N73-28383

UNSUPERVISED CLASSIFICATION AND AREAL MEASUREMENT OF LAND AND WATER COASTAL FEATURES ON THE TEXAS COAST

L. M. Flores, C. A. Reeves, S. B. Hixon and J. F. Paris, *Lockheed Electronics Company, Inc., Houston Aerospace Systems Division*

ABSTRACT

ERTS-1 multispectral scanner (MSS) digital data was used to delineate coastal land, vegetative, and water features in two portions of the Texas Coastal Zone. ERTS-1 MSS data (Scene ID's 1037-16244 and 1037-16251) acquired on August 29, 1972, were analyzed on NASA Johnson Space Center systems through the use of two clustering algorithms. Seventeen to 30 spectrally homogeneous classes were so defined. Many classes were identified as being pure features such as water masses, salt marsh, beaches, pine, hardwoods, and exposed soil or construction materials. Most classes were identified to be mixtures of the pure class types. Using an objective technique for measuring the percentage of wetland along salt marsh boundaries, an analysis was made of the accuracy of areal measurement of salt marshes. Accuracies ranged from 89 to 99 percent. NASA aircraft photography was used as the basis for determining the true areal size of salt marshes in the study sites.

1. INTRODUCTION

The purpose of this paper is to report on the application of two classification algorithms to ERTS-1 multispectral scanner (MSS) system corrected data acquired over the Galveston Bay (Texas) area on August 29, 1972 (Scene ID's 1037-16244 and 1037-16251). Specifically, two algorithms (ISOCLS and NSCLAS) available at the NASA Johnson Space Center (JSC) were used to produce classification symbol maps of portions of the Texas coastal zone. The resulting classes then were related to an information hierarchy selected for the MSC ERTS-1 coastal estuarine investigation.

HIERARCHY OF COASTAL FEATURES

Level 1	Level 2	Level 3
Water	Water Mass Features	Turbid water Productive water Nonproductive water
	Bottom Features	Vegetated Nonvegetated
Nonwater	Sand (beaches, spoil)	
	Wetland	Vegetated Open water Nonvegetated
	Other	

Photography obtained by the NASA WB-57 aircraft on August 30, 1972, was used to aid in the identification of the classes defined by the algorithms and to provide a basis for the measurement of the true area of selected features such as water bodies and marsh.

2. METHOD OF ANALYSIS

Two study sites were edited from ERTS-1 MSS system corrected data contained in frames 1037-16244 and 1037-16251. These sites covered (1) Trinity River Delta area including salt, brackish, and fresh water marshes, and (2) Galveston Island, West Bay, Pelican Island, and offshore Gulf waters.

One computer compatible tape (CCT) containing the Galveston data was sent to Purdue University. This data set could then be analyzed through the use of the remote terminal at JSC. Both sets could be analyzed by JSC systems.

Since the two study sites exhibited natural features of complex configuration and unknown spatial extent, it was decided to use classification algorithms to produce classification symbol maps of the sites. Each of the two algorithms (ISOCLS and NSCLAS) attempts to partition the data into spectrally homogeneous groups. They then generate classification maps to display the geographical regions in which these homogeneous groups of data occur. A detailed description of ISOCLS can be found in references 1 through 5. A description of NSCLAS is given in reference 6.

Photography (film 2443, color IR) of the study sites was obtained by the NASA WB-57 aircraft from 60,000 feet (MSL) on August 30, 1972, one day after the subject ERTS-1 data. The photography was rectified and enlarged to a scale of 1:41,200. Then, wetland areas were delineated by coastal analysis team members. The area of each delineation was determined through the use of a photographic data quantizer (planimeter). Field trips into the sites were made to aid in the identification of classes defined by ISOCLS in that site. Photo interpreters identified classes in both sites through the use of the aircraft photography.

In the case of the areal measurement of the wetland areas in the Galveston site, an objective technique described in reference 7 in this symposium was used to count picture elements and partial picture elements on the boundary of the marsh.

3. RESULTS

Space does not permit the inclusion of all of the color coded classification maps produced during the present analysis effort. Examples of level two maps in each site are shown in figures 1 and 2 (originals are in color). In figure 2, four classes are shown: water, forested areas, vegetation including marsh, and bare soil/beach/asphalt/concrete. In figure 3, three classes are shown: water, wetland (marsh), and other.

4. WETLAND ANALYSIS

Using the objective technique described in reference 7 (this symposium), the areas of three wetlands were determined. The results are given in table I.

TABLE I.– RESULTS OF DETERMINING AREA OF WETLANDS FROM ERTS-1 DATA THROUGH AN OBJECTIVE TECHNIQUE

WETLAND NUMBER	ESTIMATED AREA (ACRES) ISOCLS	ESTIMATED AREA (ACRES) NSCLAS	TRUE AREA (ACRES)	PERCENT ACCURACY ISOCLS	PERCENT ACCURACY NSCLAS
1	864.13	871.56	866.0	99.6	99.4
2	197.97	193.77	191.3	96.5	98.7
3	150.86	165.94	149.6	99.2	89.1

The accuracies in wetland (marsh) area measurement are quite high. Both algorithms accurately define wetland features and boundary classes. It is not possible at this time to determine the more superior algorithm; however, ISOCLS on the UNIVAC 1108 was able to handle a larger number of data points than NSCLAS on the IBM 360/67 at Purdue University.

5. MIXTURE PROBLEM

A graph depicting the position of cluster means in the first and fourth channels of the ERTS-1 MSS is shown for the Trinity site in figure 4 for a

30-cluster case, which will be the subject for the remainder of this discussion. This figure is a good representation of the data and of the results of the cluster analysis. With few exceptions the second and third channels of the MSS add little new information.

The clusters obtained here are, for the most part, mixtures of some of the major classes that exist on the surface. The remaining clusters can be uniquely related to these major classes. The following table shows those clusters that can be so related:

Cluster Identification	Class
B,R,D,N,S	Vegetation excluding trees
C	Pine Trees
4	Hardwoods
F	Beach/asphalt/concrete
U,K	Bare soil
1,G,H,L,S	Water

The clusters have been identified by assigning to them consecutive arbitrary symbols 1, 2, ..., 9, A, B, ...

Up to five clusters can be identified as belonging to water bodies. No quantitative relationship can be assigned to these clusters at this time because there was no concurrent water survey on the date of the ERTS pass. The following is a general description of the type of water clusters and where they occur:

Cluster Identification	Class or Description
1	Bay water (probably slightly turbid)
G	Turbid (fresh or saline) water
H	Highly turbid water, very shallow water, or underwater spoil banks
L	Fresh water that occurs in lakes
S	Turbid fresh water or turbid water with a small amount of vegetation (less than 10 percent).

The clustering analysis differentiates forested areas from other types of vegetation. Within the forested areas, pine trees and hardwoods can also be differentiated and assigned to separate clusters. Vegetation types other than trees cannot be separated. Specifically, the same combination of clusters occur over vegetated areas which are known to contain different classes such as Phragmites communes, Spartina spartinae, and pasture land.

One conspicuous fact that this investigation has brought forth is that a good part of the data can be related to a mixture of two or more classes found on the surface. In the case of wetlands area, the proportion of classes contributing to a particular mixture varies in a continuous fashion over the possible range. Under these circumstances there are no natural groupings of the data; so the clustering program tends to partition the data at arbitrary places.

It may be noted that data assigned to clusters which represent a mixture of classes, can be obtained by linearly combining data assigned to clusters identified with one of the major classes. This fact may be easily seen by inspecting figure 4 where, for example, data points in cluster 8 can be obtained by linearly combining data points in clusters 5 and C. In this manner, a range of ratios of a particular mixture can be extracted from the data. The following table represents this range for the water/vegetation mixture. These numbers represent an estimate only and are not the result of a statistical study.

G

Cluster Identification	Ratio of Water
6	.6-.95
Q	.5-.8
3	.3-.7
E	.2-.6
2	.1-.5
7	0-.35
5	0-.20
N	0-.10
R	0-.05

Other mixtures could consist of almost any possible combination of the major classes that have been identified. The following table presents a list of possible mixutres that have been verified:

Cluster Identification	Mixture
4	Hardwoods; pines and vegetation (excl. trees)
C	Pines; hardwoods and water
0	Dense pines; trees and water
8	Pines and vegetation (excl. trees); hardwoods vegetation (excl. water) and water
6,Q,3,2,7	Marsh; vegetation and water
I,9	Water and bare soil/beach/asphalt in increasing proportion of bare soil/beach/asphalt
P, A, T, M, J	Vegetation and bare soil/beach/asphalt/concrete in increasing proportion of bare soil/beach/asphalt/concrete
E,U,K	Vegetation, water, and bare soil/beach/asphalt/concrete in increasing proportion of bare soil/beach/asphalt/concrete

6. CONCLUSION

Two clustering algorithms were used to produce classification maps of the Trinity River delta area (Scene ID 1037-16244) and the Galveston area (scene ID 1037-16251) from data obtained on August 29, 1972. Seventeen to thirty classes were defined. Most of these classes represented mixtures of land, water, and vegetation including forest. Through the use of a cluster diagram, ranges of percent of water and wetland were assigned to some classes. Improvements would result if a statistical approach were used for the mixture problem.

7. RECOMMENDATIONS

Further analysis is needed to determine whether the position of cluster centers in the cluster diagram can be used to identify the cluster as representing a certain class of coastal features.

REFERENCES

1. Ball, G. H., and Hall, D. J.: A Clustering Technique for Summarizing Multivariate Data, Behavioral Science, Volume 12, 1967.

2. Ball, G. H., and Hall, D. J.: ISODATA, A Novel Technique for Data Analysis and Pattern Classification. Technical Report, Menlo Park, California, Standard Research Institute, May 1965.

3. Kan, E. P. F.: "The Latest Version of ISODATA(A)/ISOCLS", Lockheed Electronics Company, Inc., HASD, Houston, Texas, Technical Memorandum TM642-570, September 1972.

4. Kan, E. P. F.; and Holley, W. A. : "More on Clustering With Final Recommendations on ISODATA," Lockheed Electronics Company, Inc., Houston, Texas, Technical Rep. 640-TR-112, May 1972.

5. Minter, R. T.: "ISOCLS, Iterative Self-Organizing Glustering Program", Lockheed Electronics Company, Inc., HASD, Houston, Texas, Computer Program Documentation CPD202, October 1972.

6. Swain, Pitt.: "Pattern Recognition: A Basis for Remote Sensing Data Analysis", LARS Information Note 11572, West Lafayette, Purdue University.

7. Cousin, S. B., Anderson, A. C., Paris, J. F., Potter, J. F.: "Significant Techniques in the Processing and Interpretation of ERTS-1 Data", Lockheed Electronics Company, Inc., HASD, Houston, Texas.

G

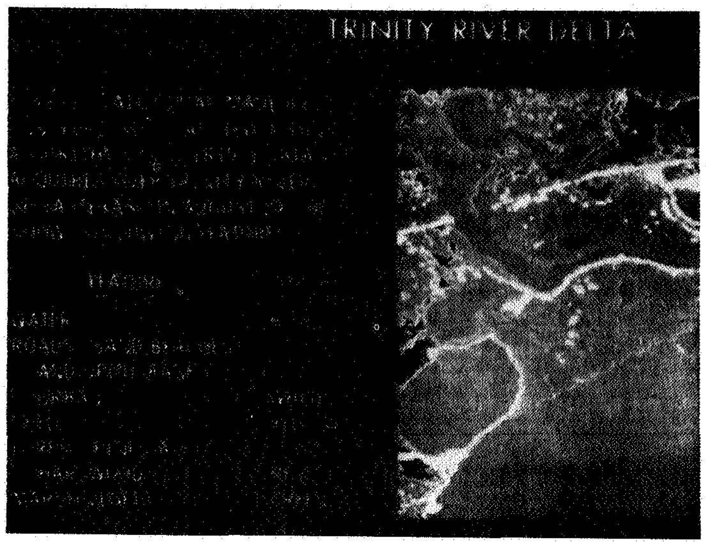
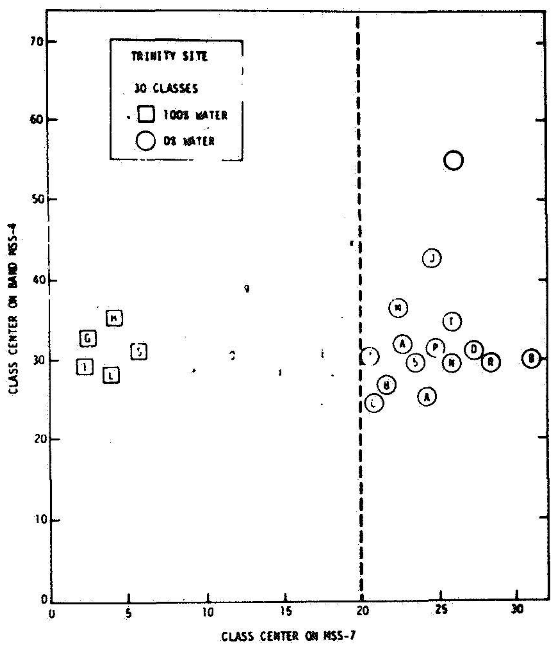


Figure 1



Location of class centers for Trinity Delta.

Figure 2

Figure 3

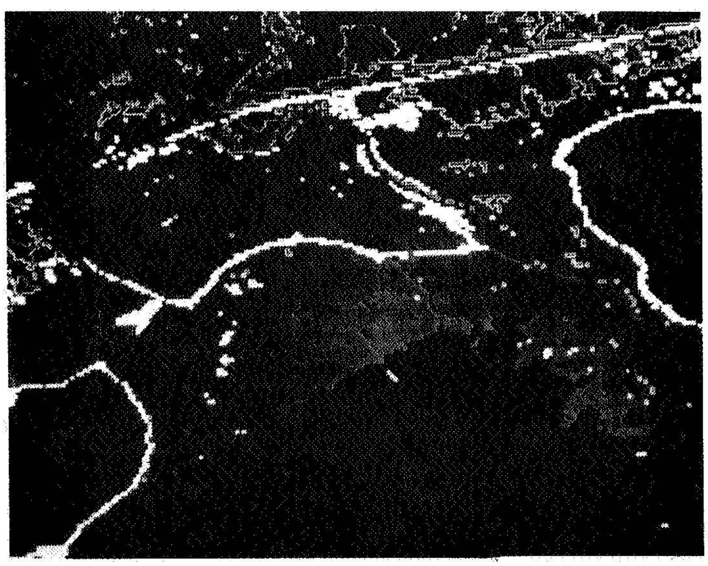


Figure 4