General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

E76-10228

TMX - 72993

Copy No. 046

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

TMX

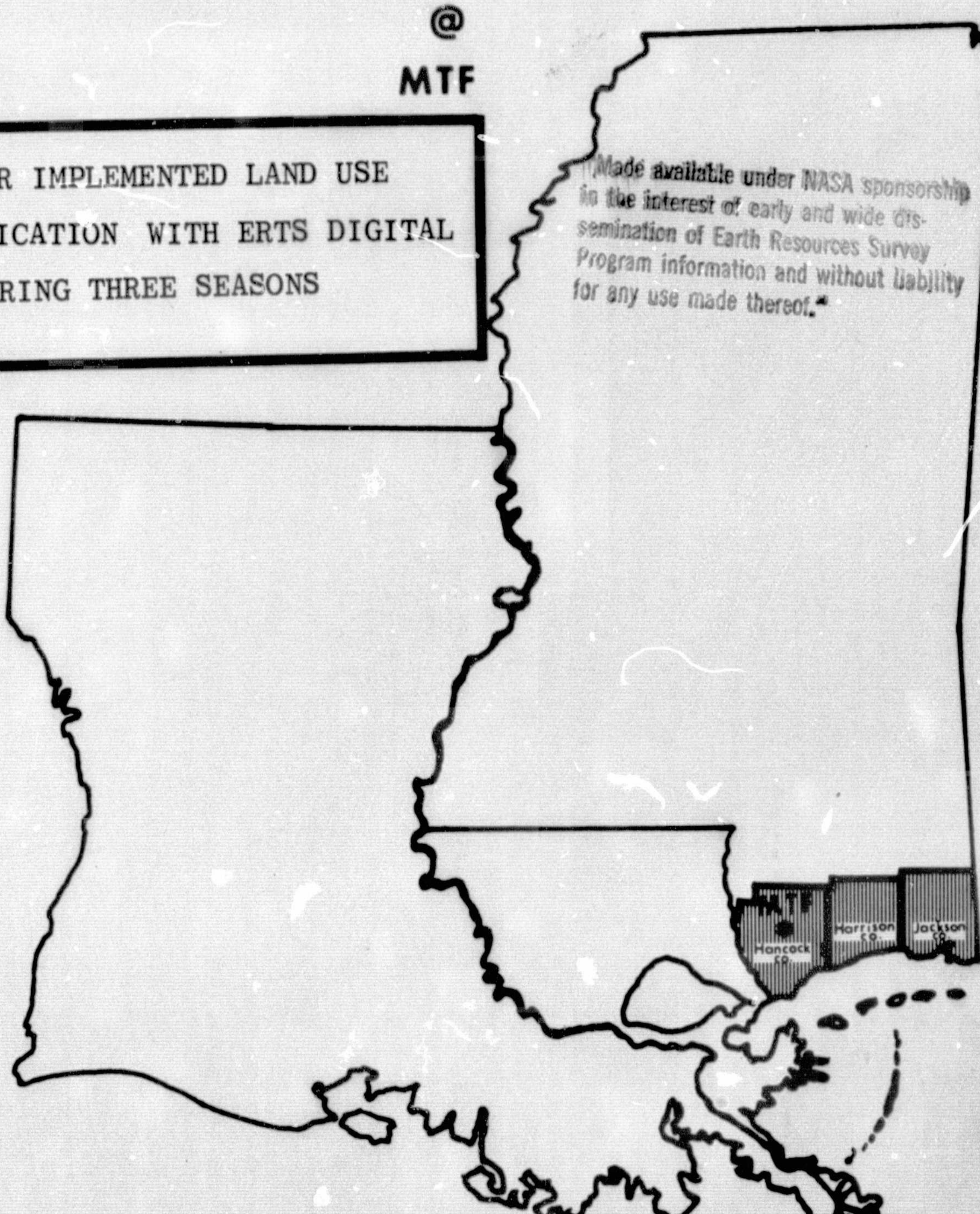
EARTH RESOURCES LABORATORY

@

MTF

COMPUTER IMPLEMENTED LAND USE CLASSIFICATION WITH ERTS DIGITAL DATA DURING THREE SEASONS

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



REPORT NO. 135
JUNE, 1974

LYNDON B. JOHNSON SPACE CENTER

(E76-10228) COMPUTER IMPLEMENTED LAND USE CLASSIFICATION WITH ERTS DIGITAL DATA DURING THREE SEASONS (NASA) 23 p HC $3.50 CSCL 05B
N76-21626
G3/43 Unclas 00228

COMPUTER IMPLEMENTED LAND USE CLASSIFICATION
WITH ERTS DIGITAL DATA DURING THREE SEASONS

By

Armond T. Joyce

Report No. 135

June, 1974

ORIGINAL CONTAINS

COLOR ILLUSTRATIONS

Original photography may be purchased from:
EROS Data Center
10th and Dakota Avenue
Sioux Falls, SD 57198

TABLE OF CONTENTS

I. Introduction.................................... 1

II. Data Processing.................................... 1

III. Data Analysis.................................... 6

IV. Conclusions.................................... 16

LIST OF TABLES AND FIGURES

Figure 1. Computer Derived Land Use Classification of ERTS-1 Data Acquired August 7, 1972, Mississippi Gulf Coast.................... 4

Table 1. Classifications Within Training Sample Areas Expressed as Percentage of Total Cells Representing a Given Surface Feature Classified as Pertaining to that Surface Feature.................................. 7

Table 2. Results of Computer Implemented Classification within Training Sample Areas for Aug. 7, 1972 Data in Percentages........... 9

Table 3. Results of Computer Implemented Classification within Training Sample Areas for Jan. 16, 1972 Data in Percentages.......... 10

Table 4. Results of Computer Implemented Classification within Training Sample Areas for May 4, 1973 Data in Percentages........... 11

ABSTRACT

Significant progress has been made in the classification of surface features (land uses) with computer-implemented techniques based on the use of ERTS digital data and pattern recognition software. The supervised technique presently used at the NASA Earth Resources Laboratory is based on maximum likelihood ratioing with a digital table look-up approach to classification. After classification, colors are assigned to the various surface features (land uses) classified, and the color-coded classification is film recorded on either positive or negative 9 1/2" film at the scale desired. Prints of the film strips are then mosaicked and photographed to produce a land use map in the format desired. Computer extraction of statistical information is performed to show the extent of each surface condition (land use) within any given land unit (e.g. training sample, township, county, etc.) that can be identified in the data. Evaluations of the product indicate that classification accuracy is well within the limits for use by land resource managers and administrators. Classifications performed with digital data acquired during different seasons indicate that the combination of two or more classifications offer even better accuracy.

I. INTRODUCTION

Earth Rescurces Technology Satellite (ERTS) data offers the land use analyst several new dimensions. A single ERTS pass results in the collection of data over a swath approximately 100 nautical miles wide, whereas imagery acquired with mapping cameras flown in aircraft commonly covers swaths from two to fifteen nautical miles. ERTS repetitive coverage on an eighteen day cycle provides possibilities for a rapid detection of cultural changes on the earth's surface as well as seasonal differences in vegetation and land use practices. In addition, the digital form of the data is conducive to automated data processing based on computerized systems.

The objective of the study reported in this paper was to perform computer-implemented land use classifications utilizing ERTS digital data and pattern recognition software for three sets of data, each pertaining to a different season of the year, and to compare the three classifications as to their portrayal of seasonal differences in vegetation and agricultural practices. ERTS digital data acquired over the Mississippi coastal plains on August 7, 1972, January 16, 1973, and May 4, 1973 were selected for the study.

II. DATA PROCESSING

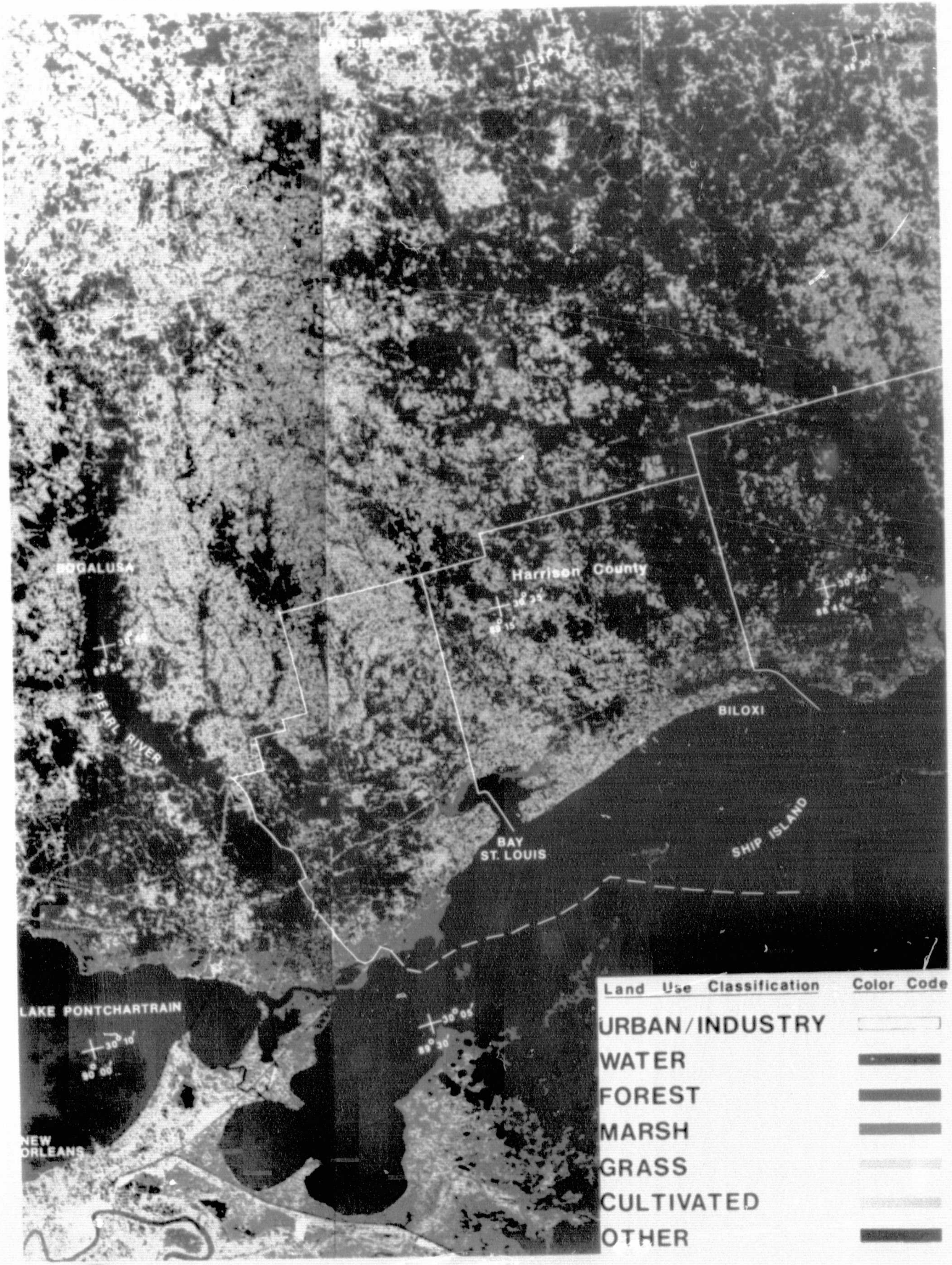
Land use classification at the Earth Resources Laboratory is performed using a Data Analysis Station (DAS) and UNIVAC 1108 software. The DAS consists of a Varian 620f computer with 16,000 16 bit words, two nine track digital tape decks, a color television display device (CRT) with light pen capability, a Singer color film recorder, a card reader and a line printer. The UNIVAC 1108 software consists of several modules which constitute a supervised maximum likelihood classification scheme based on Gaussian statistics. The modules are a statistical module, a training sample separation module, and a classification module. The procedure is described in detail in the report listed as reference No. 1 in the list of references.

The initial stage of data processing consists of reformatting the nine track ERTS computer compatible bulk data tapes received from the Goddard Space Flight Center. The reformatting operation produces a data tape in a format suitable for the 1108 software and a display tape which can be used to drive the DAS CRT or the DAS film recorder. Using the display tapes and the light pen capability of the DAS CRT, the coordinates of surface areas with known land use, called training samples, are determined. These scan line and scan line element coordinates allow the training sample areas to be located in the data in a supervised classification system. Using the training sample coordinates and the reformatted bulk data tapes, the training sample data is extracted and stored on a training sample edit tape which can be used with the UNIVAC 1108 software.

The statistical module on the UNIVAC 1108 is used to compute means and covariance matrices and to plot histograms for each training sample. The information output by the statistical module is used to edit the training sample data and is used for input to the separation module. The separation module computes a measure, "divergence," of the similarity of pairs of training samples. The measure, while quantitative, is difficult to relate to physical processes. However, it is known that the larger the measure the greater the difference between the training samples. ERL uses the measure to determine which training samples can be grouped to form a training class and which training samples cannot be grouped, but must be treated as subclasses. In particular, for the classification of the three subject ERTS frames, the divergences between all training samples which potentially belonged to a single class were computed. Those training samples which had a divergence of less than approximately 15 were grouped into a single subclass. Reference No. 2 describes this component of the pattern recognition software in detail.

As an example of training sample grouping we would consider the class "forest" from the 7 August 1972 data set as shown in Figure 1. The training information for the "forest" class consisted of fifteen training samples identified as pine and twelve samples identified as hardwood. The divergence criteria grouped these training samples into three subclasses of pine and four subclasses of hardwood. Hence, the forest classification was derived from seven forest subclasses. In general, the six class classification was derived from twenty-three subclasses which were three soybean subclasses, one corn subclass, two exposed soil subclasses, two grass subclasses, one pasture subclass, three marsh subclasses, three water subclssses, one urban industrial subclass and the previously mentioned seven forest subclasses. Figure 1, however, is the result of generalizing all twenty-three subclasses classified on computer compatible tape into seven color-coded categories.

FIG. 1 COMPUTER DERIVED LAND USE CLASSIFICATION OF ERTS-1 DATA ACQUIRED AUGUST 7,1972 - MISSISSIPPI GULF COAST



Prepared by: NASA/JSC Earth Resources Laboratory

Mississippi Test Facility Bay St. Louis, Mississippi

Based on the groupings indicated in the previous paragraph, the statistical module was used to generate information used by the classification module to classify the reformatted ERTS bulk data tapes. The classification algorithms based on pre-storing in the computer a representation of each data element and the class to which it is to be assigned. This technique eliminates the need to compute for re-occurring data elements the probability that the data element belongs to each subclass, and the comparison of all such probabilities. The classification algorithm can process one ERTS computer compatible tape in eight minutes. However, the algorithm is limited to twelve classes. Since we used twenty-three classes, two passes were required per data tape.[1] Therefore, four tapes or one ERTS frame requires about one hour to process. The resulting classification is stored on tape as a color-coded classification symbol for each data element.

The classification tape is displayed in color on the DAS CRT and is displayed on the DAS film recorder. When the classification is displayed on the film recorder, rectification allows overlaying the classification data with a map of desired scale. The rectification technique considers scan angle, scan rate, sample rate, V/H ratio of the platform, rotation rate of the earth, and the characteristics of the film recorder. A quantitative evaluation of the rectification has not yet been made, but rectified data has been overlayed with a 1:250,000 scale map on a Traverse Mercator Projection. The match between the rectified data and the map

1. After the classifications for this study were made, a new program has been developed at the Earth Resources Laboratory to increase the number of classes and reduce the processing time. (See reference No. 3)

appears to be very good in a region 25 x 100 nautical miles which corresponds to 1/4 of an ERTS frame and it is expected that the entire ERTS frame will match equally as well.

III. DATA ANALYSIS

The results of the classification within training sample areas shown in Table 1 is indicative of the classification accuracy for the entire three-county test site. Although ground evaluation has not been completed, preliminary findings show that the accuracy of the August and January classifications in areas outside of the training sample areas is not substantially different from the results of the classification within training sample areas. However, the statistics shown in Table 1 are adequate for the purpose of analyzing the data in respect to seasonal differences. In viewing the statistics in Table 1, it is evident that certain surface features were classified more accurately with one set of data than with the others.

The forested areas of the Mississippi coastal plains are mainly pine forests, but there are also large areas covered by swamp forest (mainly water tupelo, bald cypress, and willow) in the bottomlands adjacent to the major rivers. Pine tree foliage is green during the winter season (January through March) at which time most other vegetation is either dead or leafless. Most hardwood trees, as well as bald cypress, are leafless during the winter season, although there are some evergreen brush species in the understory.

TABLE 1. Classifications within training sample areas expressed as percentage of total cells representing a given surface feature classified as pertaining to that surface feature.

Surface Features (Land Use)	Aug. 1972 Data	Jan. 1973 Data	May 1973 Data
All forested	92.4	97.8	96.9
Pine forest	81.4	91.5	98.8
Swamp forest	72.7	88.5	95.7
All cultivated	84.6	84.2	89.7
Soybeans	80.0	--	--
Corn	96.0	--	--
Exposed soil	92.0	--	89.7
Winter Ryegrass	--	89.1	--
Stubble	--	69.8	--
Grass (improved and unimproved pasture)	89.0	80.4	92.5
Marsh (non-forested wetlands)	94.9	67.4	97.0
Water	97.6	98.9	99.4
Inert materials (asphalt, concrete metal, etc.)	94.9	85.6	91.7

The classification results in Table 1 show that the classification within all forest training sample areas was 92.4% for August data, 97.8% for January data, and 96.9% for May data. These results indicate that the forested area was most accurately separated from the non-forest area through use of the data acquired by the satellite during January, although the difference between the January and May classifications is not substantial. It may be noted that the statistics for "all forest" training samples are higher than the statistics for both "pine forest" and "swamp forest" for the August and January data; whereas, for the May data, the statistic for "all forest" is lower than the statistic for "pine forest" and higher than the statistic for "swamp forest." This is not a discrepancy in the statistics, but, rather, indicates that there was more difficulty in separating pine forest and swamp forest from one another with the August or January data than with the May data. This problem is apparent in the statistics that show the detailed results for each season in Table's 2, 3 and 4. The statistics in Table 1 also indicate that pine forest was most accurately classified with the May data. This observation is contrary to the result that was anticipated prior to the implementation of the classification. Prior to performing the classification with all three sets of data, it was thought that the January data would yield the most accurate results for pine forest because pine foliage is green during January at which time most other vegetation is either dead or leafless. Although additional work is needed in order to fully explain the unexpected result, it is thought that the more accurate classification of pine with the May data may be attributed to the fact that plantation grown pine was treated as a spectral subclass separately from other pine for the May

TABLE 2. RESULTS OF COMPUTER IMPLEMENTED CLASSIFICATION WITHIN TRAINING SAMPLE AREAS FOR AUGUST 7, 1972 DATA IN PERCENTAGES

	PINE FOREST	SWAMP FOREST	SOY-BEAN	CORN	EXPOSED SOIL	GRASS	MARSH	WATER	INERT MAT.	UNCLASS.
PINE FOREST	81.4	9.8								8.8
SWAMP FOREST	19.5	72.7								7.8
SOYBEAN	1.1		80.0							18.9
CORN				96.0		4.0				
EXPOSED SOIL					92.0				6.7	1.3
GRASS				2.2		89.0				8.8
MARSH							94.9		0.6	4.5
WATER								97.6		2.4
INERT MATERIALS					0.7				94.9	4.4

TABLE 3. RESULTS OF COMPUTER IMPLEMENTED CLASSIFICATION WITHIN TRAINING SAMPLE AREAS FOR JANUARY 16, 1972 DATA IN PERCENTAGES

	PINE FOREST	SWAMP FOREST	RYEGRASS	STUBBLE	GRASS	MARSH	WATER	INERT MATERIAL	UNCLASS.
PINE FOREST	91.5	6.8							1.7
SWAMP FOREST	8.3	88.5				2.1			1.1
RYEGRASS			89.1						10.9
STUBBLE				69.8	14.0			2.3	13.9
GRASS				10.7	80.4				8.9
MARSH		21.7				67.4			10.9
WATER							98.9		1.1
INERT MATERIAL				0.7	0.7			85.6	13.0

TABLE 4. RESULTS OF COMPUTER IMPLEMENTED CLASSIFICATION WITHIN TRAINING SAMPLE AREAS FOR MAY 4, 1973 DATA IN PERCENTAGES

	PINE FOREST	SWAMP FOREST	EXPOSED SOIL	GRASS	MARSH	WATER	INERT MATERIALS	UNCLASS.
PINE FOREST	98.8							1.2
SWAMP FOREST	2.6	95.7						1.7
EXPOSED SOIL	2.6		89.7					7.7
GRASS				92.5				7.5
MARSH					97.0			3.0
WATER						99.4		0.6
INERT MATERIALS			2.8				91.7	5.5

classification. It is probable that this spectral separation was possible because the pine grown in plantation form on the Mississippi Coastal plains is young, vigorously growing pine which puts forth profuse flushes of new foliage growth during the spring season. Consequently, the spectral difference between plantation pine and older naturally grown pine is more pronounced during a short spring period than during other times of the year.

The statistics in Table 1 show that swamp forest was most accurately classified with the May data. Statistics relative to the May classification in Table 4 show that a small percentage of swamp forest was classified as pine, apparently resulting from a spectral similarity between new leaves on the swamp forest trees and new flushes of foliage growth in young plantation grown pine. However, in the August classification results shown in Table 2, there is a larger percentage of misclassification between swamp forest and pine; and, in the January classification results shown in Table 3, there is a larger percentage of misclassification between swamp forest and marsh (non-forested wetlands). In the case of the former, it appears that, by August, the new flushes of spring foliage in the plantation grown pines have changed to become more spectrally similar to the foliage on the older pine trees, and that by August, changes have occurred in the swamp forest foliage so as to cause spectral overlap with the pine forests. During January, the swamp forest trees are leafless and should be spectrally distinct from vegetation with green foliage, but are spectrally similar to dead foliage. Statistics relating to the January classification in Table 3 show a significant misclassification between the leafless swamp forests and dead foliage of marsh vegetation.

During August 1972, on the Mississippi coastal plains, the main agronomic crop was soybean, although there was some corn and some exposed soil in cultivated areas. During January, some of the cultivated area contains winter ryegrass in a green, growing condition; and the remainder of the cultivated area contains stubble (dead soybean or corn stalks) or dead annual weeds. Improved permanent pastures and unimproved native grass areas contain dead grass foliage. During May most cultivated areas have been plowed or planted, but are essentially exposed soil. Pasture grasses and native grasses are in a green, growing condition during May.

Statistics in Table 1 show that there was no significant difference between the classifications made with the August and January data within all training sample areas corresponding to the "cultivated" category. The results show 84.6% and 84.2% respectively. During August, all vegetation is in a green, growing condition. Statistics in Table 2 show that there was some misclassification between corn and grass, and some betweeen soybeans and pine. During January, winter ryegrass is the only cultivated crop that occurred in a green growing condition, but dead vegetative material occurs in the pasture and marsh areas at the time that the swamp forest trees are leafless. As seen in Table 3, the principle misclassification within training samples corresponding to the cultivated category occurred between stubble and dead grass. Table 1, for the May data, shows 89.1% classification accuracy for the cultivated category, surpassing the accuracy of both the August and January data. As is evident, the statistic for the cultivated category resulting from the May classification corresponds to the statistic for exposed soil inasmuch as only exposed soil occurred within training sample areas applicable to the cultivated category. During May, all cultivated areas are in some stage of soil preparation or are planted, and do not contain any significant amount of growing plant material.

Marsh vegetation (nonforested vegetated wetlands) is in a vigorously growing state during August; whereas, except for a few evergreen brush species in some areas, marsh areas contain dead plant material during January. During May, even though the marsh vegetation typical of the spring and early summer is present, this green, growing vegetation is still overtopped by the dead material remaining from the previous growing season. Consequently, even though the statistics in Table 1 indicate that the May data yielded the most accurate classification, this is only true in respect to all marsh vegetation viewed as one category. If an attempt was to be made to classify various species associations of marsh vegetation separately from one another, then, it would be more reasonable to attempt such a classification with August data rather than with May data. The use of May data apparently allows a separation of all marsh vegetation treated as one class from all other non-marsh categories, because, during May, areas outside of the marsh contain either green, growing vegetation or exposed soil both of which are spectrally dissimilar from dead plant material overtopping the spring growth of marsh vegetation.

The low accuracy of the marsh classification with January data can be attributed to a spectral similarity between dead marsh vegetation and leafless swamp forest with a flooded condition.

A computer-implemented classification, as used in this study, is based on separating surface conditions that have different spectral characteristics caused by differences in reflected energy as measured from above. Urban, commercial, industrial, and residential land uses can be separated from the other land uses only inasmuch as their surface conditions with inert materials (asphalt, concrete, metal, wood, etc.) are spectrally different from vegetation or other material (sand, water, etc.) Residential or urban areas that have foliated trees overtopping the buildings as well as lawns and shrubs occupying surface areas as seen from above are likely to be classified as vegetation. Such was the case for the August classification. It is apparent when viewing the Biloxi and New Orleans areas in Figure 1 that only the urban and industrial centers that are essentially devoid of all vegetation were classified as pertaining to urban and/or industrial land uses. However, the color assigned to inert materials (asphalt, concrete, etc.) and the colors assigned to other surface conditions (grass, trees, etc.) that may be in the urban environment will form color patterns on a color presentation that can be interpreted so as to enable delineations of urban areas, especially to separate urban commercial and industrial centers with large concentrations of inert surface materials, residential areas with associated vegetation, and other land uses. In this context, the classification results within training sample areas is meaningless to the accuracy of the classification of the total urban, commercial, industrial, or residential area. However, inasmuch as the hardwood trees that overtop one-story or two-story buildings are leafless and lawn grasses are dead during January, a larger portion of the urban areas outside of training sample areas was classified as having inert material (asphalt, concrete, etc.) on the surface when utilizing ERTS data acquired in January than when utilizing data acquired in August or May when all trees have green foliage.

IV. CONCLUSIONS

Analysis of the results of computer implemented classifications within training sample areas for three sets of data corresponding to three distinct seasons of the year indicate that certain surface features (land uses) can be classified most accurately with season specific data. Although the ground evaluation of the May classification has just begun, the classification results for training sample areas indicate that a classification carried-out for the purpose of classifying broad generalized categories such as "marsh," "cultivated," "forest," "grass," and "water" could be accomplished best with data acquired in early May rather than August or January. However, inasmuch as there is considerable variation among the three classification results for any one surface feature, it appears that an integration of two or more sets of seasonal data would yield the most accurate surface classification. Cloud-free data from the fall season was not available for inclusion in this study, but it is possible that certain surface features, such as swamp forest trees which are in a state of foliage color change during fall, could be most accurately classified with data acquired at that time.

On the basis of the three sets of data used for this study, it appears that pine forest could be most accurately separated from other surface features with data acquired during January (typical of the winter season) or May (typical of the spring season) for the Mississippi coastal plains. However, it is thought that the time "window" for data acquisition during the spring season is likely to be much shorter than for the winter season.

The three sets of data used in this study show swamp forest was most accurately classified with the May data.

Marsh vegetation (non-forested wetlands) as a category was most accurately classified with May data, but, if an attempt to classify individual species associations within the marsh were to be made, this could be done better with August data.

Classifications of grass areas and cultivated areas as a category were performed most accurately with May data. However, if classifications of individual agronomic crops were desired, such classifications could only be performed for corn and soybeans with August data and for winter ryegrass with January data.

These data indicate that there is no significant difference among the three sets of data in respect to the classification of open water bodies.

Surface features with inert materials such as asphalt or concrete in urban-industrial areas can be identified most accurately with January data when those materials may be only overtopped by leafless trees and when grasses or shrubbery in yards is dead.

It is important to note that the procedure used in this analysis does not take account of the extensiveness with which each surface material occurs in a given area. Therefore, these conclusions are meaningful to potential classification accuracies for large area classifications to the degree that one accounts for the importance or general geographic extent of each surface feature within a given area.

LIST OF REFERENCES

1. Whitley, S.L. 1973. A Procedure for Automated Land Use Mapping Using Remotely Sensed Multispectral Scanner Data. NASA, JSC, Earth Resources Laboratory, Bay St. Louis, Ms. Report No. 081.

2. Jones, W.C. 1973. Choice:36 Band Feature Selection Software with Applications to Multispectral Pattern Recognition. NASA, JSC, Earth Resources Laboratory, Bay St. Louis, Ms. Report No. 059.

3. Jones, W.C. 1974. Implementation of an Advanced Table Look-up Classifier for Large Area Land-Use Classification. NASA, JSC, Earth Resources Laboratory, Bay St. Louis, Ms. Report No. 115.