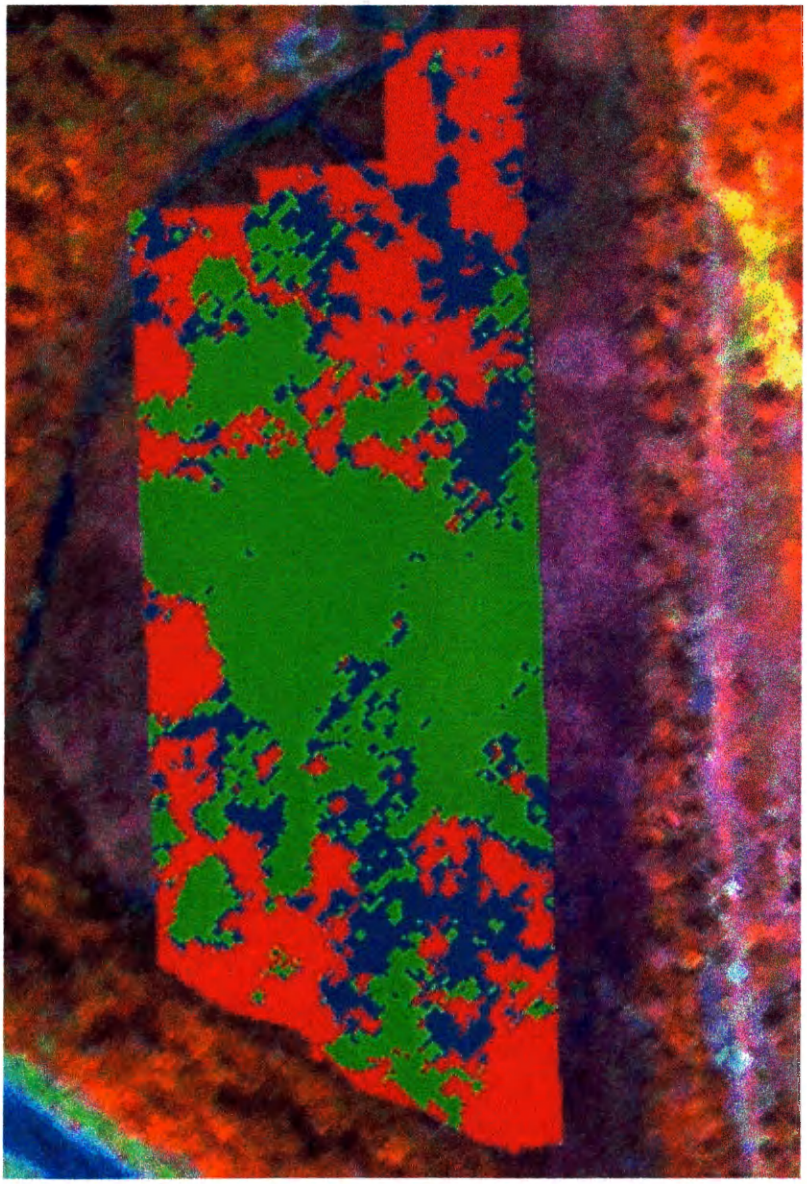


Figure 10. Supervised Classification Produced Using K-Means Clustered Data, All of Available Training Data, Condensed Dataset.
Blue = Class 1 (*S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*)
Green = Class 2 (*P. australis*)
Red = Class 3 (*Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina)*

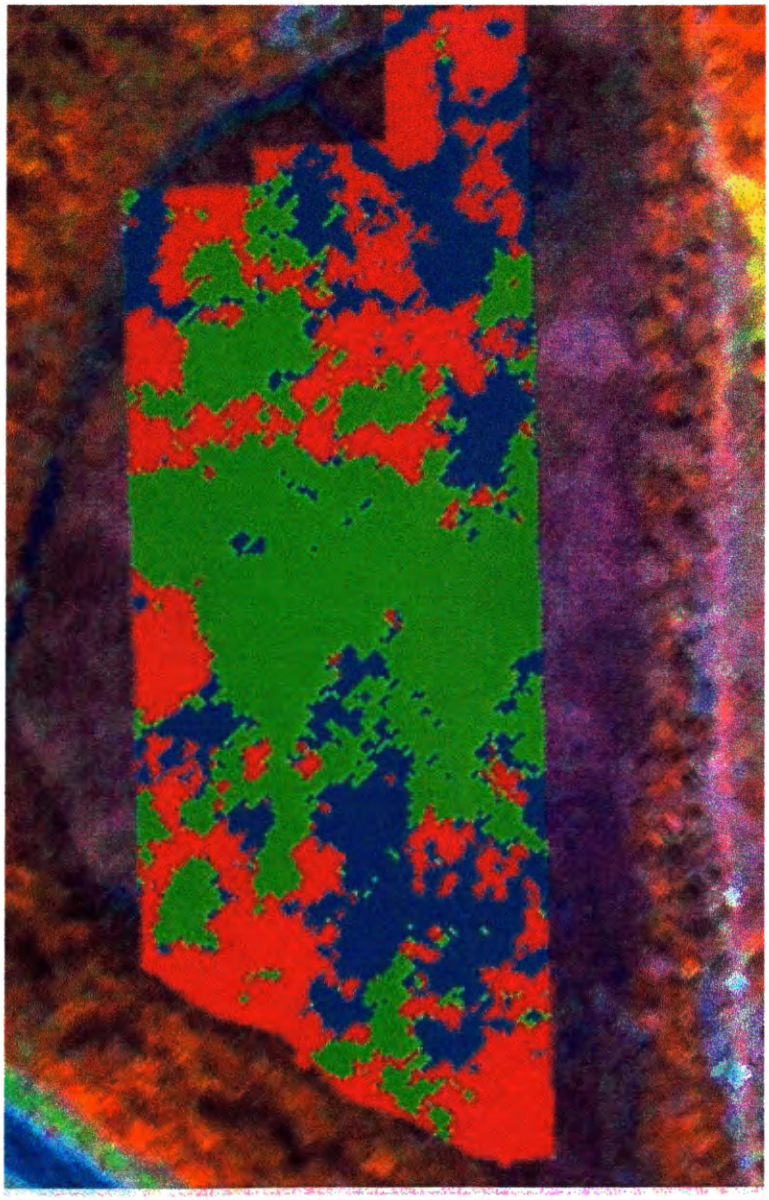


Figure 11. Supervised Classification Produced Using Divisive Hierarchical Clustered Data, Half of Available Training Data, Complete Dataset.
Blue = Class 1 (*Scirpus* spp., *S. cynosuroides*, *A. cannibina*, *Typha* spp.)
Green = Class 2 (*S. alterniflora*, No Cover more prominent, *A. cannibina*)
Red = Class 3 (*P. australis*)

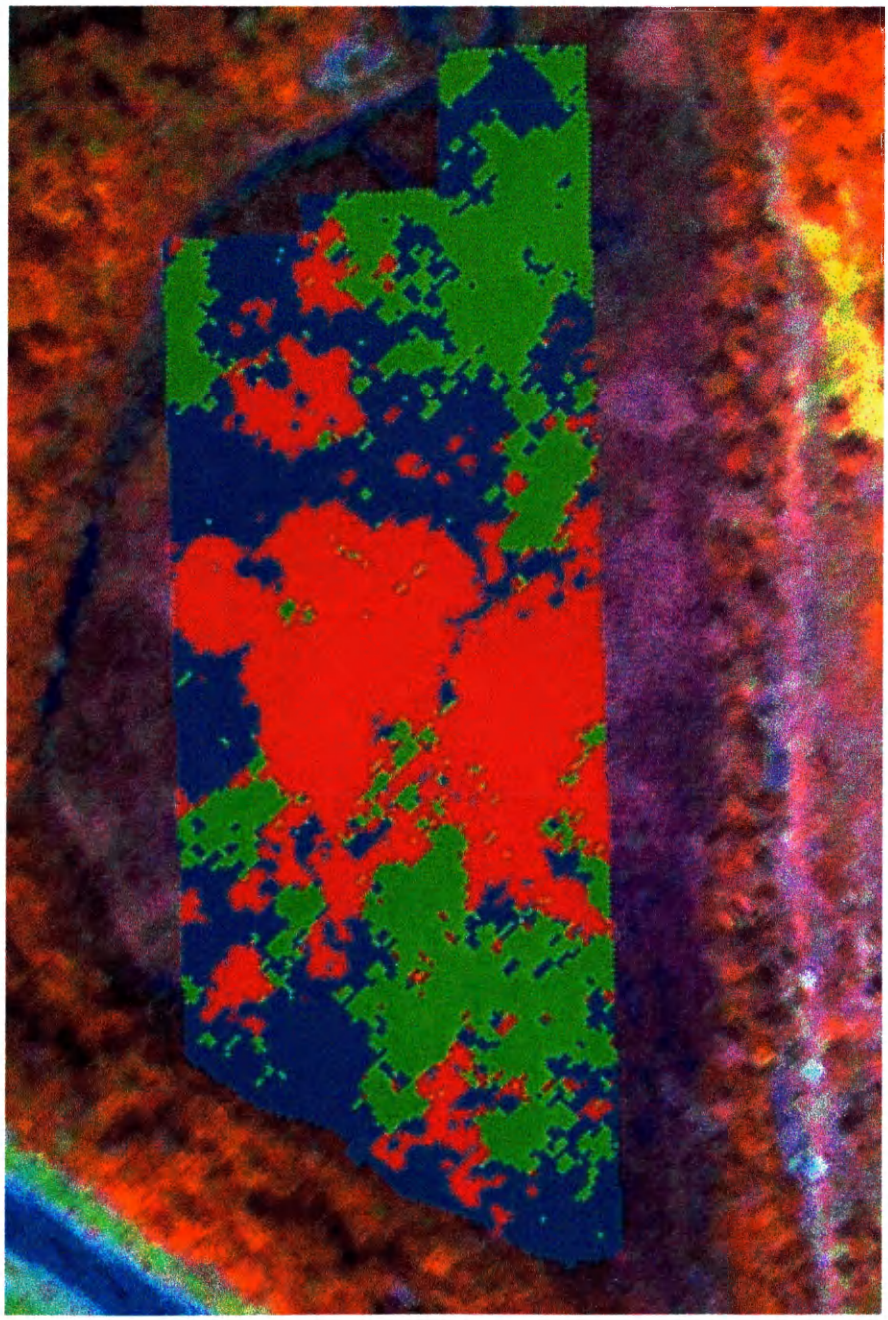


Figure 12. Supervised Classification Produced Using Divisive Hierarchical Clustered Data, Half of Available Training Data, Condensed Dataset.
Blue = Class 1 (*Scirpus* spp., *S. cynosuroides*, *A. cannibina*, *Typha* spp.)
Green = Class 2 (*S. alterniflora*, No Cover more prominent, *A. cannibina*)
Red = Class 3 (*P. australis*)

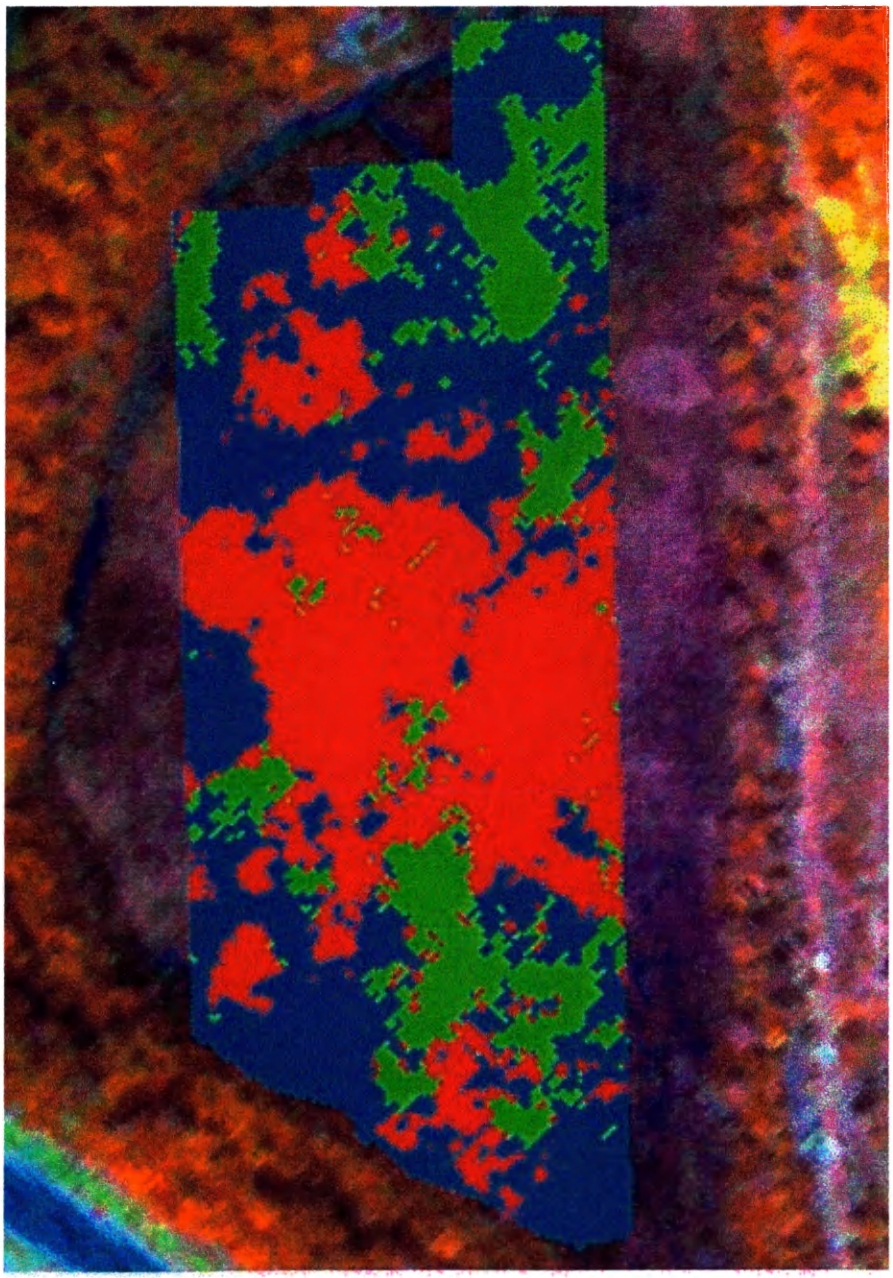


Figure 13. Supervised Classification Produced Using Divisive Hierarchical Clustered Data, All of Available Training Data, Complete Dataset.
Blue = Class 1 (*Scirpus* spp., *S. cynosuroides*, *A. cannibina*, *Typha* spp.)
Green = Class 2 (*S. alterniflora*, No Cover more prominent, *A. cannibina*)
Red = Class 3 (*P. australis*)

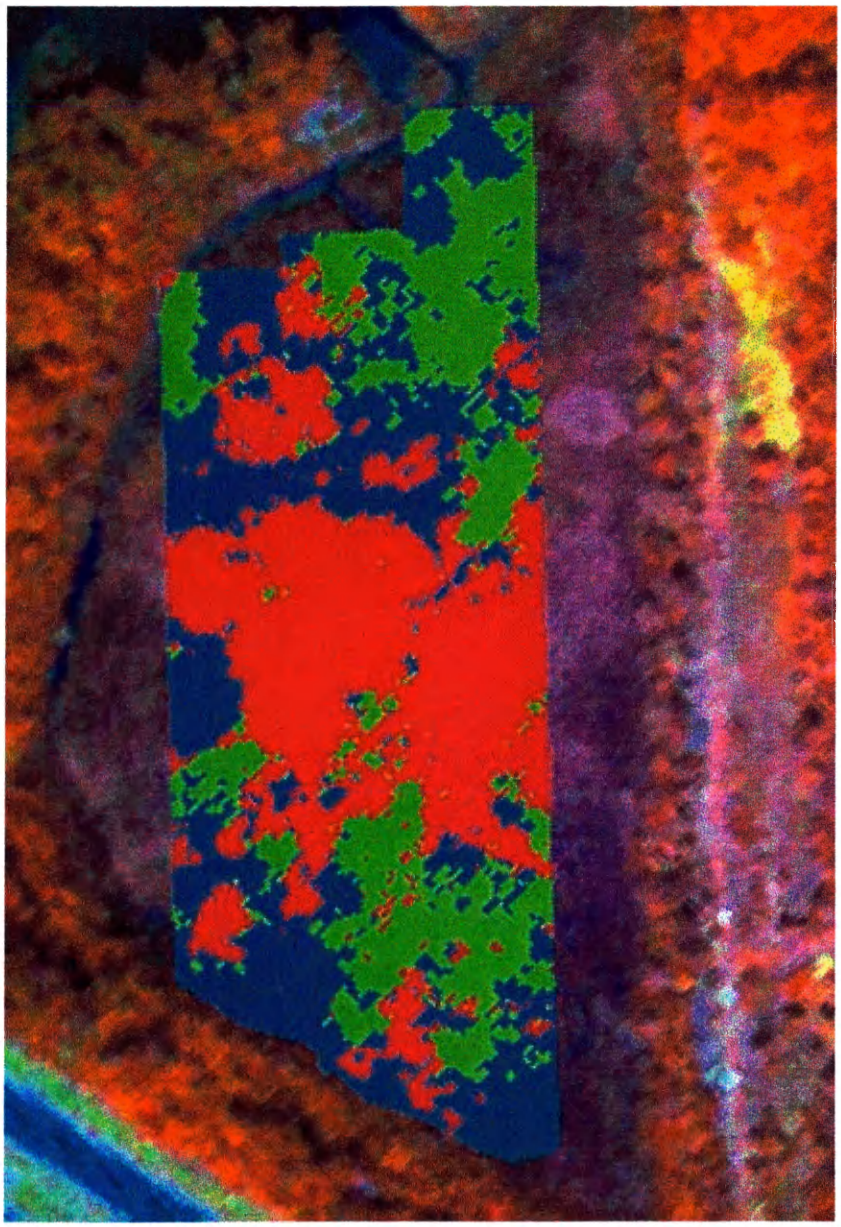


Figure 14. Supervised Classification Produced Using Divisive Hierarchical Clustered Data, All of Available Training Data, Condensed Dataset.
Blue = Class 1 (*Scirpus* spp., *S. cynosuroides*, *A. cannibina*, *Typha* spp.)
Green = Class 2 (*S. alterniflora*, No Cover more prominent, *A. cannibina*)
Red = Class 3 (*P. australis*)

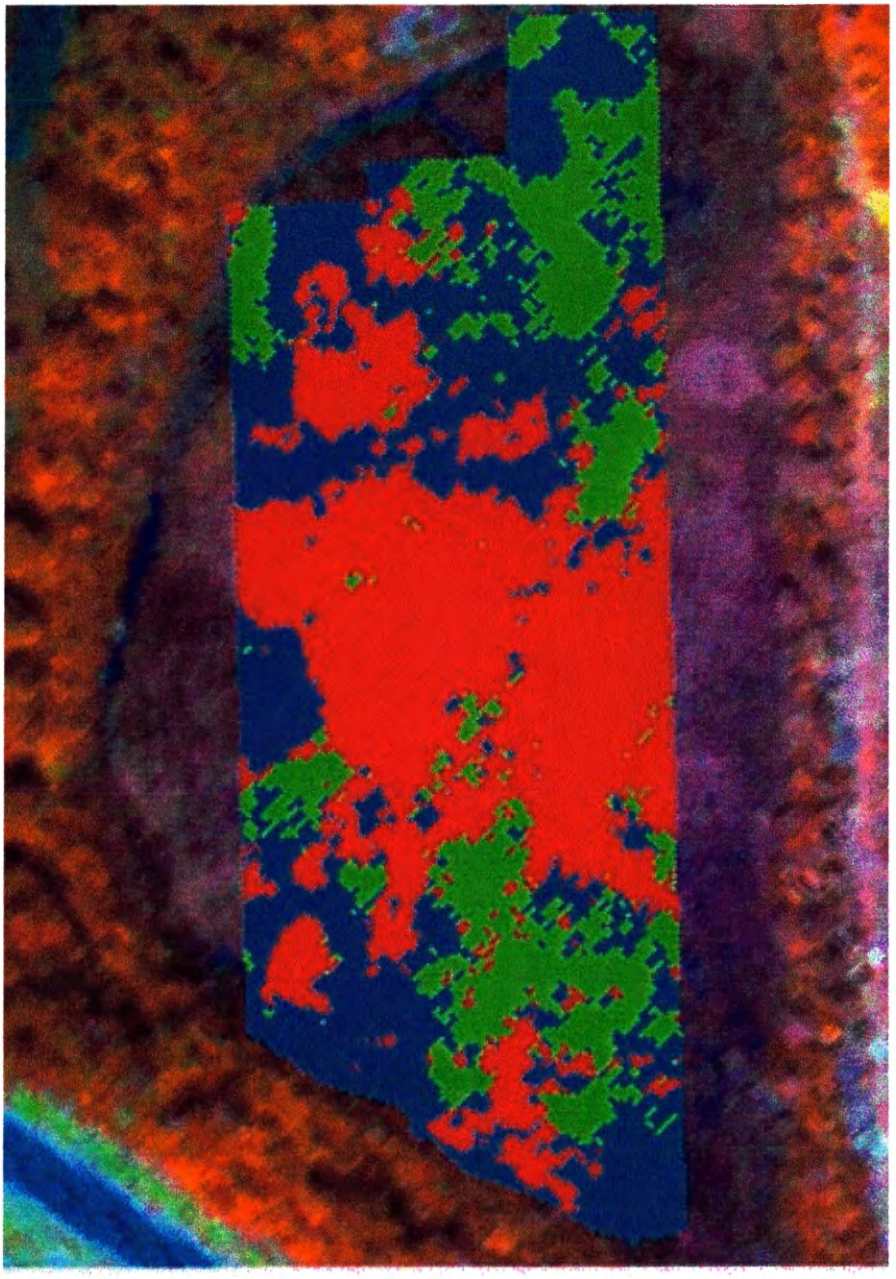


Figure 15. Supervised Classification Produced Using K-Means Clustered Data, All Available Training Data, Complete Dataset as a Verification of Results.
Blue = Class 1 (*S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*)
Green = Class 2 (*P. australis*)
Red = Class 3 (*Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina)*

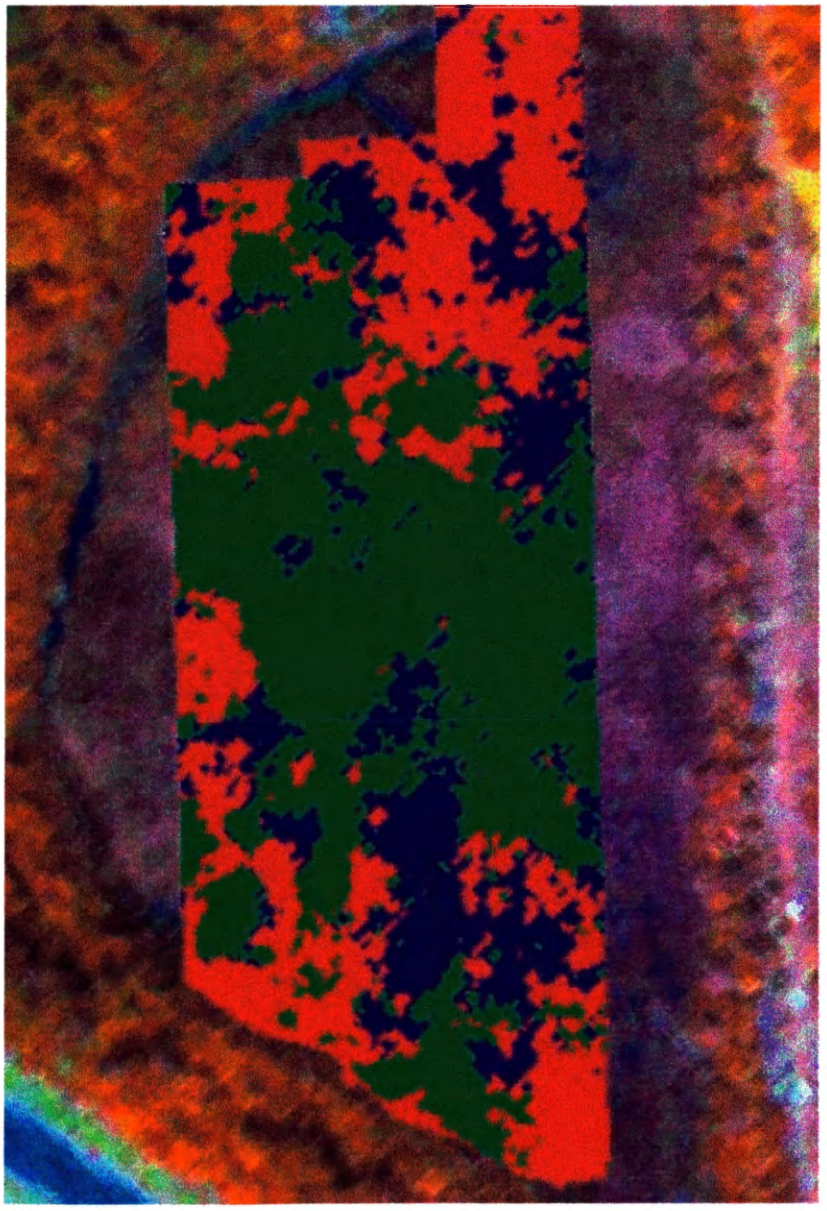
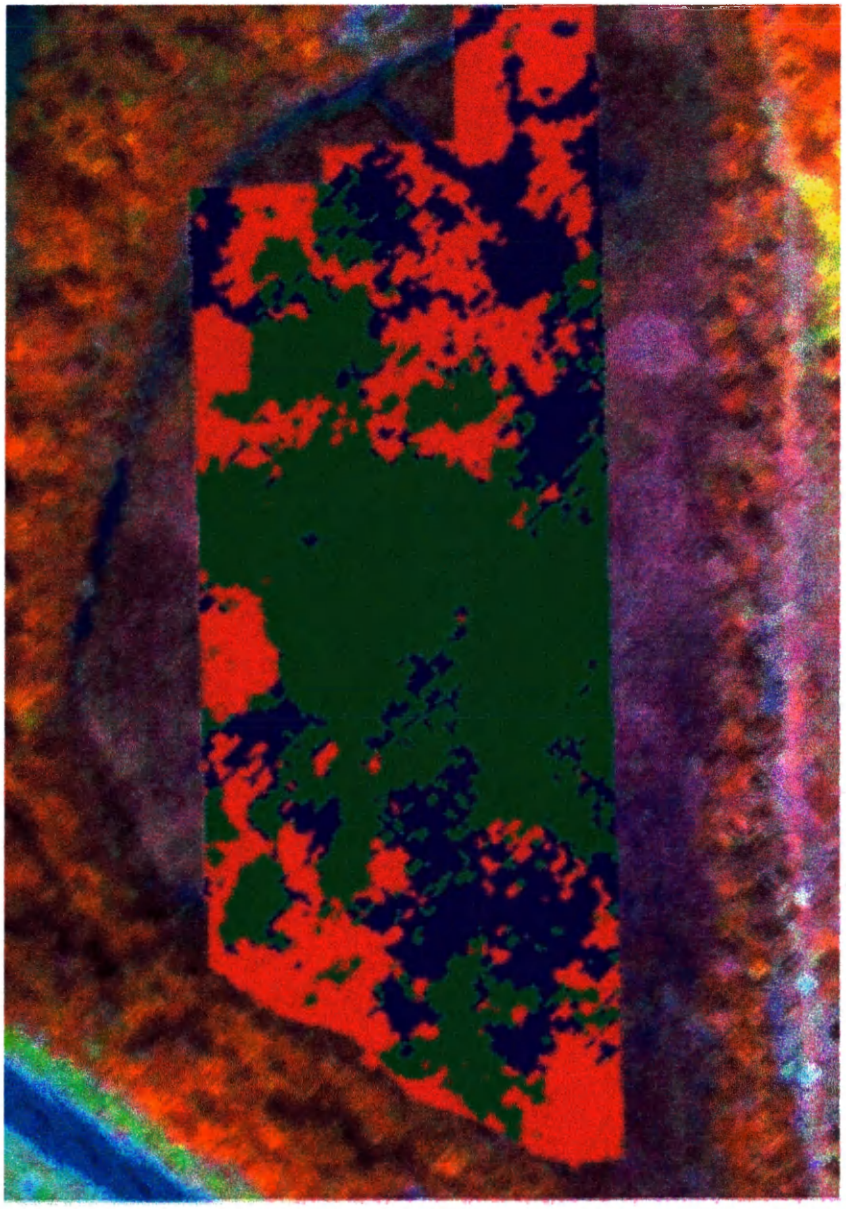
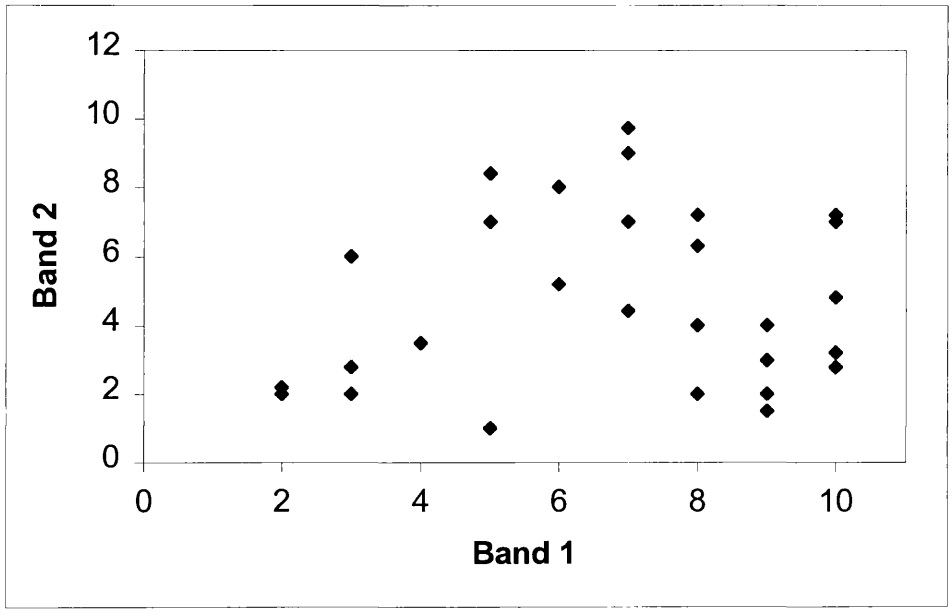


Figure 16. Supervised Classification Produced Using K-Means Clustered Data, All Available Training Data, Complete Dataset as a Verification of Results.
Blue = Class 1 (*S. alterniflora* (main dominant), *A. cannibina, Scirpus* (not as prominent as in Class 3), *P. australis*)
Green = Class 2 (*P. australis*)
Red = Class 3 (*Scirpus* spp. (main dominant), *S. alterniflora*, *S. cynosuroides, A. cannibina)*



Example 1. Sample Classification Distribution



Example 2. Sample Error Matrix

		Reference Data	(Ground-truthed)			
		W	R	T	S	Row Total
	W	50	0	5	0	55
(Pixel)	R	2	25	0	0	27
Classified Data	T	20	5	100	0	125
	S	0	1	7	20	28
	Column Total	72	31	112	20	235

APPENDIX

Unsupervised Classification Assessed with K-Means Clustered Data

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	14	31	26	71	
Classified	Class 2	6	21	0	27	
	Class 3	0	11	33	44	
	Column Totals	20	63	59	142	
	Kappa	0.27		Overall Accuracy		0.48
	Theta One	0.48				
	Theta Two	0.28				
	Theta Three	0.33				
	Theta Four	0.62				
	Kappa Var.	0.00				
	Z score	4.06				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.20		Class 1	0.70	
	Class 2	0.78		Class 2	0.33	
	Class 3	0.75		Class 3	0.56	

Unsupervised Classification Assessed with Divisive Hierarchical Clustered Data

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	24	20	26	70	
Classified	Class 2	5	17	5	27	
	Class 3	4	8	32	44	
	Column Totals	33	45	63	141	
	Kappa	0.29		Overall Accuracy		0.52
	Theta One	0.52				
	Theta Two	0.32				
	Theta Three	0.36				
	Theta Four	0.56				
	Kappa Var.	0.00				
	Z score	4.45				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.34		Class 1	0.73	
	Class 2	0.63		Class 2	0.38	
	Class 3	0.73		Class 3	0.51	

Supervised classification produced using k-means clustered data, half of available training data, complete dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	14	13	12	39	
Classified	Class 2	24	40	1	65	
	Class 3	25	6	7	38	
	Column Totals	63	59	20	142	
	Kappa	0.12		Overall Accuracy		0.43
	Theta One	0.43				
	Theta Two	0.35				
	Theta Three	0.34				
	Theta Four	0.58				
	Kappa Var.	0.00				
	Z score	1.90				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.36		Class 1	0.22	
	Class 2	0.62		Class 2	0.68	
	Class 3	0.18		Class 3	0.35	

Supervised classification produced using k-means clustered data, half of available training data, condensed dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	27	6	7	40	
Classified	Class 2	21	39	4	64	
	Class 3	14	14	9	37	
	Column Totals	62	59	20	141	
	Kappa	0.28		Overall Accuracy		0.53
	Theta One	0.53				
	Theta Two	0.35				
	Theta Three	0.41				
	Theta Four	0.58				
	Kappa Var.	0.00				
	Z score	4.31				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.68		Class 1	0.44	
	Class 2	0.61		Class 2	0.66	
	Class 3	0.24		Class 3	0.45	

Supervised classification produced using k-means clustered data, all available training data, complete dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	19	3	5	27	
Classified	Class 2	24	43	2	69	
	Class 3	20	13	13	46	
	Column Totals	63	59	20	142	
	Kappa	0.29		Overall Accuracy		0.53
	Theta One	0.53				
	Theta Two	0.33				
	Theta Three	0.40				
	Theta Four	0.62				
	Kappa Var.	0.00				
	Z score	4.53				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.70		Class 1	0.30	
	Class 2	0.62		Class 2	0.73	
	Class 3	0.28		Class 3	0.65	

Supervised classification produced using k-means clustered data, all available training data, condensed dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	23	6	4	33	
Classified	Class 2	20	41	4	65	
	Class 3	19	12	12	43	
	Column Totals	62	59	20	141	
	Kappa	0.30		Overall Accuracy		0.54
	Theta One	0.54				
	Theta Two	0.34				
	Theta Three	0.40				
	Theta Four	0.58				
	Kappa Var.	0.00				
	Z score	4.70				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.70		Class 1	0.37	
	Class 2	0.63		Class 2	0.69	
	Class 3	0.28		Class 3	0.60	

Supervised classification produced using divisive hierarchical clustered data, half of available training data, complete dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	20	15	19	54	
Classified	Class 2	9	21	9	39	
	Class 3	4	9	35	48	
	Column Totals	33	45	63	141	
	Kappa	0.31		Overall Accuracy		0.54
	Theta One	0.54				
	Theta Two	0.33				
	Theta Three	0.37				
	Theta Four	0.49				
	Kappa Var.	0.00				
	Z score	4.93				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.37		Class 1	0.61	
	Class 2	0.54		Class 2	0.47	
	Class 3	0.73		Class 3	0.56	

Supervised classification produced using divisive hierarchical clustered data, half of available training data, condensed dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	21	21	22	64	
Classified	Class 2	5	14	4	23	
	Class 3	7	10	37	54	
	Column Totals	33	45	63	141	
	Kappa	0.27		Overall Accuracy		0.51
	Theta One	0.51				
	Theta Two	0.33				
	Theta Three	0.37				
	Theta Four	0.56				
	Kappa Var.	0.00				
	Z score	4.14				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.33		Class 1	0.64	
	Class 2	0.61		Class 2	0.31	
	Class 3	0.69		Class 3	0.59	

Supervised classification produced using divisive hierarchical clustered data, all available training data, complete dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	23	12	19	54	
Classified	Class 2	5	19	7	31	
	Class 3	5	14	37	56	
	Column Totals	33	45	63	141	
	Kappa	0.34		Overall Accuracy		0.56
	Theta One	0.56				
	Theta Two	0.34				
	Theta Three	0.39				
	Theta Four	0.52				
	Kappa Var.	0.00				
	Z score	5.28				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.43		Class 1	0.70	
	Class 2	0.61		Class 2	0.42	
	Class 3	0.66		Class 3	0.59	

Supervised classification produced using divisive hierarchical clustered data, all available training data, condensed dataset

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	20	18	18	56	
Classified	Class 2	5	15	5	25	
	Class 3	8	12	40	60	
	Column Totals	33	45	63	141	
	Kappa	0.29		Overall Accuracy		0.53
	Theta One	0.53				
	Theta Two	0.34				
	Theta Three	0.39				
	Theta Four	0.55				
	Kappa Var.	0.00				
	Z score	4.55				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.36		Class 1	0.61	
	Class 2	0.60		Class 2	0.33	
	Class 3	0.67		Class 3	0.63	

Supervised classification produced using k-means clustered data, all available training data, complete dataset as a verification of results

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	23	7	2	32	
Classified	Class 2	19	44	5	68	
	Class 3	21	8	13	42	
	Column Totals	63	59	20	142	
	Kappa	0.34		Overall Accuracy		0.57
	Theta One	0.56				
	Theta Two	0.34				
	Theta Three	0.43				
	Theta Four	0.60				
	Kappa Var.	0.00				
	Z score	5.32				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.72		Class 1	0.37	
	Class 2	0.65		Class 2	0.75	
	Class 3	0.31		Class 3	0.65	

Supervised classification produced using k-means clustered data, half of available training data, complete dataset as a verification of results

			Reference			
		Class 1	Class 2	Class 3	Row Totals	
	Class 1	26	5	3	34	
Classified	Class 2	18	44	3	65	
	Class 3	19	10	14	43	
	Column Totals	63	59	20	142	
	Kappa	0.38		Overall Accuracy		0.60
	Theta One	0.59				
	Theta Two	0.34				
	Theta Three	0.44				
	Theta Four	0.58				
	Kappa Var.	0.00				
	Z score	6.13				
	User's Accuracy			Producer's Accuracy		
	Class 1	0.76		Class 1	0.41	
	Class 2	0.68		Class 2	0.75	
	Class 3	0.33		Class 3	0.70	

LITERATURE CITED

Anderson, J.E. 1997, January 16. Course Lecture for MS 592: Remote Sensing of Aquatic Resources.

Anderson, J.E. 1997, January 23. Course Lecture for MS 592: Remote Sensing of Aquatic Resources

Aronoff, S. 1985. The Minimum Accuracy Value as an Index of Classification Accuracy. *Photogrammetric Engineering and Remote Sensing* 51(1): 99-111.

Barnard, Jr., T.A., C.H. Hershner, and M. Greiner. 1997. The Oldest Operating Wetland Mitigation Bank in the U.S. Wetlands Program Technical Report 97-7. Virginia Institute of Marine Science. College of William and Mary.

Bishop, Y.M.M., S.E. Feinberg, and P.W. Holland. (from Rosenfield and Fitzpatrick-Lins 1986). 1975. *Discrete Multivariate Analysis-Theory and Practice.* Cambridge, Mass., The MIT Press.

Carter, G.A. 1991. Primary and secondary effects of water content on the spectral reflectance of leaves. *American Journal of Botany* 78(7): 916-924.

Congalton, R.G., Oderwald, R.G., and R.A. Mead. 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing* 49: 1671-1678.

Congalton, R.G. 1988. A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing* 54(5): 593-600.

Congalton, R.G. and K. Green. 1993. A Practical Look at the Sources of Confusion in Error Matrix Generation. *Photogrammetric Engineering and Remote Sensing* 59(5): 641-644.

Congalton, R.G. 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of the Environment* 37: 35-46.

Congalton, R.G. and K. Green. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. New York: Lewis Publishers. 137 p.

Daubenmire, R.F. 1968 (referenced from Mueller-Dombois and Ellenberg 1974). Plant communities: A textbook of plant synecology. Harper & Row. New York. 300 p.

Dicks, S.E. and T.H.C. Lo. 1990. Evaluation of Thematic Map Accuracy in a Land-Use and Land-Cover Mapping Program. Photogrammetric Engineering and Remote Sensing 56(9): 1247-1252.

Environmental Laboratory. 1987. "Corps of Engineers Wetlands Delineation Manual," Technical Report Y-87-1, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

Fitzpatrick-Lins, K. 1981. Comparison of Sampling Procedures and Data Analysis for a Land-Use and Land-Cover Map. *Photogrammetric Engineering and Remote Sensing* 47(3): 343-351.

Gauch, Jr., H.G. Multivariate Analysis in Community Ecology. 1995. Cambridge University Press, New York, New York. 298 p.

Ginnevan, M.E. 1979. Testing Land-Use Map Accuracy: Another Look. *Photogrammetric Engineering and Remote Sensing* 45(10): 1371-1377.

Hardisky, M.A., M.F. Gross, and V. Klemas. 1986. Remote Sensing of Coastal Wetlands. *Bioscience* 36(7): 453-460.

Hay, A.M. 1979. Sampling Designs to Test Land-Use Map Accuracy. *Photogrammetric Engineering and Remote Sensing* 45(4): 529-533.

Hinson, J.M., C.D. German, and W. Pulich, Jr. 1994. Accuracy Assessment and Validation of Classified Satellite Imagery of Texas Coastal Wetlands. *Marine Technology Society Journal* 28(2): 4-9.

Jakubauskas, M.E., Whistler, J.L, Dillworth, M.E., and E.A. Martinko. 1992. Classifying remotely sensed data for use in an agricultural nonpoint-source pollution model. *Journal of Soil Water Conservation* 47: 179-183.

Lillesand, T.M. and R.W. Kiefer. 1994. Remote Sensing and Photo Interpretation. John Wiley & Sons: New York, NY, USA.

Ma, Z. and R.L. Redmond. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing* 61: 435-439.

Magurran, A.E. 1988. *Ecological Diversity and Its Measurement*. Princeton University Press: New Jersey. 179 p.

Marks, M., B. Lapin, and J. Randall. 1994. *P. australis (P. communis)*: Threats, Management, and Monitoring. *Natural Areas Journal* 14(4): 285-294.

MathSoft, Inc. May 1999. S-Plus User's Guide. Seattle: MathSoft, Inc. 558 p.

Mitsch, W. and J.G. Gosselink. 1993. Wetlands. Van Nostrand Rhinehold, New York, NY, USA.

Mueller-Dombois, D. and H. Ellenberg. 1974. Aims and Methods of Vegetation Ecology. John Wiley & Sons, NY, USA.

Murtha, P.A. 1982. Detection and analysis of vegetation stress. *In:* Johannson, C. & J. Sanders (eds.). Remote Sensing for Resource Management. SCSA. Ankeny. IA, USA.

Perry, J. and C. Hershner. (in press). Temporal Changes in the Vegetation Pattern in a Tidal Freshwater Marsh. *Wetlands*.

Rice, R.L. 1967 (referenced from Mueller-Dombois and Ellenberg 1974). A statistical method of determining quadrat size and adequacy of sampling. *Ecology* 48: 1047-1049.

Robinove, C.J. 1986. Spatial Diversity Index Mapping of Classes in Grid Cell Maps. *Photogrammetric Engineering and Remote Sensing* 52(8): 1171-1173.

Rosenfield, G.H. 1986. Analysis of Thematic Map Classification Error Matrices. *Photogrammetric Engineering and Remote Sensing* 52(5): 681-686.

Rosenfield, G.H. and K. Fitzpatrick-Lins. 1986. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering and Remote Sensing* 52(2): 223-227.

Rosenfield, G.H., K. Fitzpatrick-Lins, & H.S. Ling. 1982. Sampling for Thematic Map Accuracy Testing. *Photogrammetric Engineering and Remote Sensing* 48(1): 131-137.

SAS Institute, Inc. 1994. SAS/STAT User's Guide, v. 2 (version 6). Cary, North Carolina: SAS Institute, Inc. 1686 p.

Short, N.M. 1982 (from Rosenfield and Fitzpatrick-Lins 1986). 1982. *The Landsat Tutorial Workbook – Basics of Satellite Remote Sensing*: Greenbelt, Md., Goddard Space Flight Center, NASA Reference Publication 1078.

Silberhorn, G.M. 1982. Common Plants of the mid-Atlantic Coast: A Field Guide. Baltimore: Johns Hopkins University Press. 256 p.

Smith, C., N. Pyden, and P. Cole. 1994. ERDAS Field Guide, 3rd ed. Atlanta: ERDAS, Inc.

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

Stehman, S.V. 1997. Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sensing and the Environment* 62:77-89.

Story, M. & R.G. Congalton. 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing* 52(3): 397-399.

Swain, P.H. and S.M. Davis (referenced from Smith *et al.* 1994). 1978. Remote Sensing: The Quantitative Approach. New York: McGraw Hill Book Company.

Turk, G. (from Rosenfield and Fitzpatrick-Lins 1986). 1979. GT Index: A measure of the success of prediction: *Remote Sensing of Environment* 8: 65-75.

Zar, J.H. 1984. *Biostatistical Analysis*, 2nd ed. Prentice Hall: New Jersey. 718 p.

VITA

Jason Samuel Goldberg

Born in Syracuse, New York, 25 August 1974. Graduated valedictorian from Ridgewood High School in 1992. Earned B.S. in Marine Science and Biology, *magna cum laude*, from University of Miami, Florida, in 1996. Spent 1995 studying abroad at James Cook University in Townsville, Queensland, Australia. Entered M.S. program in College of William and Mary, School of Marine Science in 1996. Entered M.P.P. program in College of William and Mary, Thomas Jefferson Program in Public Policy, in 1997. Received Knauss Sea Grant Fellowship in 2000 with Congressman Neil Abercrombie (D-HI). Trying to boldly go where a few have gone before, as well as a few new places.