DEFINITION OF OPTIMUM ISOCLS PARAMETERS FOR RSIS INTERACTIVE DATA ANALYSIS, TEXAS COASTAL APPLICATIONS TEST SITE

Robert W. Baumgardner, Jr., and Robert J. Finley

Bureau of Economic Geology
The University of Texas at Austin
Austin, Texas 78712

Assisted by

Marcie E. Machenberg and Adel Moustafa

Prepared for the

Texas Natural Resources Information System

TDWR/TNRIS IAC No. (80-81)-1676

November 1980

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

IMAGE ANALYSIS

Data tapes for a specified window are processed through a program called ISOCLS (Iterative Self-Organizing Clustering). The user supplies several input parameters including the spectral distance between clusters (DLMIN) and the maximum standard deviation for the values within a cluster (STDMAX). The purpose of this study was to vary these two parameters and analyze the differences in images produced from the same data.

It has been standard procedure to interrupt the data processing scheme (fig. 1) after the ISOCLS program has been run. At that time clusters which have a small number of pixels and/or high standard deviations are deleted. The remaining pixels are examined as follows. The mean reflectance value in channel 2 (band 5) is plotted against twice the mean reflectance value in channel 4 (band 7) for each of the clusters. The data points tend to fall along lines radiating from a point near the vertical (channel 2) axis of the graph. The lines are generally defined as water, vegetation and substrate lines (fig. 2) (Weisblatt, personal communication, 1980). Those clusters, generally between the vegetation and substrate lines and not aligned with these three groups, are "hybrid" clusters, which appear to consist of mixtures of bare substrates and vegetation.

Channel 4 values generally reflect moisture content of the target. Low channel 4 values indicate high moisture content. Channel 2 values indicate whether the target appears light (high channel 2 value) or dark (low channel 2 value) in the red portion of the visible spectrum.

Water clusters have very low channel 4 values, and are distributed subparallel to the channel 2 axis depending on turbidity. Higher channel 2 values indicate more turbid water. Clusters defined as vegetation

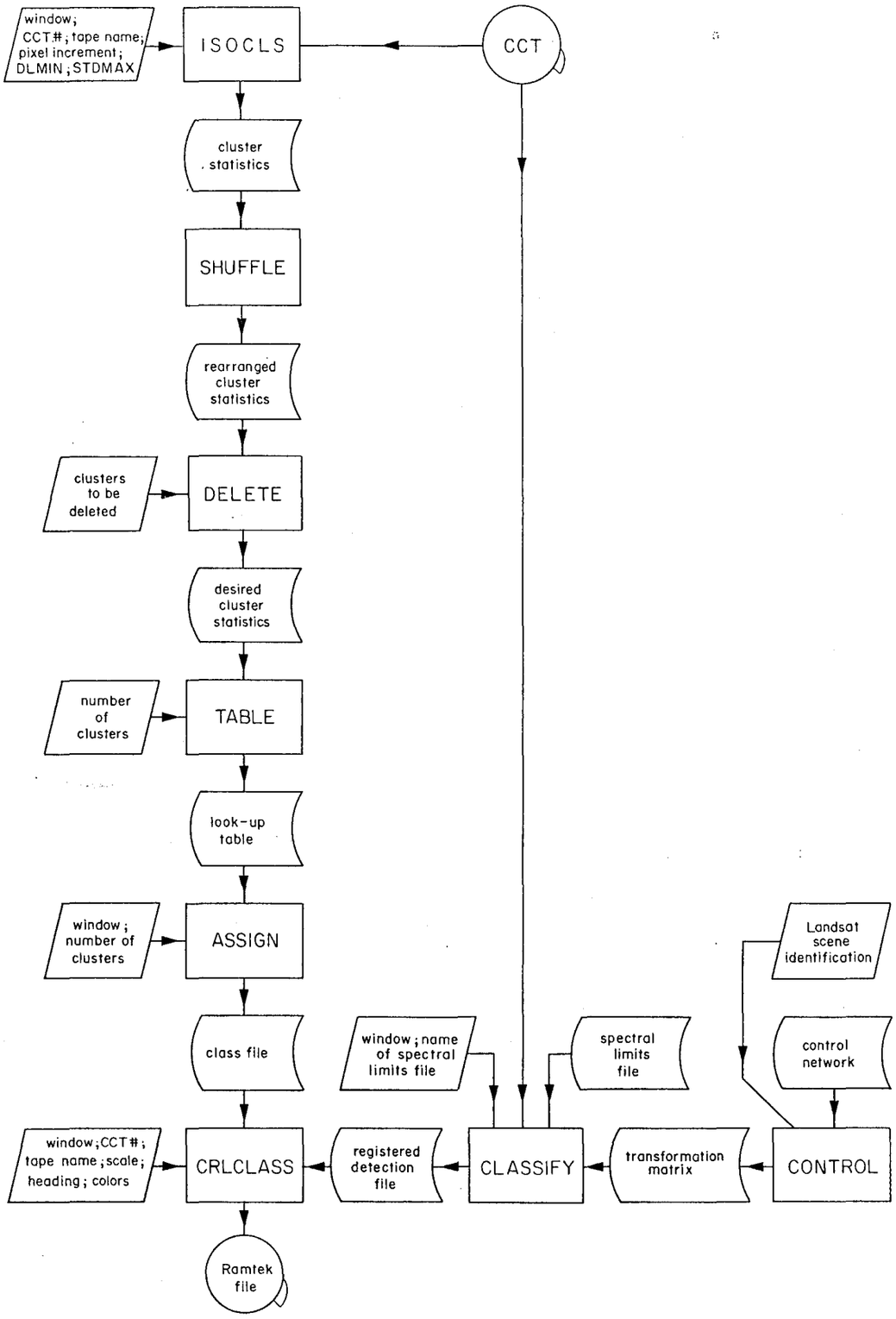


Figure 1. Flow diagram of Landsat data through classification programs to output on the Ramtek Graphics Display Terminal. Processing is usually interrupted after ISOCLS to examine cluster statistics and assign colors to clusters.

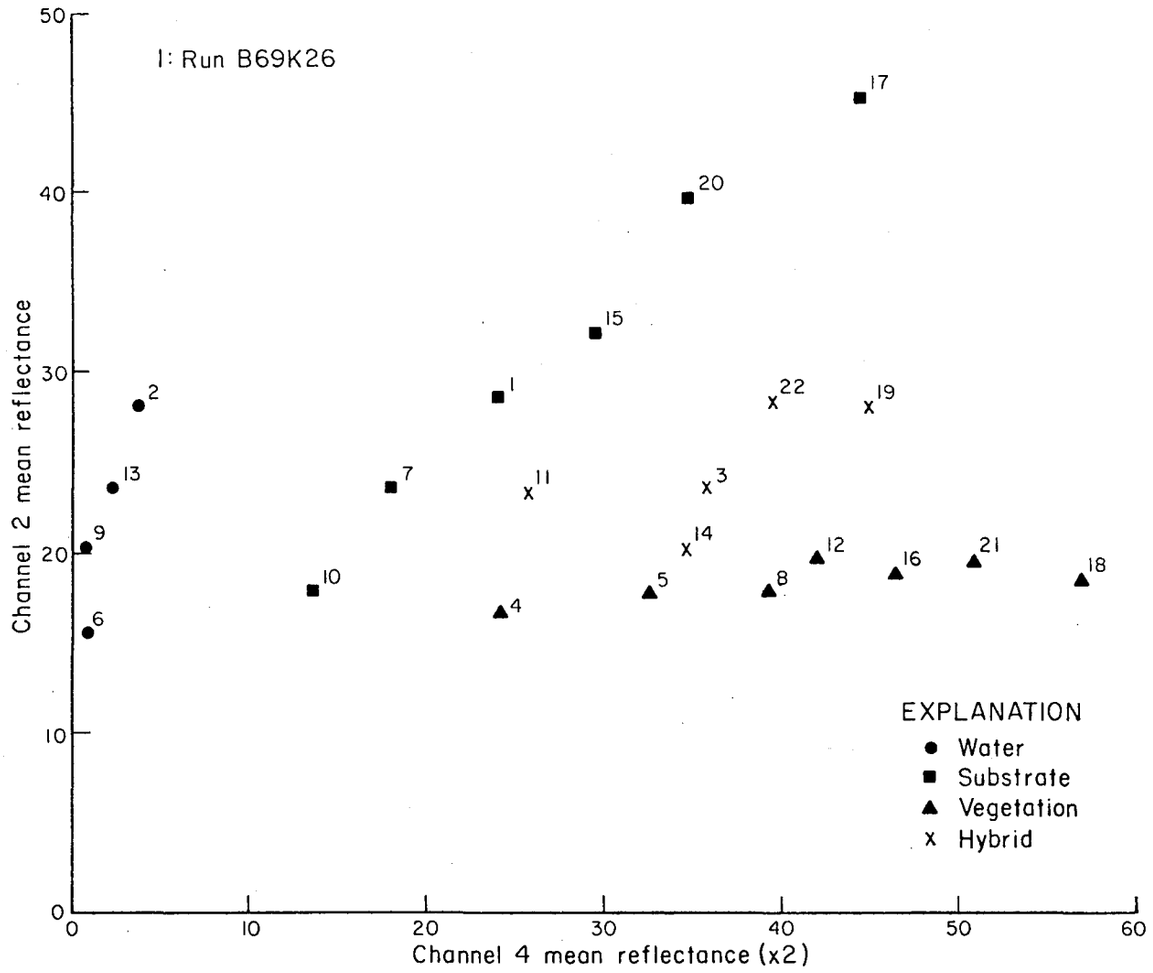


Figure 2. Plot of mean reflectance values in channel 2 and channel 4 for run B69K26. Channel 4 values are doubled. Four major types of clusters are seen here: water, substrate, vegetation, and hybrid. All but the hybrid clusters define lines radiating from a point near the y-axis. This and figures 4 through 6 are discussed in the text.

describe a line subparallel to the channel 4 axis (fig. 2). Clusters in this group have a wide range of values in channel 4, reflecting changes in moisture content, but the range of values in channel 2 is narrow, indicating that all these clusters have little difference in reflectance in the 0.6 to 0.7 μm wavelength range.

Clusters defined as "substrate" describe a line about halfway between the water and vegetation lines. Clusters with low values in both channels may be a combination of organic-rich soil, water, and vegetation. Clusters with high values in both channels are probably dry, sparsely vegetated surfaces such as industrial complexes or sandy, barren areas.

After a channel 2 versus channel 4 plot is complete the user knows approximately what each cluster represents. Using this information he/she assigns to each cluster a color selected from the color menu, a list of 62 available colors (table 1). By assigning similar color schemes to several different ISOCLS iterations, a number of different displays of the same Landsat scene can be studied sequentially at the Ramtek display terminal.

This method allows the user to efficiently determine the effect that changing STDMAX and DLMIN values have on the representation of the same reflectance data. He/she can then select the values for these two parameters that are best for his/her needs.

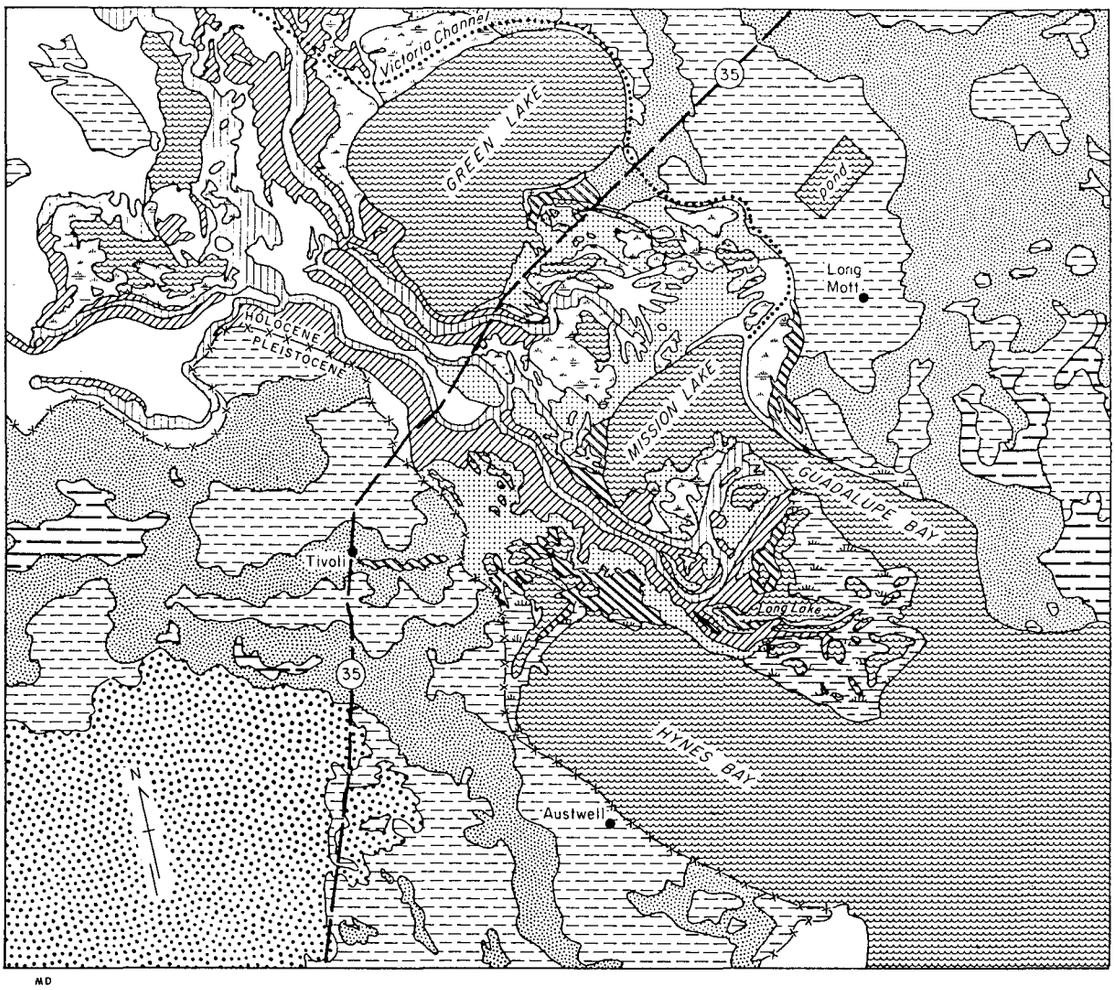
For Phase I of this study the parameter combinations used are shown in table 2. Landsat Scene No. E-21348-16023, for October 1, 1978, was used for all four ISOCLS runs. A window centered on the Guadalupe River delta was selected to include areas of rangeland, cropland, wetlands, and water (fig. 3). These runs were examined in detail and compared

Table 1. Standard color menu used for assigning colors to clusters generated by ISOCLS program.

Color No.	Gun Setting	Color Name	Color No.	Gun Setting	Color Name	Color No.	Gun Setting	Color Name
1	000015	True blue	22	000905	Dull green	43	150000	Red
2	000013		23	000704		44	110000	
3	000011		24	000503	Gray-green	45	090000	
4	000009		25	001500	Bright green	46	050000	Burgundy
5	000007		26	001100		47	030000	Red-brown
6	000005	Midnight blue	27	000700		48	150008	Hot pink
7	000003	Blue-black	28	000300	Forest green	49	110006	Light scarlet
8	000815	Light blue	29	081500	Lime green	50	090005	
9	000713	Light blue-gray	30	061100		51	050003	Magenta
10	000611		31	040700		52	150015	Mauve
11	000407		32	030500	Olive green	53	090009	
12	000305	Dark blue-gray	33	151500	Yellow	54	070007	
13	001515	Aqua	34	131300		55	050005	Lavender-mauve
14	001313		35	090900		56	080015	Lavender
15	001111		36	050500	Yellow-brown	57	060011	Light purple
16	000909		37	030300	Brown	58	040007	
17	000707		38	150800	Orange	59	030005	Dark purple
18	000505	Gray-aqua	39	130700		60	141414	Light gray
19	001508	Pale green	40	110600	Ochre	61	080808	
20	001307		41	050300	Chocolate	62	050505	Dark gray
21	001106		42	030200	Miss. brown			

Table 2. Parameter combinations for ISOCLS program runs using a single Landsat data set covering the Guadalupe River delta and vicinity.

Run No.	Parameter			Total No. of Clusters	No. of Pixels in Cluster		Type of Clusters			
	NMIN	DLMIN	STDMAX		Largest	Smallest	Water	Sub-strate	Vegeta-tion	Hybrid
1:B69K26	50	2.0	3.0	24	3,335	40	4	6	7	5
2:B69K36	50	3.5	3.0	20	5,238	61	3	5	5	3
3:B69K37	50	4.0	3.0	15	6,464	55	3	5	5	0
4:B69K45	100	5.0	3.8	8	7,948	389	2	3	3	0



EXPLANATION

- Unclassified
- Water
- Marsh, salt water, mud and locally sand substrate
- Marsh, fresh water, mud and locally sand substrate
- Delta plain mud and sand, sparsely mud-covered
- Delta plain mud and sand, grass-covered
- Interdistributary mud and sand veneer
- Interdistributary mud
- Distributary and fluvial sands and silts
- Marine deltaic sand, delta front, and reworked delta facies
- Levee, silt, mud, and sand, sparsely grass-covered
- Levee, silt, mud, and sand, tree covered
- Pleistocene / Holocene boundary

Figure 3. Environmental geology of the "window" in Landsat scene E-21348-16023 for October 1, 1978, examined in this study. Modified after McGowen and others, 1976.

to determine the effects of different parameter combinations on the number, size, and information content of clusters.

1:Run B69K26

For Run B69K26 DLMIN was set at 2.0, and STDMAX was 3.0 (table 2). The channel 2 versus channel 4 plot for this run is shown in figure 2. Of the original 24 clusters generated by the ISOCLS program, two were deleted owing to their small number of pixels and high standard deviations. These pixels were grouped as unclassified data and assigned a color from the color menu. The remaining 22 clusters were plotted as described above. Of these 22, four were classified as water (2, 6, 9, 13), six were classified as substrate (1, 7, 10, 15, 17, 20), seven were classified as vegetation (4, 5, 8, 12, 16, 18, 21), and five were "hybrid" clusters (3, 11, 14, 19, 22) (fig. 2).

Twenty-two clusters were more than enough for delineation of broad categories such as rangeland and cropland. Not all clusters could be interpreted as unique land cover/land use types on Landsat imagery. Apparently, many clusters represent only slight changes in moisture content or leaf canopy cover not associated with any distinctive type of vegetation or land use already mapped or distinguishable on available aerial photographs. Field checking has not been carried out.

The four water clusters delineated various degrees of turbidity from clear (6) to moderately turbid (2). The clearest water was found in the rectangular ponds near the chemical plant north of Long Mott (fig. 3). The output of this scene was photographed on the Ramtek screen and the slide was magnified and compared with the Port Lavaca Sheet of the

Environmental Geologic Atlas of the Texas Coastal Zone (McGowen and others, 1976).

The substrate clusters were found mostly in areas of barren cropland, although the two with the highest reflectance (17 and 20) were found at the chemical plant site. Only cluster 10 was identified with a specific land cover type, that being grass-covered levee.

The vegetation clusters were associated with rangeland, for the most part, and with vegetated Holocene fluvial deposits including the Guadalupe River delta. Clusters 4 and 16 were associated with the tree-covered levees in the delta and west of Mission Lake. Cluster 5 appeared to correspond with grass-covered levees and fresh-water marsh. The remaining four vegetation clusters (8, 12, 18, 21) could not be clearly associated with a recognizable land cover type because they occurred over widely scattered, small areas which gave a mottled appearance to the image display.

The hybrid clusters (3, 11, 14, 19, 22) comprised boundaries and were scattered among other clusters. Cluster 19 clearly defined the town of Austwell, although other pixels with this color were scattered about the image.

In general, run 1 was useful for separating different types of vegetation such as tree- and grass-covered levees. But it was not as good for defining land/water boundaries of small water bodies. In some cases, small lakes were not assigned blue pixels, but were assigned the same color as the darkest bare substrate, cluster 10 (fig. 2). The poor definition of land/water boundaries was probably a result of boundary pixels being assigned some other color than the two available for water and the dominant surrounding land cover type. The incorrect color

assignment was caused by the reflectance characteristics of the lake being more like those of the bare substrate cluster than the water cluster for the clusters as defined in this run. This run delineates roads, such as Highway 35 west of Austwell, better than run 4, which had only eight clusters, probably for the same reason that land/water boundaries were obscured by pixels of different colors.

2:Run B69K36

For this run DLMIN was 3.5 and STDMAX was 3.0 (table 2). Sixteen of the original 20 clusters were retained, three of which belong to the water group (fig. 4). Five clusters were grouped as vegetation, five were substrate, and three fell into the hybrid category.

This run has six fewer clusters than run 1. The vegetation and hybrid categories lost two clusters each, while water and substrate decreased by one each.

As in run 1, water clusters defined areas of varying turbidity ranging from clear (6) to moderately turbid (2). Clear water was again found in the ponds near the chemical plant north of Long Mott.

All substrate clusters (1, 7, 10, 11, 14) are associated with cropland. On a Landsat false-color composite they all appear as various shades of blue-gray to bluish-white, denoting various degrees of moisture content in substrate.

Like run 1, this scene was photographed on the Ramtek screen and compared with the Environmental Geologic Atlas of the Port Lavaca area. All vegetation clusters (4, 5, 8, 12, 15) are associated with rangeland. Cluster 4 corresponds to tree-covered levees in the Holocene fluvial areas and the Guadalupe River delta (fig. 3). The remaining four clusters are

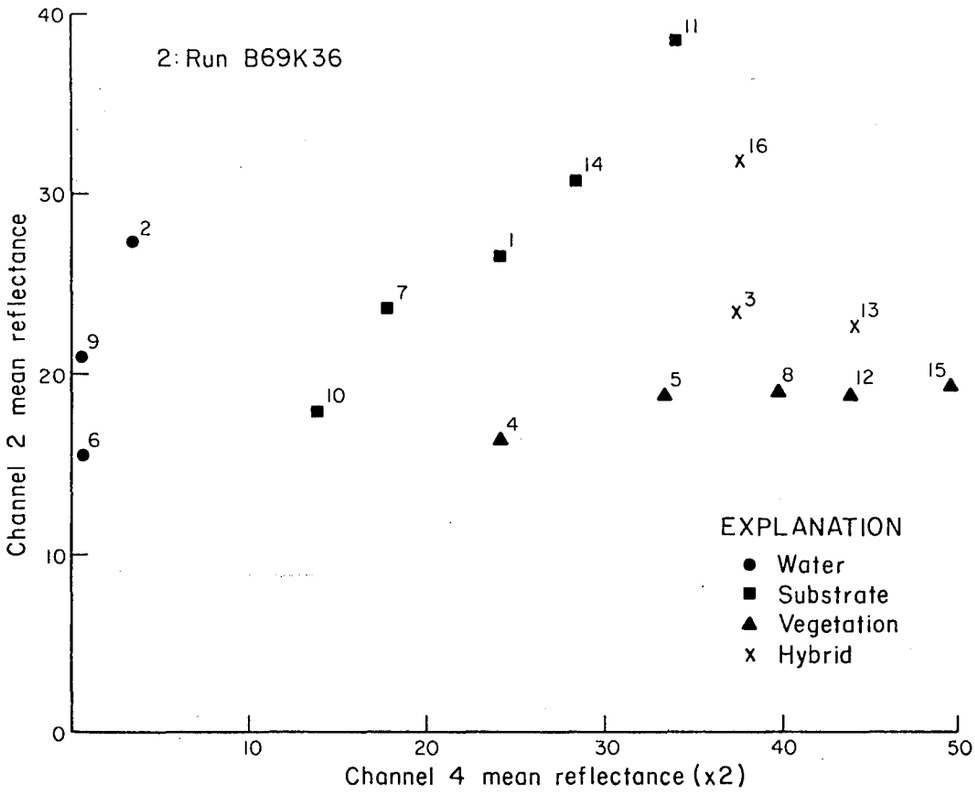


Figure 4. Plot of mean reflectance values in channel 2 and channel 4 for run B69K36. Channel 4 values are doubled. Four major types of clusters are seen here as in figure 2.

not closely associated with any particular land cover, but are scattered among each other and the other clusters in small groups.

Hybrid clusters define boundaries between other clusters (3, 13), delineate roads (16), and are scattered throughout the scene. Cluster 13 delimited the town of Austwell, showing it off to advantage among pixels of very different colors.

The accuracy of land/water boundaries of small lakes such as Long Lake is better than that of run 1. But, some smaller lakes are assigned the same color as bare substrate. In addition, the boundary between the Pleistocene uplands and the Holocene fluvial system north of Tivoli is less obvious. The delineation of different vegetation on levees is retained mostly unchanged from run 1, and Highway 35 still shows up west of Austwell.

3:Run B69K37

This run had a STDMAX of 3.0 and a value for DLMIN of 4.0 (table 2). Fifteen clusters resulted, of which 13 were retained, not including unclassified data. Run 3 clusters are aligned closer to the three major groups than on any of the other runs (fig. 5); three water clusters (3, 6, 8), five vegetation clusters (2, 4, 5, 10, 12), and five substrate clusters (1, 7, 9, 11, 13) were produced by this run. The basic difference between this and run 2 is the absence of hybrid clusters. The number of clusters in each of the three major groups is unchanged.

As in previous runs, the substrate clusters are most closely associated with cropland and wetlands, while vegetation clusters are most closely associated with rangeland. The absence of hybrid clusters means

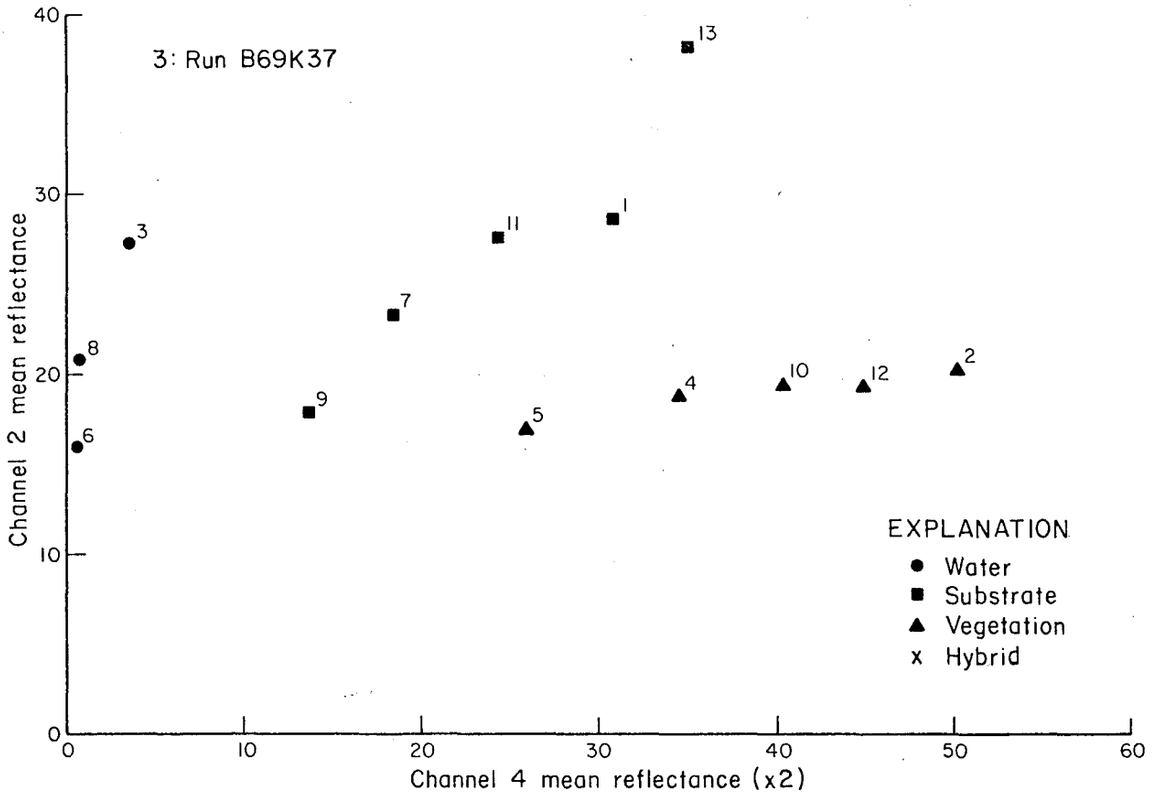


Figure 5. Plot of mean reflectance values in channel 2 and channel 4 for run B69K37. Channel 4 values are doubled. No hybrid clusters appear in this run.

that the delineation of some differences in reflectance is diminished, but there is no overall decrease in information content of the scene.

Long Lake and other small, unnamed lakes in the Guadalupe River delta and north of Hynes Bay are somewhat better defined than in run 1. Additionally, some blue pixels appear in the centers of these small lakes, rather than being totally classified as wet, bare substrate.

Austwell is clearly, though not uniquely, defined, and Highway 35 is partly delineated west of the town. The industrial areas are identified by the largest groups of pixels of the color that characteristically defines them. This class also appears as bad scan lines. Marsh and grassy levee are lumped together as cluster 9, but are separated from the tree-covered levee land cover (cluster 7).

4:Run B69K45

This run had only eight clusters, the smallest number of any run in this study. This was attempted deliberately to define a minimum number of clusters that would be useful for image analysis. DLMIN and STDMAX were assigned values of 5.0 and 3.8, respectively (table 2). For the first time NMIN was increased from 50 to 100 to reduce the number of clusters generated by setting a higher limit on the least number of pixels that could be included in one cluster. But, the smallest cluster had 389 pixels, so it is not known how effective NMIN was in this run.

Two water clusters (3, 7), three substrate clusters (1, 4, 8), and three vegetation clusters (2, 5, 6) were defined (fig. 6). No hybrid clusters were defined, just as in run 3 (table 2). Small lakes, including Long Lake in the Guadalupe River delta, were better defined on this

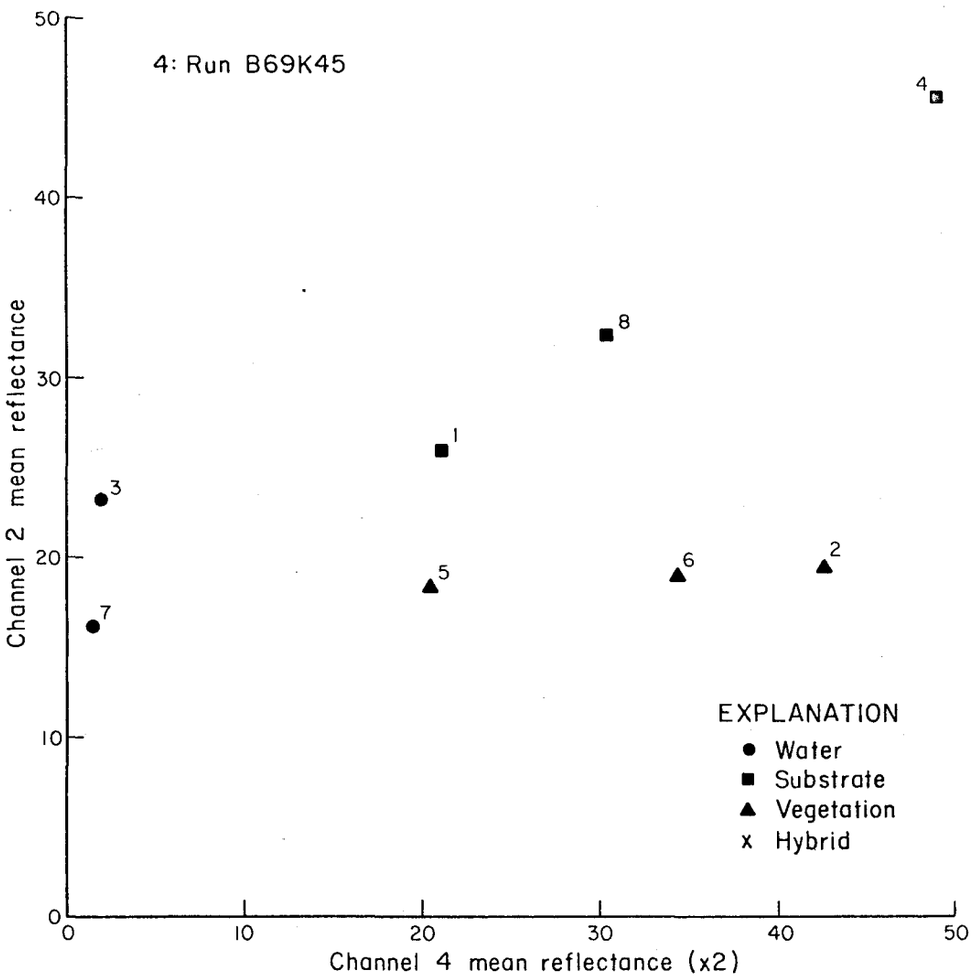


Figure 6. Plot of mean reflectance values in channel 2 and channel 4 for run B69K45. Channel 4 values are doubled. No hybrid clusters appear in this run.

run than on run 2. They were assigned a blue color in this run rather than the color corresponding to wet, bare substrate.

Unlike previous classifications, the correspondence between substrate clusters and cropland and between vegetations clusters and rangeland is not clear. Cluster 5 is mostly associated with cropland and marsh, and lies in a position on figure 6 that is halfway between the vegetation and substrate lines as seen on the other graphs (figs. 2, 4, 5). Clusters 1 and 8 are associated with cropland, while clusters 2 and 6 are associated with rangeland (fig. 6). Cluster 4 corresponds to white areas on the Landsat false-color composite, including the chemical plant near Long Mott, and is scattered among the areas occupied by cluster 2. It also clearly defines Austwell.

The smaller number of clusters necessarily obliterates some of the detail of other classification displays having more clusters. For example, the distinction between grass-covered and tree-covered levees northwest of Green Lake is lost. In addition, Highway 35 west of Austwell no longer is defined by a line of pixels of different classes than the surrounding area.

SUMMARY

The smaller number of clusters in run 4 tends to blur distinctions between vegetation types such as tree- and grass-covered levees. It follows that the distinction between very different land cover/land use categories such as bare, muddy substrate and water becomes more clearly defined with a smaller number of possible clusters. Regardless of the number of clusters, areas of anomalous reflectance, such as the chemical

plant near Long Mott and the town of Austwell, stand out in sharp contrast with their surroundings. However, they may not be uniquely identified by a color in the classification displays with fewer clusters, although they are in displays with more clusters. For optimal delineation of land cover/land use types that can be associated with recognized environmental geologic units, run 3, B69K37, may have the best combination of input parameters seen in this study. The result is a number of clusters large enough to distinguish between similar vegetation types and small enough to define shallow lakes as turbid water, not as bare, wet substrate.

REFERENCES

McGowen, J. H., Proctor, C. V., Jr., Brown, L. F., Jr., Evans, T. J., Fisher, W. L., and Groat, C. G., 1976, Environmental geologic atlas of the Texas Coastal Zone - Port Lavaca area: The University of Texas at Austin, Bureau of Economic Geology, 107 p.