N73-28329

Paper I 8

SIGNIFICANT TECHNIQUES IN THE PROCESSING AND INTERPRETATION OF ERTS-1 DATA

S. B. Cousin, A. C. Anderson, J. F. Paris and J. F. Potter, *Lockheed Electronics Company, Inc., Houston Aerospace Systems Division*

ABSTRACT

The discipline oriented investigations underway at the Johnson Space Center (JSC) using ERTS-1 data provide an appropriate framework for the systematic evaluation of the various elements comprising a prototype multispectral data processing and analysis system. In particular such a system may be thought of as the integration of (1) a Preprocessing Subsystem, (2) a Spectral Clustering Subsystem, (3) a Correlation and Classification Subsystem, (4) Mensuration Subsystem, and (5) an Information Management Subsystem (see fig. 1). Specific elements of this system are already operational at JSC.[1]

It is in the context of this system that technique development and application is being pursued at JSC. Aircraft, ERTS and EREP data will be utilized to refine the subsystem elements for each of the data acquisition systems or system combinations that are optimally suited for a specific Earth Resources application. The techniques reported here are those that have been developed to date during the utilization of ERTS-1 data in this processing and analysis system.

1. THE PREPROCESSING SYSTEM

The Preprocessing Subsystem accepts the multispectral data from the Data Acquisition Subsystem and incorporates the measured sensor and platform characteristics (e.g., detector calibration, platform navigational parameter data, etc.) to yield absolute radiance values for each sensor resolution element. Ideally, this subsystem would include an atmospheric correction scheme to relate the sensor readings to absolute scene reflectance values. Correlation to ground control points and reference to UTM coordinates can also be accomplished at this stage. At the very least an overlay grid to allow easy reference to specific scan line and pixel numbers could be introduced by this subsystem. In the case of the ERTS-1 multispectral scanner (MSS) initial preprocessing is done by the Goddard Data Processing Facility. Since the MSS is constructed with six detectors in each of its four spectral bands, twenty-four distinct calibrations must be accounted for in this preprocessing. The detectors are swept by a calibrated source on every other retrace scan and the detector outputs are equated to the pre-measured calibration source radiance. This calibration data is continuously averaged, so that slow changes in the detectors can be calibrated out of the scene data. The success of this phase of preprocessing is dependent on (a) precise pre-measurement of the calibration source radiance, (b) accurate knowledge of the source radiance as a function of time during the calibration scan, and (c) conformance of the operating source to the prelaunch measured performance. Striping which has been observed in all four channels of the ERTS MSS data can be attributed to such correction inaccuracies. This striping is particularly evident when the scene is a large, homogeneous area such as a lake. Although the striping is due to differences in detector outputs amounting to between one and five quantum levels in the data, it is sufficient to affect sensitive spectral clustering algorithms as shown in the cluster map of figure 2.

A large rectangular area of Lake Livingston, Texas, was selected to study the statistics of the striping in the data (Frame 1037-16244) and to develop a striping removal technique. It was first established that the best available estimates for the correct average data values for water in the four ERTS channels are the composite averages given by the detectors in each channel. If each detector's average output for water were moved to the composite average, the stripes would be removed and the composite average unchanged. A stripe removal procedure was therefore developed to suitably correct the recorded ERTS integer values. Twenty-four accumulator registers were established in the processing computer. Each time a new reselm was considered, the quantity $(A_T - A_i)$, representing the established difference between the *ith* detector average and the composite average, was added into the appropriate accumulators. Whenever an accumulator absolute value exceeded 0.5 the reselm data value for that detector was changed by the integer value closest to the current accumulator value. The accumulator was then updated by an equal but opposite amount. If the new data value was negative or greater than 127 (63 for the infrared channel) it was moved back into the allowable data range and the accumulator was corrected accordingly. In this procedure, offset correction errors are reduced using a portion of the scene itself as a calibration target. The procedure was then applied uniformly to all the data in the frame. The smoothed data was processed once more using the clustering algorithm and resulted in the cluster map shown in figure 3. The reduction in striping is clearly evident. Complete preprocessing of the data should also include a correction for gain calibration error (stripes attributable to inaccurate gain corrections have also been observed). Such a procedure, which will be important in high reflectance target scenes, is now in the process of being developed. The impact of these striping effects must be assessed by each investigator to determine if such smoothing procedures aid in his application.

The atmosphere can make a significant contribution to the apparent signature of a target viewed from space. Since the atmospheric aerosol content changes from place to place, identical targets located at different places may classify differently in automatic pattern recognition studies. The same is true for identical targets viewed at different times. Figure 4 (ERTS Frame 1038-16303 over Lake Somerville, Texas) shows an example of such unwanted atmospheric effects. The programs described here (PREPS-ROTAR) correct the sensor response data by converting the data to target reflectance, a target parameter that is independent of the atmosphere.

The atmospheric correction process in use at JSC consists of two program modules. The first module, PREPS is used to pre-calculate a family of curves relating sensor output to target reflectance. The PREPS program is based on the following assumptions: (a) the sensor is pointing in the nadir direction, (b) the target is a Lambert reflector whose reflectance is constant across a given ERTS/MSS band, and (c) the atmosphere consists of two homogeneous layers, a Rayleigh scattering molecular layer on top and a Mie scattering aerosol layer near the earth's surface. At present, calculations have been made for an aerosol layer with the characteristics of a continental type haze. Other haze models will be included in the near future.

Within the limits of the above assumptions, accurate numerical solutions to the radiative transfer problem are obtained. These solutions, which include all orders of multiple scattering, give the radiance at the sensor as a function of wavelength, target reflectance, haze level in the atmosphere and solar zenith angle. They are then normalized, multiplied by the sensor response function and integrated across the sensor bandwidth to obtain the sensor output as a function of target reflectance using the atmospheric haze level, solar zenith angle, instrument mode (compressed or linear) and instrument gain (high or low) as fixed parameters.

For any MSS frame to be corrected, the appropriate solar zenith angle, instrument mode and instrument gain are selected. A "response tape" is produced for each desired set of these parameters. The response tape contains curves (i.e., tables), which relate instrument response to target reflectance for a given haze level. The haze level is specified by giving the haze optical depth τ at wavelength 0.5 μm. The optical depth at other wavelengths is determined by the haze model. A typical set of response curves is shown in figure 5. The three curves shown correspond to $\tau = 0.0$, 0.424, and 0.848.

To correct each reselm in the MSS frame the second module (ROTAR) is used. The input to the ROTAR program consists of the appropriate response curves and the ERTS data to be corrected. The user can divide the frame into rectangular areas and assign values of τ to each area. ROTAR generates a "correction curve" for each optical depth input to the program. The curve relates the instrument output to the target reflectance for each specific value of optical depth. It is obtained by interpolating (or extrapolating) to the designated values of τ

from the input response tables. Each data point is corrected by determining the target reflectance from the appropriate correction curve. The corrected data can then be analyzed with pattern recognition programs. Figure 6 shows a LARSYS grey scale map of data from the Somerville frame corrected by ROTAR. The numbers represent the percent reflectances of the picture elements in that channel.

A capability has been developed to measure the aerosol optical thicknesses at various positions in the Houston Area Test Site (HATS) during each ERTS overpass. This is accomplished by the use of PREPS photometers whose calibrated response yield the optical thickness values. These values are then used in performing the atmospheric corrections. At this point additional effort is required to develop this capability to a reliable operational system. The PREPS/ROTAR calculations have been compared with target reflectances measured on the ground. The results, which are shown in figure 7, appear accurate to within two percent. (Note: The allowable range of target reflectance is 0 to 100 percent of the Lambert perfect reflector value.) These results are under more detailed error analysis at this time. Both the ground reflectance and the aerosol optical depth measurements are subject to known systematic errors and will require further analysis for final verification.

Additional development in this preprocessing subsystem is therefore required in correcting for the gain stripping effect, in reference grid overlaying, and in further testing of the atmospheric correction scheme.

2. THE CLUSTERING SUBSYSTEM

The preprocessed ERTS data represents a class of multivariate data for which clustering subsystems have been under development for several years. Image enhancement techniques may be used to bring out subtle spectral variations or data groupings will be clustered to aid in the selection of "training fields" for use in the classification subsystem. An example of such a technique is the ISODATA algorithm.[3] It is an iterative technique for grouping each multivariate data point with other similar data points. On each iteration, all of the data points are assigned to one of the existing clusters and each cluster is examined to determine if it should be split, left alone, or combined with some other cluster. A revised set of clusters is generated and the process is repeated until the resulting clusters stabilize or some other criterion set by the investigator has been satisfied (i.e., maximum number of iterations or clusters). The fundamental assumption in the algorithm is that the data is comprised of well separated clusters of similar points. That is, there are regions in the multivariate space where the data points are dense along with areas between those regions where the data points are spare (refer to figure 8a in this discussion). The ISODATA version in use at JSC[4,5] assigns points to the existing clusters by a "city block" distance measure. Each cluster is defined by the location of its center point. The distance from a data point to a cluster is defined as the sum of the distances parallel to each of the coordinate axes. While points are being assigned to cluster, information is accumulated for recalculating cluster centers and the standard deviations in each variable for each cluster. The splitting and combining decisions are made on the basis of these standard deviations and the inter-cluster distances. If the standard deviation in one of the variables becomes larger than a fixed input value (STDMAX), the cluster is split into two clusters with centers at plus and minus one standard deviation from the old center. If clusters are too close together, they are combined and a single center is computed.

The ERTS data, recorded as integers, are similar to the results of a Poisson process which is, typically, the counting of discrete events during a fixed time period (see figure 8b for the low reflectance target analogy). For bright targets, the ERTS data appear to be normally distributed about a mean value. For mean values above 10, however, the Poisson distribution is indistinguishable from the normal distribution. The useful features of the Poisson distribution is that once the mean value is known, everything else about the distribution is known. The probability of occurrence of each integer count is known; the standard deviation is known. There is no such relationship for normal distributions. Using these facts, the split criterion in ISODATA can be established as "splitting occurs if one of the standard deviations becomes larger than the square root of the mean for that channel in that cluster." In practice, the allowable standard deviation is made proportional to the square root of the mean with the constant of proportionality selectable by the investigator. Values of the proportionality constant of 0.8 and 1.0 have been tested and yield reasonable clustering results. The criterion for combining clusters was also modified to be consistent with this splitting criterion. Surfaces derived from the standard deviations in each cluster were investigated by entering a t value, 1.0 for 1σ, 2.0 for 2σ, etc. The algorithm assumes that the t surface is an ellipsoid in the multivariate space and is oriented with its axes parallel to the coordinate axes. It then tests the surfaces to see if they intersect on the line joining the cluster means. If they do, the clusters are combined.

Results obtained in the testing of this algorithm to date (figs. 2, 3, and 4b are examples) indicate its flexibility and its capability to generate meaningful cluster results regardless of the range of the data values considered. Additional testing of this technique is required to assess its utility in various applications.

3. THE CORRELATION AND CLASSIFICATION SUBSYSTEM

This subsystem associates material or feature designations to each set of pixel energy responses through the use of ground information available about the scene being processed or by the comparison of these responses with predetermined material spectral signatures. This can be accomplished in either a supervised classification mode, by developing classification criteria based on the "training field concept" successfully utilized by LARS at Purdue University, or in an unsupervised classification mode, by relating spectral cluster members to known, correlatable scene features, or by associating the cluster center to historically established spectral signatures. To aid in ground correlations, techniques are under development at JSC to locate specific ERTS resolution elements accurately on aerial photographs. One technique, being developed specifically in urban scenes, consists of relating specific reselms clustered as extremely high reflectance responses to the specific ground features known to cause such a response. When several appropriately spaced objects are positvely correlated with pixels in this way, a relative scale between the aerial photography and ERTS cluster maps can be established. Such geographic correlation, also being explored using high contrast boundaries in other than urban scenes, can be used as the basis for establishing polynomial pixel translation functions to allow the direct registration of the classification or clustering maps onto rectified large-scale photography or base reference maps. Once such techniques can be refined, detailed analysis of the spectrally clustered data can yield estimates to the proportionate composition of resolution elements in heterogeneous areas. One such analysis has already been performed using a land/water interface as the correlation mechanism and the comparison of calculated water area to the true area as a measure of success.

Data collected over the Galveston Bay (Texas) area on August 29, 1972 (Scene ID's 1037-16244 and 1037-16251) were analyzed through the use of the clustering algorithm. Nineteen water bodies ranging in area from 2.8 to 607.2 acres were chosen for analysis. NASA aircraft photography of the area from 60,000 feet was used as a basis for determining the true area of each of the nineteen water bodies. An examination of the locations of the cluster/classification centers in two ERTS-1 channels (MSS bands 4 and 7) as shown for example in figure 9, revealed that the non-water classes (circled symbols) lie above a threshold of 20.0 in MSS band 7. The pure water classes (boxed symbols) lie close together near 2.8 in MSS band 7. Thus, the other classes can be assummed to be simpled mixtures of water and non-water classes such that a percentage water amount can be assigned to each mixture class based on a distance measure between the two pure classes.

By considering only those mixture classes that were spatially adjacent to pure water classes, the total number of picture elements representing each of the nineteen water bodies was determined. A linear regression analysis of the results revealed that the surface area (A, acres) of a water body was given as $A = 0.60958 + 1.09344 \Sigma_i E_i$ where E_i is the fraction of the *ith* resolution element composed of water. The standard error of the estimate was found to be 6.7 acres. The correlation coefficient of the estimate was determined as 0.998 for the sample tested. The slope of the regression curve was 1.093 which is consistent with the value of ground coverage in acres per pixel obtained by other methods. A further analysis revealed that the absolute error (5.2 acres) was independent of the size of the water body. On the basis of this limited analysis, this regression estimate introduces no systematic errors and can be used practically for estimating the areal extent of water bodies. The estimate can be further tested, but can be expected to give the best results in terms of percentage accuracy for larger sized water bodies.

4. THE MENSURATION SUBSYSTEM

This subsystem summarizes the classified multispectral data in terms of the mensuration parameters appropriate to a specific user application. Regression formulae of the type discussed above can be used to estimate the total geographical area directly observed to be devoted to a particular land use or a specific land feature. If the total data analysis system is based on a statistically rigorous model, extrapolation algorithms can be applied in this subsystem to infer information with known accuracy and cost function levels about the total area of interest based on the limited sample data actually processed (e.g., total estimate of crop yield based on the observed acreage within the measured geographical area and the observed crop vigor and maturity). The wide coverage of ERTS data can hopefully reduce the need for such extrapolations as the capability to efficiently process ERTS data tapes further develops. In this subsystem, additional information is normally required to relate the classified remotely sensed observations to the user desired management parameters. This may be accomplished through the use of phenomenological or mathematical models to which the classified data provides input or through the correlation of the observed data with other conventionally obtained statistical data (viz., census data, historical records on dollar yield per bushel, etc.). Techniques for this subsystem have yet to be demonstrated using ERTS data, but possible approaches are being explored at JSC in the context of the joint USDA/NASA data utilization program. Since this subsystem must be closely tailored to operational requirements, user agencies must play a significant role in selecting the techniques most appropriate in this subsystem for their application.

5. THE INFORMATION MANAGEMENT SUBSYSTEM

This subsystem stores all summary data in retrieval files, updates these files at the user's option and generates outputs in formats appropriate to specific user requirements. Such a system, with high-speed response and wide output flexibility has been developed and demonstrated at JSC. This subsystem referred to as the Regional Information Management System (RIMS)[6] is based on a one square kilometer grid and contains inventory information on land-use in the HATS area. A typical output form this system using the interpretation of aircraft data as the source is shown in figure 10. Information cells representing 100 percent forestry, water, urban and agriculture have been retrieved from the data base and displayed in this geographical format. The techniques for this subsystem must also be closely tailored to the user's specific operational requirements. User definition and documentation of such requirements is essential before meaningful techniques development for this subsystem can be developed. Indications are that ERTS data can provide useful data for the generation and update of such display information.

In summary, a systematic approach toward techniques development has been adopted at JSC. The system breakdown presented herein provides a framework against which current and future techniques development effort can be related and reported. The significant techniques refinements accomplished to date at JSC have been summarized in this context with the specific areas requiring additional development and extensive user agency activity highlighted.

REFERENCES

1. S. B. Chism, C. L. Hughes. "Screening and Data Handling Plan for ERTS-1 Computer Compatible Tape and Aircraft 24-Band Scanner Data." Lockheed Electronics Company, Inc., 640-TR-132, January 1973.
2. "ERTS Data Users' Handbook." Goddard Space Flight Center, Document 71SD4249.
3. Ball, G. H. and Hall, D. J. ISODATA, A Novel Technique for Data Analysis and Pattern Classification, Technical Report, Standard Research Institute, Menlo Park, California, May 1965.

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

5. Minter, R. T. "ISOCLS, Iterative Self-Organizing Clustering Program." Lockheed Electronics Company, Inc., Houston, Texas, CPD 202, October 1972.

6. Allison, J., Argo, W. and Shelton, M. "Users' Guide to a Regional Information Management System (RIMS)." To be published.

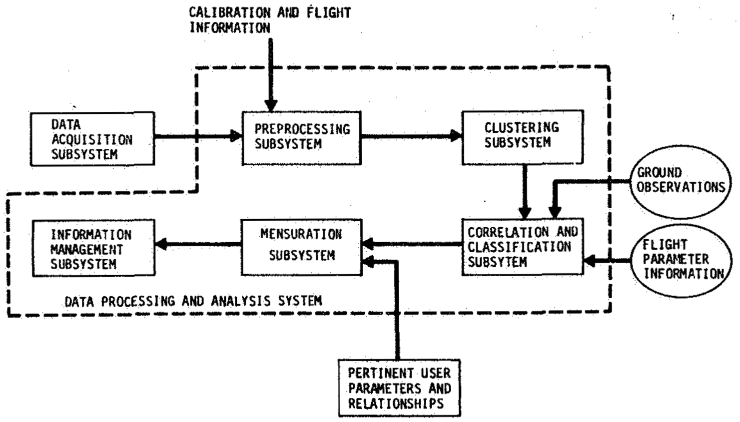

Figure 1 - Elements of a multispectral data processing and analysis system.

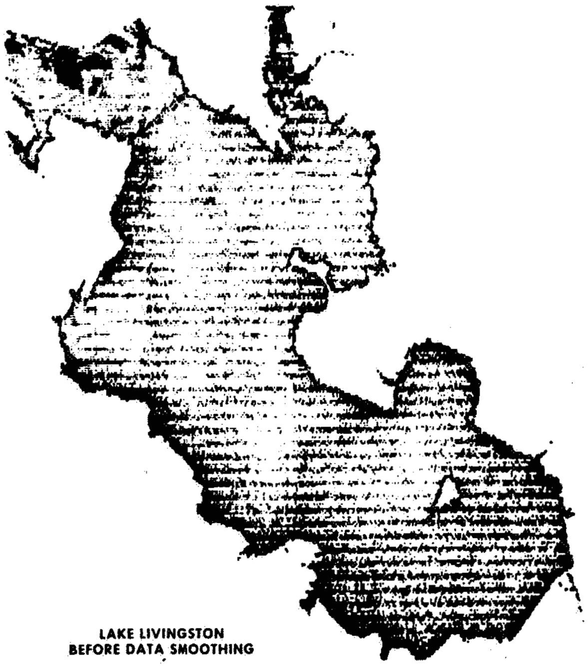

Figure 2

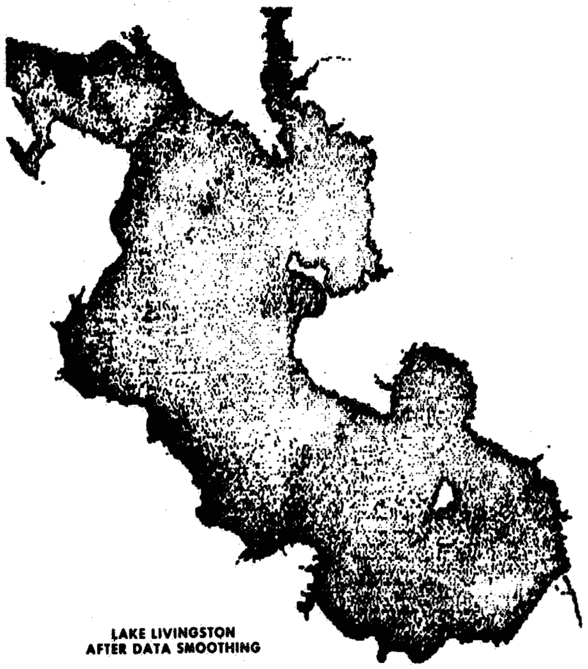

Figure 3

F

Figure 4a - ERTS MSS frame of August 30, 1973, of Lake Somerville. Thick haze over the lake is evident.

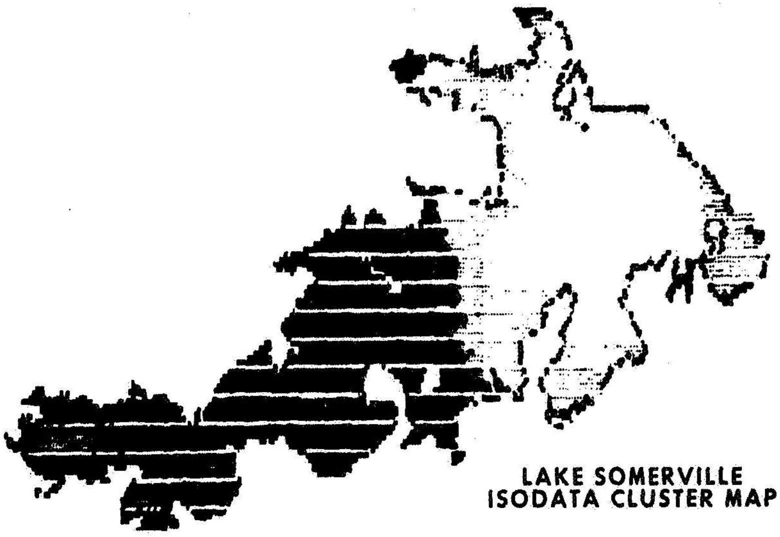

Figure 4b - ISODATA cluster map of Lake Somerville from the MSS data of August 30, 1973. Haze layer resulted in the lake being classified as several targets. On clear days the same program classifies the lake as a single target.

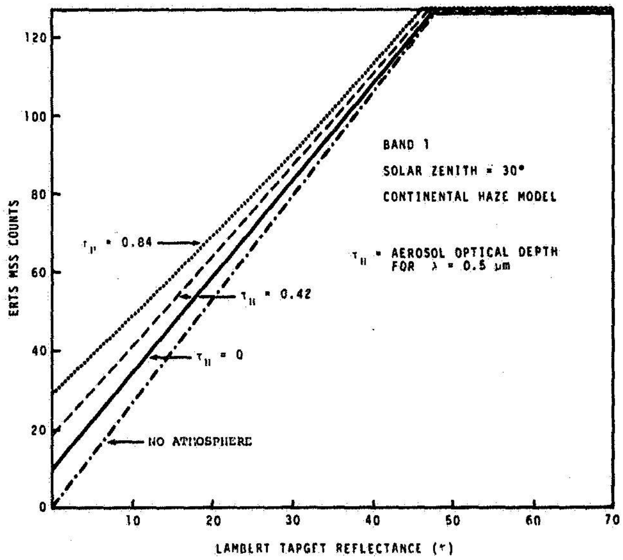

Figure 5 - Response curves for the ERTS MSS.

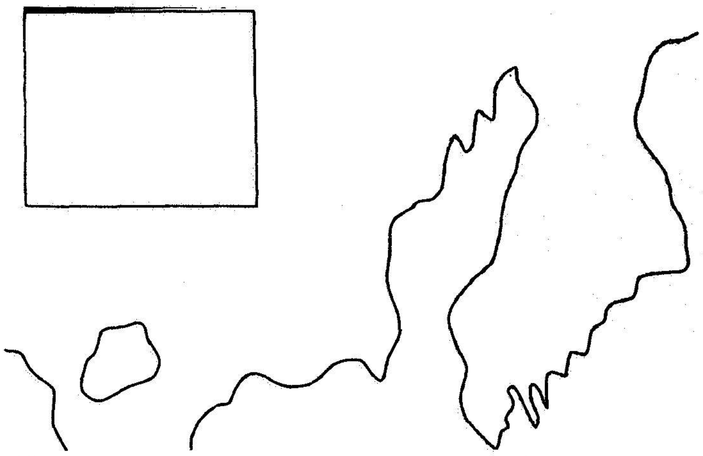

Figure 6 - Corrected gray scale map for a portion of the Somerville frame.

REFLECTANCE OF WATER, LAKE LIVINSTON, AUGUST 29, 1972

NADIR LOOK ANGLE

BAND	GROUND MEASURED VALUE (%)	MSS-ROTAR VALUE (%) $r_H = 0.5$
1	2.7	1
2	1.7	0
3	0.8	0

REFLECTANCE OF WATER, LAKE SOMERVILLE, AUGUST 30, 1972

NADIR LOOK ANGLE

BAND	GROUND MEASURED VALUE (%)	MSS-ROTAR VALUE (%) $r_H = 0.7$
1	3.5	2
2	2.5	1
3	1	0

Figure 7 - Comparison of reflectances obtained from ERTS MSS data corrected with ROTAR to ground measuring values.

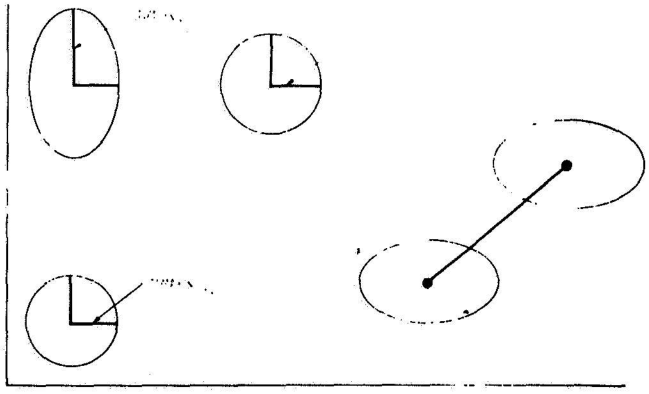

Figure 8 - Typical cluster parameters and illustration of their influence.

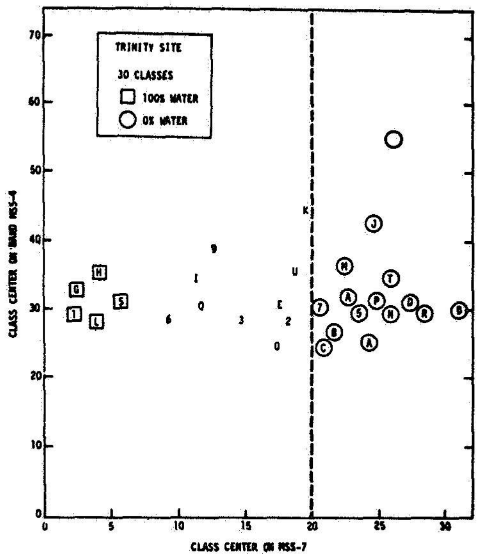

Figure 9 - Location of class centers for Trinity Delta.

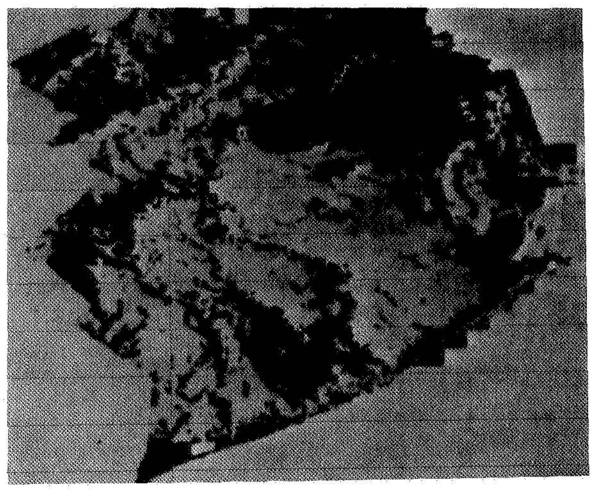

Figure 10 - Typical RIMS output of the HATS Area.