CLASSIFICATION OF WETLANDS VEGETATION USING SMALL SCALE COLOR INFRARED IMAGERY

ANNUAL REPORT
October 9, 1973 to December 20, 1974

Prepared Under Contract No. NAS6-1913 by
Chesapeake Bay Center For Environmental Studies
Smithsonian Institution

For
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WALLOPS FLIGHT CENTER
WALLOPS ISLAND, VIRGINIA 23337

February 1975
CLASSIFICATION OF WETLAND VEGETATION USING
SMALL SCALE COLOR INFRARED IMAGERY

Annual Report
October 9, 1973 to December 20, 1974

Contract No. NAS6-1913
National Aeronautics and Space Administration
Wallops Station, Wallops Island, Virginia 23337

Chesapeake Bay Center for Environmental Studies
Smithsonian Institution

Francis S. L. Williamson, Ph.D.
Principal Investigator

Staff
B.L. Rice, Ph.D., Plant Ecologist
Deborah (Ford) Cassel, Research Technician
The purpose of the research reported here is to consider classification systems for Chesapeake Bay wetland vegetation. This work is an extension of previous studies carried out by Smithsonian Institution (S.I.) investigators (Welles, et al., 1973), and is concerned with the problem of mapping and classifying vegetation types using high altitude color infrared aerial photography.

There is a growing body of literature which indicates that vegetation can be classified, and even measured quantitatively with imagery from satellites equipped with multispectral scanner systems (e.g., Carter and Schubert, 1974; Bently, 1974; Klemas et al., 1974; Rouse et al., 1974; Tueller et al., 1974; Williamson, 1974). However, although a satellite provides repetitive, high quality imagery covering very large areas, the possibility exists that cloud cover or orbital timing could prevent coverage of any given area at a critical time. Aerial photography may be more suitable for detailed surveys, especially those that intend to census species distributions or short term environmental disturbances. Driscoll et al. (1974) have used microdensitometry and large scale color infrared aerial photos to discriminate individual shrub and tree species within a pinyon pine-juniper plant community in Colorado. However, their attempts to differentiate between different sites or cultural treatments within native grasslands were not so successful. Driscoll and Coleman (1974) also tested the ability of photointerpreters to identify plant species on large scale color infrared photos; seven of eleven species could be correctly identified over 80 percent of the time. Physiognomically, our test sites on the Chesapeake Bay would most closely resemble those on their native grasslands, so we would be testing much the same techniques on a slightly
different habitat. Seher and Tueller (1973) considered color infrared photos useful for evaluating marsh vegetation in Nevada; they point out that while large scale (1:1000) gives greater accuracy than small scale (1:10,000), coverage was much better with small scales.

There are therefore good reasons for continuing research on vegetation classification systems using color infrared photography, and attempting to obtain accuracy of species identification at smaller scales. The work completed in 1973 by Smithsonian personnel used photointerpretive techniques on low (1828 m) and high (18,000 m) altitude color infrared and natural color photography. After photointerpretation, vegetation identifications were checked in the field. Six 2.59 km² vegetation maps were produced for Eastern Shore marshes. One of these mapped areas, Farm Creek Marsh, was selected to provide trial training sets for 1974 classification procedures.

For the 1974 research effort Smithsonian Institution and the National Aeronautics and Space Administration (NASA) agreed to emphasize the use of automated data analysis in classifying wetlands vegetation. The data processing programs to be used were those developed by the Laboratory for the Applications of Remote Sensing (LARS) at Purdue University (Lindenlaub, 1973). LARS programs have primarily been used for classification of agricultural crops, using large scale multispectral scanner imagery. We wished to test the value of these programs in classifying natural vegetation, using digitized data from small scale aerial photography.

RESEARCH GOALS

Since the purpose of the 1974 research program was to develop a classification system for Chesapeake Bay wetlands derived from the correlation of film density classes and actual vegetation classes, two basic questions had to be
answered. First, does a classification map produced by the LARS system give a reasonably accurate picture of actual vegetation? Secondly, can classification information obtained from imagery and data analysis of one geographic area be used to predict vegetation types in other areas at other times?

Time was a limited factor for the research project so it was necessary to set up a precise work/timing schedule. Test areas were selected and vegetation sampling transects established in May 1974. New photo data was to be flown over these test sites, and was to be reduced and put on computer compatible tapes by the NASA/Wallops scanning digital microdensitometer. Classification of the data would be performed by S.I. investigators using the remote terminal to LARSYS by NASA/Wallops. Classification sets would be determined by the LARS programs and these would be compared to species composition, cover, and height of vegetation in the test sites. Existing imagery and the vegetation map of Farm Creek Marsh would be used to determine the optimal number of classes, and to aid in determining if the computer maps were a believable product. If this approach produced a reliable correlation between film density classes and real vegetation classes within the test sites, then the parameters for the classification system would be documented and extended to other wetland vegetation.

DESCRIPTION OF TEST SITES, VEGETATION SAMPLING & DATA ANALYSIS METHODS

The Farm Creek Marsh test area is described in Welles et al. (1973). The area had been mapped and classified according to plant species cover. Two new test sites were selected on the Eastern Shore of Chesapeake Bay and detailed vegetation analysis carried out for correlation with automated data analysis classifications of the same areas. Both are high marshes with limited areas of low marsh species, and both are under the management of the Maryland Department of Natural Resources.
Bestpitch Marsh, Dorchester County, Maryland

This site is located on USGS Blackwater River and Chicacomico quadrangles. Blackwater National Wildlife Refuge lies to the west, and the Transquaking River on the east and south (Fig. 1). Mixed hardwood and conifer forest form the northern boundary. Plant species are typical of irregularly flooded brackish marshes.

Two vegetation transects, designated Bestpitch North and Bestpitch South and each 0.61 km long, were established in May 1974 with the assistance of the Maryland Dept. of Natural Resources. The transects run 315° NW, roughly parallel to each other, and are 0.32 km apart (Fig. 2). The ends and middle of each are marked by 2.4 m lengths of gaspipe. At 0.14 km intervals crosstransects are marked with 1 m lengths of gaspipe. For convenience in data analysis, the transects at right angles to the main transect were considered in two sections, each 30 meters long. There were thus 12 subtransects in each transect (24 total), which were sampled every 2 meters with a 0.1 m² rectangular (20 cm x 50 cm) open frame. The subtransects of each main transect were numbered from 1 to 12, beginning at the east end. Odd numbers were north and even numbers south of the main transect. As the two transects were parallel, the data could be analyzed either in an east-west direction, or from north to south. Bestpitch Marsh is at least partially burned every winter for management reasons, and is also used by recreational hunters.

Pigeonhouse Creek (Dames Quarter) Marsh, Somerset County, Maryland

The Pigeonhouse Creek test site (USGS Monie quadrangle) is bounded on the southwest by Pigeonhouse Creek, and on the south by state road 363 (Fig. 3). The vegetation transect is marked by a 1 meter length of gaspipe about 50 meters north of the road and runs north approximately 0.5 km. There are a total of ten 30 meter long crosstransects marked at right angles to the main transect, evenly spaced. This marsh is also burned periodically at undetermined
Figure 1. General location of Bestpitch Marsh test site showing the extent of the wetlands. Forested areas are shown as the figured pattern, ponds and rivers are shaded, and the remainder is marsh.
Figure 2. Bestpitch Marsh transects. Each subtransect was numbered and sampled separately at 2m intervals.
Figure 3. Location of Pigeonhouse Creek Marsh. Tree areas are shown as the figured pattern. The halves of each of the five cross-transects were numbered and sampled separately at 2m intervals.
intervals. It contains a larger number of shrubby species than Bestpitch Marsh, possibly because of the habitat created by extensive mosquito ditching in the past. The sampling procedure was the same as that followed in Dorchester County.

Vegetation Analysis

Aerial photography records the integration of soil, water, and vegetation features within the field of view. When wetlands managers discuss their management practices and policies, they often speak in terms of available amounts of desirable plant species, or in terms of species cover. For example, burning of marshes is often done to encourage the dominance of *Scirpus olneyi*, which is a favored food for muskrats. Plant species cover is, for the above two reasons, a reasonable parameter for analysis and was selected as the basis for naming the classes determined by automated data processing and in the field.

The data obtained for each 0.1 m$^2$ sampling plot were: percent mud or water; percent live plant cover by species; percent dead plant cover by species; height of each species; estimated density of each species. These plant features were selected as being the most likely to affect reflectance values for classifications performed at various times of year and to be instrumental in determining the clustering sets. Field data were collected in June, July and August. Imagery had been requested to coincide with field data collection dates? There were a total of 360 data points (plots) for Bestpitch Marsh and 150 for Pigeonhouse Creek. These data were averaged for each subtransect and the live cover values used as input for a Bray-Curtis ordination program. The species selected for detailed analysis at Bestpitch Marsh were *Spartina patens*, *Scirpus olneyi*, *Distichlis spicata*, *Juncus roemerianus*, *Typha angustifolia*, and *Kosteletzkya virginica*. *S. patens*, *D. spicata*, *J. roemerianus*, *S. olneyi*, *Iva frutescens*, and *Solidago sempervirens* were analyzed for Pigeonhouse Creek.
The Bray-Curtis method of vegetation analysis is one of several methods which arrange vegetation units in some sort of order in one or more dimensions (Bray and Curtis, 1957). It often follows that the order is determined by environmental factors affecting the plant species, and which may be of considerable interest to ecologists. The approach is in opposition to classification systems which arrange vegetation units in discrete classes (Daubenmire, 1966). Gauch and Whittaker (1972) have shown that a Bray-Curtis ordination is most suitable for quantitative data of the sort that we collected. Cornell University has developed a "canned" computer program for Bray-Curtis ordination which was used for our data analysis.

During July the vegetation sampling was supplemented by helicopter flights over the two test sites at 150, 460, and 1525 meters altitude. This aided in determining the vegetation type boundaries over which the transects crossed. Opinions differ as to the optimal season for obtaining good remote sensing data on vegetation. Carter and Schubert (1974) considered October best for satellite multispectral scanner data from saline coastal marshes. Seher and Tueller (1973) used data taken during August to analyze low level aerial photos of Nevada wetlands; they were considering submerged aquatic species as well as marsh species. Since our approach was to attempt to distinguish vegetation classes on the basis of species cover it would have been useful if seasonal phenological changes aided in differentiating the types. Data was therefore collected before the growing season was well underway, and twice during the growing season.

Automated Data Processing of Aerial Photography

Color infrared aerial photos (transparencies) provided by NASA Wallops were processed with a scanning digital microdensitometer. The original photo scale was 1:130,000. A 23 x 23 cm transparency was attached to a rotating drum and the microdensitometer:
1) viewed a 50 micron diameter portion of the photograph image at spectral intervals of $0.67\mu - 0.88\mu$, $0.58\mu - 0.67\mu$, and $0.51\mu - 0.58\mu$. These, with film sensitivity levels and dye components, correspond to infrared, red, and green color bands.

2) The optical density of the image was read by means of a scanning optical system.

3) The sample was scanned at a uniform rate.

4) Data was presented to digital tapes for computer reduction and analysis. These data tapes were reformatted at the NASA Wallops computer facility and then sent to Purdue University for inclusion in the LARS system. Once at Purdue, the data could be accessed by S.I. investigators using the LARS terminal at NASA Wallops Island. Total elapsed time for obtaining the data after flights were made over test sites was never less than 3 months.

The scale of the imagery after it completely processed and printed out on a Data 100 lineprinter was 1:2560 in the horizontal scale and 1:2050 in the vertical scale. The scale variation is caused by the rectangular shape of the lineprinter symbols.

RESULTS AND DISCUSSION

Shortly after the 1974 imagery became available the NASA Wallops microdensitometer broke down and thus prevented the digitizing of data for the two new Eastern Shore test areas. However, two consecutive frames of the August 1973 flight over Farm Creek Marsh had been digitized and data was available for computer analysis. The equipment failure led to a situation whereby detailed computer analysis and classification could only be done for a marsh which had rough field data, while no computer analysis was possible for the marshes with detailed vegetation information. As a result, this report deals briefly with generalizations about small scale aerial photography as applied to the questions posed in the introduction. Comments on the success of vege-
tation classifications are primarily concerned with the limited Farm Creek data sets. Although much of the information discussed here may be well known to wetland managers and researchers, we have noticed that many potential problems are not mentioned in the literature and it seems worthwhile to document them.

The vegetation map produced in 1973 by S.I. (Welles et al., 1974 and Fig. 4 in this report) was used to identify the vegetation classes on the computer-generated LARS maps. The field data consisted of estimated plant cover over very large areas, so only an unsupervised computer classification was done. Supervised classifications require accurate knowledge of vegetation types and boundaries. Color infrared frames 3648 and 3649 of U2-72-147 (original scale 1:130,000) were classified separately, using seven classes in a clustering program. After class statistics were established for the training fields, the entire digitized area was classified. This approach provided for 6 classes of vegetation, the number on the field map, and a water class. The patterns on the computer printout were compared to those on the 1:12,000 field map, and each computer class assigned a vegetation identification. The vegetation types were named for the plant species which had the greatest amount of cover within the distinguishable unit.

In general, water courses were easily identified as such. Water in shallow ponds and mud flats were often classified as *Spartina alterniflora* or vice versa. *Spartina alterniflora* tends to occur in wet areas (low marsh), and there are often large open spaces of mud or water in the community. The class designations and corresponding vegetation types are fairly consistent for both frames. The most obvious difference between the two frames arises from the switching of classes 4 and 6, which could lead to the rather serious misclassification of *Spartina patens* as *S. alterniflora* or water. Table 1 compares the optical density classes with the corresponding vegetative cover in Farm Creek.
Figure 4.

FARM CREEK MARSH

- **Spartina patens/Distichlis spicata** 80-100%
  - Mud/Water 0-20%

- **Spartina alterniflora** 20%
  - Mud/Water 80%

- **Spartina patens/Distichlis spicata** 0-100%
  - Scirpus Olneyi 0-100%
  - Juncus Roemerianus 0-100%
  - Spartina alterniflora 0-50%
  - Scirpus robustus 0-30%
  - Mud/Water 0-30%

- **Spartina cynosuroides** 100%

- **Juncus Roemerianus** 100%
  - Scattered ponds

- **Trees & Shrubs**
Figure 4.

FARM CREEK MARSH

SCALE 1:1747
Table 1. Comparison of a seven cluster classification of frames 3648 and 3649 of flight U2-72-147 over Farm Creek Marsh.

<table>
<thead>
<tr>
<th>Vegetation Class</th>
<th>Frame 3648</th>
<th>Frame 3649</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spartina patens</td>
<td>3, 4, 5</td>
<td>3, 5, 6</td>
</tr>
<tr>
<td>Spartina alterniflora</td>
<td>6, 7</td>
<td>4, 7</td>
</tr>
<tr>
<td>Water</td>
<td>6, 7</td>
<td>4, 7</td>
</tr>
<tr>
<td>Scirpus olneyi, mix</td>
<td>2, 3</td>
<td>2, 3</td>
</tr>
<tr>
<td>Juncus roemerianus</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

The clustering program combines the optical density values of three separate bands. A key can be constructed, using the separated spectral values for these bands, and thus eliminating the class overlap which is illustrated in Table 1. However, a key which separates the classes for frame 3648 is not suitable for class separation on frame 3649. This indicates that considerable caution should be used in transferring information from frame to frame in the same flight, to say nothing of distance or seasonal data transfers. Although the same class symbol may have been maintained for the vegetation class, the levels of response in the three color bands were not the same. The response variation between the two frames is illustrated for the infrared color band, 0.67μ - 0.88μ, in Figure 5.

Selected small areas of S. patens, S. alterniflora, and J. roemerianus and water were measured on the computer generated map for frame 3649. These were compared to the same areas on the 1973 field map. The computer classification underestimated J. roemerianus and S. patens by 20 - 50%, and overestimated S. alterniflora by 20% and water by 2%. The field maps were
Figure 5. Comparison of classifications using seven clusters on two consecutive frames of U2-72-147, Farm Creek Marsh. Class one has the lowest optical density and class seven the greatest density.
considered "correct"; however, this assumption is not necessarily true. Since there were almost no pure stands in this marsh, it is probable that the differences stemmed from the computer classification placing mixed classes in different categories than the field workers.

In spite of these inconsistencies the overall printout patterns corresponded well with those on the field vegetation map. The classification seems capable of discriminating between two low marsh species, J. roemerianus and S. alterniflora and the high marsh class of S. patens. [Spartina patens always contains a varying percentage of Distichlis spicata, and this class may actually be predominantly the latter species.] There is also good discrimination of a class designated as Scirpus olneyi which contains mixtures of various other species.

Due to the microdensitometer equipment failure, no digitized data was available for either the Bestpitch or Pigeonhouse Creek test sites. Therefore, the color infrared aerial photography (1:3000) obtained over the sites in July 1974 was used to map vegetation types for the Bestpitch transects. This map, considered in conjunction with the summer's field data, illustrates the nature of the problems of vegetation patterning on computer classifications. The map units and boundaries were distinguished by color and texture. The vegetation classes were named for the species with the greatest percentage of cover as determined from field data. Bestpitch was selected for detailed analysis, rather than Pigeonhouse Creek, since time only permitted mapping of one area and Bestpitch contained a larger number of data points. Figure 6 shows the patterning of vegetation types in the Bestpitch area; some of the 30 meter subtransects cross as many as four vegetation types and five vegetation boundaries.
Figure 6. Vegetation patterns at Bestpitch Marsh sampling site. Typed by dominant species cover; photointerpretation of color infrared imagery at 460m.

Spartina patens & Distichlis spicata

Scirpus olneyi

Spartina cynosuroides

Juncus roemerianus

Typha angustifolia or domingensis

Spartina alterniflora

Mixture of species (too small to map individually).

Water and/or Mud

Cloud shadow
A map, whether hand drawn or computer processed, arranges vegetation units in discrete classes. In contrast to this method, techniques are available for arranging the units in some sort of order, in one or more dimensions (Bray & Curtis, 1957, Whittaker, 1967). We tried to test whether an ordination method would produce information which would permit improved identification of the vegetation units; e.g., by indicating an environmental gradient such as moisture or salinity, which could be correlated with specific plant characters which showed up on photographs. Whittaker and Gauch (1972) have shown that a Bray-Curtis ordination is most suitable for quantitative data of the sort we had collected. Accordingly, we selected cover values for six most common species in the Bestpitch transects to be used as importance values in a Bray-Curtis ordination (Bray and Curtis, 1957). Several different end pairs were used as the X and Y values for the 2 dimensional ordinations. Analysis of the July cover values for *Spartina patens*, *Distichlis spicata*, *Juncus roemerianus*, *Scirpus olneyi*, *Typha angustifolia*, and *Kosteletzkya virginica* in the 24 subtransects showed almost no linear ordination tendencies when plotted in two dimensions (Fig. 7). This indicates that in this area there are few predictable physical gradients present which could be used to predict vegetation identifications in the visible unit classes. It is possible that the transects were not long enough to show gradients, or that gradients are not of useful magnitude in such a flat brackish marsh with a total relief of less than 2 meters.

However, the Bray-Curtis ordination of Figure 7 shows some very interesting clustering tendencies, not unlike those which may be seen in a two dimensional "ordination" of classes of spectral bands $0.67 \mu m - 0.88 \mu m$ vs. $0.56 \mu m - 0.67 \mu m$ (Fig. 8). The Bestpitch subtransects are numbered 1 through 24, and each was given a vegetation class designation based on the plant species with the greatest percentage
Figure 7. Bray-Curtis ordination of July cover values for 24 subtransects in Bestpitch Marsh. Species names indicate the most common dominate types in each group surrounded by dotted lines. Numerical values on the ordinate and abcissa are generated output from a computer program and do not necessarily indicate percentage cover.
Figure 8. Spectral values of seven vegetation classes with a response of 0.67-0.88 μm plotted against 0.58-0.67 μm. Digitized data from color infrared imagery of Farm Creek Marsh, flight U2-72-147, frame 3649.
It can be seen that the transects which tend to be in the same cluster usually have the same dominant species.

The positions of the clusters in 2 dimensional space were compared to the clusters from the spectral ordination. It must be noted that the areas from which the clusters are generated (Farm Creek and Bestpitch) are different in character and have slightly different species compositions. The axis values are not comparable, only the positional relationships of the species clusters relative to each other in their respective ordinations. The comparison suggests that a small sample of field data compared in such a manner to spectral data from the same area might be useful in increasing accuracy of vegetation identification.

Patterns formed by species in the marsh are extremely variable both in size and position. The relationship of pattern size to size of a data point (resolution element) will have an effect upon the accuracy of species identification on computer generated maps. For example consider two units, *Spartina patens* and *Juncus roemerianus*, formed into random patterns. If the pattern units are large (i.e., larger than the resolution element for digitized data) then the possibility of misclassification is greatest only at the interfaces of the two types, and there will be a large area inside the boundaries which may be accurately identified. On the other hand, if the size of the units approaches the size of the computer generated resolution element (pixil), and there are many boundaries with mixing of species, then the possibility of misclassification is greatly increased, and any estimates of species areas will be in error.

This situation emphasizes the difficulty of classifying natural vegetation by the pattern recognition methods such as LARS, which are so successful on cultivated plant species. With either crop classification, or...
classification of marsh vegetation, the primary problem is one of distinguishing species which grow in a dense, homogeneous fashion, often with few phenological differences and in areas of low relief. Identification of plant species is not quite so difficult in arid regions, where phenology, plant spacing and geographical relief are more varied. Crops are planted in reasonably ordered arrangements and subjected to predictable, more or less well known cultural practices. It has been suggested that spectral signatures for crop species be determined and these used to classify vegetation in the future (Wiegand et al., 1974). This is probably not presently feasible for native vegetation.

One of the common management practices in wetlands is burning during the winter months. This will change the spectra of a vegetation type in unpredictable ways by removing dead plant material; the amount removed, and where, often depends upon wind and tide conditions at time of burning. A further confounding factor in spectral analysis of wetland data is the tidal characteristics of any given marsh. Tide cycles may be calculated, and lag times for water movement known for the marsh, and appropriate corrections made when time of imagery is known. However, in the Chesapeake Bay region "wind tides" are often of greater magnitude than the lunar tides. For example, on a clear sunny July day a strong southeast wind created standing water to a depth of 15 cm in the ordinarily dry S. patens and D. Spicata areas of Bestpitch Marsh. Imagery taken on that day would have produced data which could have misclassified high marsh species (e.g. S. patens, D. Spicata) as low marsh species (e.g. S. olneyi, J. roemerianus). The wind tide effects on spectra and vegetation classification would be both unpredictable and unquantified.
CONCLUSIONS

We tested LARS data processing programs for remotely sensed data on color infrared aerial photos from a small wetlands area, and performed detailed vegetation ordinations on field data from another marsh site. Small scale (1:130,000) imagery can, after being digitized and computer processed, distinguish between species associations of Spartina patens/Distichlis spicata, Juncus roemerianus, Spartina alterniflora, and mixtures which are predominantly Scirpus olneyi. Estimates made from the computer-produced map of the areas covered by each association are only approximate and may err by as much as 50%. Two consecutive frames of a marsh scene were classified individually. The class designations for vegetation types were the same for both frames, with the exception of the reversal of one of the water classes with one of the Spartina patens classes. The spectral values (spectral signatures) for classes were not the same for both frames, suggesting that standardization of signatures for native plant associations will be difficult.

Analysis of the vegetation field data shows intensive patterning of vegetation associations (units). Because of the number of boundaries involved, and their effect on pattern recognition classifiers such as LARS, misclassification of data points is a greater problem with natural vegetation than with agricultural crops. The number and type of misclassification errors will depend on pattern size and shape, boundary sharpness, the number of vegetation units, and size of the imagery resolution unit.


