D34

N78-14564

SATELLITE LAND USE ACQUISITION AND APPLICATIONS TO HYDROLOGIC PLANNING MODELS

V. R. Algazi and Minsoo Suk

Department of Electrical Engineering
University of California
Davis, California 95616

ABSTRACT

In this paper, we report on a developing operational procedure for use by the Corps of Engineers in the acquisition of land use information for hydrologic planning purposes. The operational conditions preclude the use of dedicated, interactive image processing facilities. Given the constraints, an approach to land use classification based on clustering seems promising and is being explored in detail. The procedure is outlined and examples of application to two watersheds given.

1. HYDROLOGIC ENGINEERING PLANNING MODELS

The objective of our work is to develop operational procedures for use by the Corps of Engineers across the United States in the acquisition of land use information. The primary source of data is LANDSAT digital data and the intended use of the land use information is for hydrologic planning purposes. Land use information allows the Corps of Engineers to assess the flood hazard, general damage potential, and environmental status of watersheds. Use of this information is illustrated in a recently completed pilot research Flood Plain Information (FPI) study [1] by the Hydrologic Engineering Center (HEC) of the Corps of Engineers. In the study, data management and analytical techniques, integrating the use of spatial gridded geographic data files (called Grid Cell Data Bank), are incorporated into hydrologic computer models denoted HEC-1 and STORM.

The analysis methods of the HEC study interpret the hydrologic, economic, and environmental consequences of alternative land use patterns in combination with other physical characteristics of the watershed, such as soil class, land slope, erosion index and topography. Land use information is the key factor in performing the analysis in that it is used as the primary indication of the watershed conditions and of its response to precipitation.

The acquisition of land use information by conventional methods such as manual classification using aerial photographs or ground surveys are often time consuming for large watersheds or inadequate, not providing accurate spatial information of land use. Remote sensing data can provide land use information accurately and in a timely fashion for hydrologic planning purposes. By proper use of high speed digital computers, highly accurate and point-by-point information of land use can be extracted from the remotely sensed data.

2. LAND USE CATEGORIES AND REMOTE SENSING FOR HYDROLOGIC APPLICATIONS

Since the land use pattern is an important factor in hydrologic, economic and environmental analysis, the development and use of a reasonable set of land use categories is quite important. Hardy and Anderson [2] have recommended a standard set of land use categories for use with remote sensing data. Ragan [3], in applications to water resources, has used a modified subset of land use categoires* of Hardy and Anderson, and has shown that remote sensing

*Land use categories used by Ragan in his work are: Forested area, highly impervious, grassed area , residential, streets and highways, bare land, streams, ponds or pools.

data can provide land use information. The land use pattern was then used by Ragan to determine hydrologic parameters in urban hydrology.

On the other hand, we note that the FPI study by HEC has applications to economic and environmental analysis as well as to hydrology. Thus, the objectives and criteria which determine a set of land use categories in this study are different from what was used in previous work. Quoting the criteria applied by HEC to determine a rational set of land use categories:

" • The categories should be reasonably compatible with local and other agency land use classification schemes

• It must be reasonably possible to classify the land use within the study area by conventional or automated means

• The land use categories should allow rational, consistent determination of flood hazard, economic and environmental effects of land use change

• The land use categories should be compatible with those needed by certain available computer models

• The land use categories should provide a complete umbrella of classifications so that further breakdown of land use within each category would be possible if deemed necessary in future studies"

The different concerns in land use for each application are well expressed in another quotation from the HEC report:

> " ... from the hydrologic viewpoint, the concern in a land use sense is with moisture retention/precipitation excess and basin response characteristics which are related to impervious cover and land surface management measures. From the economic viewpoint the damage potential and disruption of community activities is a function of urban development in general and the size, density, and type of structures and contents. From the environmental viewpoint, the concern is mostly with the intensity of development and the potential for adverse impacts (such as pollution) that could derive therefrom."

In the specific application of the FPI study to the Trail Creek Watershed in Georgia, HEC has adopted the set of land use categories shown in Table 1. These categories represent a compromise between the general criteria mentioned above and the technical requirements needed for applications to hydrology and economic and environmental studies.

Note that, for the economic and environmental analysis, detailed land use information in urban areas is quite important, which is not the case for hydrologic analysis. The requirements of accurate urban land use classification, such as differentiation between commercial and industrial areas, and differentiation of housing density of residential areas, are quite difficult to meet from remote sensing data.

3. OUTLINE OF AN OPERATIONAL PROCEDURE

Since the final operational procedure should be easily applicable by the field engineers of the Corps of Engineers, this precludes the use of any dedicated and highly interactive image processing hardware (such as G.E. Image 100 or the Bendix M-DAS System). The procedure should require only a limited number of iterations and not rely on the use of full scale color image display. The emphasis is thus in the use of general purpose computers and line printers for intermediate and final output products.

With these specific objectives and constraints, we have first tried a maximum likelihood classification algorithm to determine whether it can be adapted into an operational procedure. We have encountered several difficulties in using a maximum likelihood classifier. Among these are: (1) The choice of land use categories on which the classifier will be trained: The set of land use categories should be complete in the sense that every part of the watershed should belong to one of these categories. We have found the selection of land use categories difficult. (2) Training areas: To have a reliable estimate of statistics, a large number of sample points (corresponding to large size training areas) is required. In actual applications, it is not easy to find training areas of large size for certain land use categories (principally in urban areas). Further, the determination of the exact outlines and coordinates on the LANDSAT image for each training field is also difficult in the absence of an interactive color image display.

These difficulties make a maximum likelihood classifier unattractive or impossible as an operational procedure with a minimum amount of interaction. Unsupervised classification algorithms appear to be more suitable to our objectives. The clustering approach is a well known unsupervised classification algorithm, which does not require a priori knowledge of land use categories nor locating training areas. At this stage of our work, we are exploring in detail an operational procedure for satellite land use classification based on the clustering approach. The steps and sources of data for the proposed operational procedure are as follows:

a. Digital Specification of the Watershed. Information on the watershed obtainable from maps, such as the watershed boundary and major roads, is entered on a grid oriented data base using an x-y digitization tablet or by key punching the data on cards.

b. Preprocessing of the Data. The original LANDSAT data is transformed using a principal component or Karhunen Loeve (KL) transformation. Only two of the transform components are required for classification and this step results in a reduction in computing costs. This step is optional and can be bypassed for a small watershed.

c. Clustering. We have implemented and tried a clustering program based on the ISOCLS package developed for NASA Johnson Space Center.

d. Classification. The data is classified after clustering by labeling each cluster as belonging to one of the land use categories. This requires ground truth information in the form of maps and aerial photographs. After examining all the available information such as the display of the centers of resulting clusters, maps, and aerial photographs, one of the following decisions is made for each cluster: (1) The cluster belongs to a specific land use category. (2) It is a mixture of two or more land use categories or the information at hand is not sufficient to label the cluster, i.e., the cluster is either in conflict or inconclusive in nature. (3) The cluster is of no importance or not valid. We assign that cluster to an "other" category (none of the desired land use classes). The ground truth information such as maps and aerial photographs is used to label the clusters in the following manner. From the examination of the computer printout of clustering results, several spatially contiguous areas (each having more than M points) within each cluster are chosen and the corresponding LANDSAT data is brought in registration with maps and aerial photographs. By studying corresponding areas on all available data we make one of three decisions for each cluster as outlined above. We can also use the reverse process, i.e., define some ground truth points or areas on maps and photographs, and transform those points or areas to LANDSAT image coordinates. We can then label the data clusters. The registration procedure of maps, aerial photographs and LANDSAT data will be discussed later.

e. Reclustering. For the points belonging to the second group of clusters, i.e., clusters in conflict or inconclusive, a reclustering step is applied. First, points belonging to this group are selected from the original LANDSAT data; then the clustering algorithm is applied again to those points. The purpose of this reclustering is to more finely subdivide the data in the difficult areas to allow unequivocal labeling of clusters. After reclustering, all the resulting clusters are labeled using the procedures described in Step d, with the only difference that we now try to label all clusters. Clusters which cannot be labeled properly are assigned to the "other" class. However, we expect that very few points will belong to this group. A schematic diagram showing the steps above is given in Figure 1.

f. Geometric Correction and registration of Maps, Aerial Photographs and LANDSAT Data: Geometric correction, using principally a least square geometric correction program requires that a number of control points be obtained from all the sources of data and entered in numerical form into a program. Obtaining such ground control points for LANDSAT data is an important porblem which remains to be solved for the case in which no high quality, high resolution display of LANDSAT data is available. Alphanumeric printouts of portions of raw LANDSAT data which accentuates landform information is being considered as a possible source of ground control points.

4. PILOT STUDY

The operational procedure described above was applied to two watersheds of different size and geographic location; the Trail Creek Watershed in Georgia and the Castro Valley Watershed in California. Here we explain the results to date.

4.1. TRAIL CREEK WATERSHED

The Watershed is located in Clarke County and in the city of Athens, Georgia. It is relatively small, approximately 30 square kilometers and has been further subdivided into 21 subbasins in an HEC study.

Because of the need to establish a basis of comparison on pixel by pixel for our work using satellite data, we proceeded to do a manual classification of the land use in the watershed using NASA Research Aircraft images (Scene ID 6274 000 50047, 74/4/24). The manual classification serves as a principal basis for the verification of remote sensing classification results. Results of the manual classification are displayed in Figure 2. A tabulation of the percentage of each land use class is shown in Table 2.

4.1.1. MAXIMUM LIKELIHOOD CLASSIFICATION

In order to test the suitability of well developed classification algorithms, we first examined a maximum likelihood classifier. A particular version of LARSYS program was chosen primarily because of its availability and because of the well established application of maximum likelihood classifiers in agricultural land use classification [4].

We have attempted to classify an October scene (Scene ID 8180415322, 74/10/5) of the LANDSAT image of the Trail Creek Watershed using a maximum likelihood classifier. The steps in the classification procedure are: (1) Define a reasonable set of land use categories, (2) locate one or several training fields for each class on LANDSAT imagery and identify the coordinates of each training field, (3) run the classification program, and (4) process the result for display and tabulation. The classification result is displayed in Figure 3 and summarized in Table 2, along with other results.

4.1.2. CLUSTERING APPROACH

As explained previously, we encountered difficulties in the use of the maximum likelihood classifier, such as the choice of land use categories and defining training areas.

Considering these difficulties and our objective to develop an operational procedure with a minimum amount of interaction, we shifted emphasis from supervised to unsupervised classifiers. The clustering approach to land use classification is based on classifying first the data into machine classes or clusters according to machine measure of homogeneity without injecting into the process the human preconception of what the land use categories should be. Then a human being interacts with the machine to interpret and refine the results of the machine classification. At this second stage, the prior knowledge of land use and the relative importance of achieving accurate classification results for each land use category play an important role.

We have used the clustering approach for the Trail Creek Watershed. The same October scene used in the maximum likelihood classification was used again. The procedure used for our pilot study consists of the following steps:

(1) Principal component transformation of the data. The original LANDSAT image is transformed for data compression using the Karhunen Loeve transformation.

(2) Clustering of the data. The first two components of transformed data KL1 and KL2 are clustered. A display of the centers of resulting clusters are given in Figure 4.

For the purpose of comparison, we tried to label all clusters with land use categories as best we could without using the reclustering step. The result of the classification is summarized in Table 2, and designated "one step clustering."

(3) Initial classification. As described in step d of the operational procedure, we divide the clusters into three groups, shown also in Figure 4.

(4) Reclustering. For the clusters marked "reclustering" in Figure 4, we applied the

clustering program again using a different set of parameters. The final result of steps (3) and (4) are shown in Figure 5 and Table 2.

The result obtained by this operational procedure can be compared to the ground truth and to the maximum likelihood classification result. Even though clustering is significantly better than maximum likelihood classification, improvements might not be as apparent in direct numerical comparison of percentage of land use in each class, since generally numerically compiled results are the average of many detailed effects. However, the following conclusions seem justified.

(a) The clustering approach results in a significant improvement over the maximum likelihood classification when we examine the detailed classification point-by-point on an image.

(b) The clustering approach is much more flexible in the sense that the classes are assigned after the fact.

(c) Reclustering result is a significant improvement over the one step clustering classification.

(d) It still appears to be necessary to devise some kind of consolidation program to remove extraneous and isolated misclassified points.

Note also that we have used finally, only 6 land use categories instead of the 10 categories used by HEC. The following comments are pertinent:

(a) The separation of industrial and commercial classes. These two categories may not be differentiated accurately from remote sensing data. By applying a spatial consolidation algorithm, a partial success appears possible. For example, in clustering, we have a good indication that downtown commercial areas and large size parking spaces around shopping centers may be identified. Further work is needed.

(b) Density of residential areas. That depends largely on the definition of low, medium and high density residential areas. Residential areas, generally, tend to be clustered as newly developed or old residential areas on the basis of surroundings rather than housing density. More work on fairly large urban areas is needed to determine whether density of residential areas can be determined by using remote sensing data.

(c) Separation among agricultural, pasture and developed open space. We have not paid too much attention to this problem as yet. Even with lots of care and attention, it seems difficult to separate these classes even from high-flight images.

4.2. CASTRO VALLEY WATERSHED

Because of our interest in developing techniques useful in all parts of the country and because of the need to firm all details of the procedure, we are working on several watersheds across the United States. We are completing work on the Castro Valley Watershed in the San Francisco Bay Area. The Castro Valley Watershed is very small (12.8 sq. kilometers) and highly urbanized. It has been studied in great detail by the Corps of Engineers. We have applied the operational procedure based on the clustering approach and conducted also classification by photo interpretation using aerial photographs backed up by a ground visit to Castro Valley. On the basis of results to date, we are able to classify this watershed into six different classes and with an accuracy comparable to what was achieved for the Trail Creek Watershed. Shadows are much more pronounced in the Castro Valley Watershed and may result in classification problems.

5. DISCUSSION

We feel that we are developing a technique which should be usable operationally and we are trying to finalize the procedure so that the Corps of Engineers can use it with only line printer output. We are also planning to study in the coming few months 2 to 4 additional watersheds of different sizes (ranging from 250 to 750 square kilometers) and of different terrain across the United States to determine the applicability of the operational procedure to different geographic conditions and to different constraints on availability of data

In order to make the operational procedure complete, there are several remaining problems to be solved. First, we need to improve classification accuracy within urban land use classes.

This problem appears to be difficult to solve perfectly and seems to be the major challenge for future work and possibly for higher resolution sensors. We also expect to encounter problems for large size watersheds in terms of computation time. And last, but not least, we have a continuing problem of compiling the results. In order that the classification based on LANDSAT data be integrated into a Grid Cell Data Bank, the original LANDSAT pixels should be distorted, scaled and resampled. In spite of these remaining difficulties, we are hopeful that the procedure developed will be widely useful in the common situations where specialized interactive computing equipment is not available.

6. REFERENCES

1. Phase 1 Oconee Basin Pilot Study: Trail Creek Test, Hydrologic Engineering Center, U.S. Army Corps of Engineers, Davis, California, September 1975.

2. Hardy, E. E. and J. R. Anderson, "A Land use Classification System for Use with Remote Sensor Data," Conf. Proc. Machine Processing of Remotely Sensed Data, pp. 2A 1-6, Purdue University, West Lafayette, Indiana, October 1973.

3. Ragan, R. M., T. J. Jackson, et. al., "Use of LANDSAT and High Altitude Aircraft Remote Sensing in Urban Hydrology," Department of Civil Engineering, University of Maryland, College Park, Md., January 1976.

4. Fu, K. S., D. A. Landgrebe, and T. L. Phillips, "Information Processing and Remotely Sensed Agricultural Data," Proc. IEEE, Vol. 57, No. 4, pp. 639-653, April 1969.

7. TABLES

1. NATURAL VEGETATION

   Heavy weeds, brush, scrub areas, forest woods

2. DEVELOPED OPEN SPACE

   Lawns, parks, golf courses, cemeteries

3. LOW DENSITY RESIDENTIAL

   Single Family: 1 unit per 1/2 to 3 acres; average 1 unit per 1-1/2 acres. Areal Breakdown: 5% structures; 10% pavement; 50% lawns; 37% vegetation. Proportion developed = 60%

4. MEDIUM DENSITY RESIDENTIAL

   Single Family: typical subdivision lots; 1 unit per 1/5 to 1/2 acres; average 1 unit per 1/3 acre. Areal Breakdown: 10% structure, 15% pavement, 45% lawns, 30% vegetation. Proportion developed = 70%

5. HIGH DENSITY RESIDENTIAL

   Multi-Family: row houses, apartments, townhouses, etc., structures on less than 1/5 acre lots; average 1 unit per 1/8 per acre. Areal Breakdown: 25% structures; 15% pavement; 35% lawns; 25% vegetation. Proportion developed = 100%

6. AGRICULTURAL

   Cultivated land, row crops, small grain, etc.

7. INDUSTRIAL

   Industrial centers and parks, light and heavy industry. Average 1 plant per 8 acres. Areal Breakdown: 20% pavement, 50% structures, 30% open space. Proportion developed = 100%.

8. COMMERCIAL

   Shopping centers and "strip" commercial areas. Average 3 structures per acre. Areal Breakdown: Structures 30%, lawns 5%, vegetation 10%, pavement 55%. Proportion developed = 80%

9. PASTURE

   Livestock grazing areas, ranges, meadows, agricultural open areas, abandoned crop land

10. WATER BODIES

    Lakes, large ponds, major streams, rivers

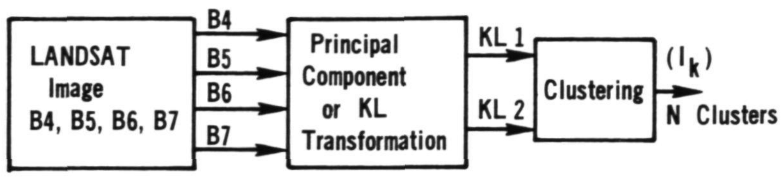
TABLE I. HEC LAND USE CATEGORIES, TRAIL CREEK WATERSHED

Land Use	Percent of Area			
	Ground Truth	Maximum Likelihood Classification	One Step Clustering	Operational Procedure
Natural Vegetation	50.17	36.19	48.69	45.06
Residential	(low density) 2.45 (medium density) 6.79 (high density) 0.11	23.78	16.60	15.31
Dev. Open Space Agricultural Pasture	0.49 28.73 3.04	8.82 19.17	26.69	32.24
Industrial Commerical	2.59 1.55	6.08 1.97	5.84	5.21
Water Bodies	0.57	1.02	0.97	0.97
Trailer Parks Highways Open Space	2.47 1.06	2.98	1.21	1.21

TABLE II. AREAL PERCENTAGE OF LAND USE AS DETERMINED FROM LANDSAT IMAGE - TRAIL CREEK WATERSHED.

8. FIGURES

Preprocessing and Clustering



Classification

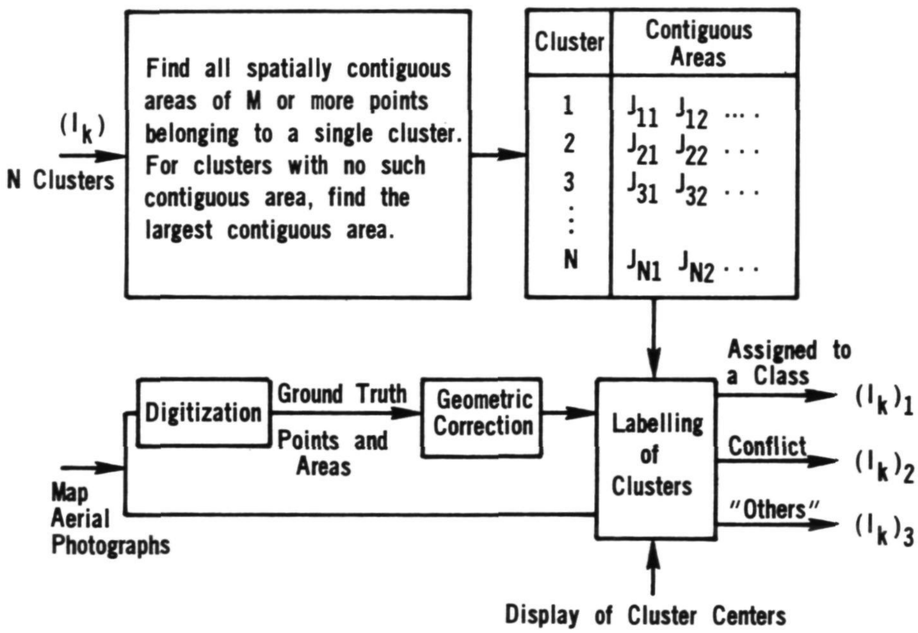


Figure 1. An Operational Procedure for Land Use Classification.

Reclustering

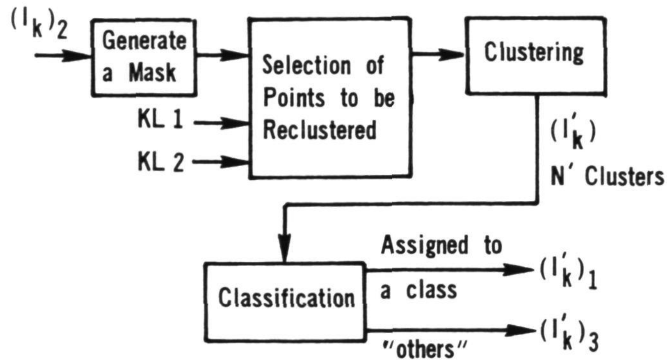


Figure 1. (Cont.)

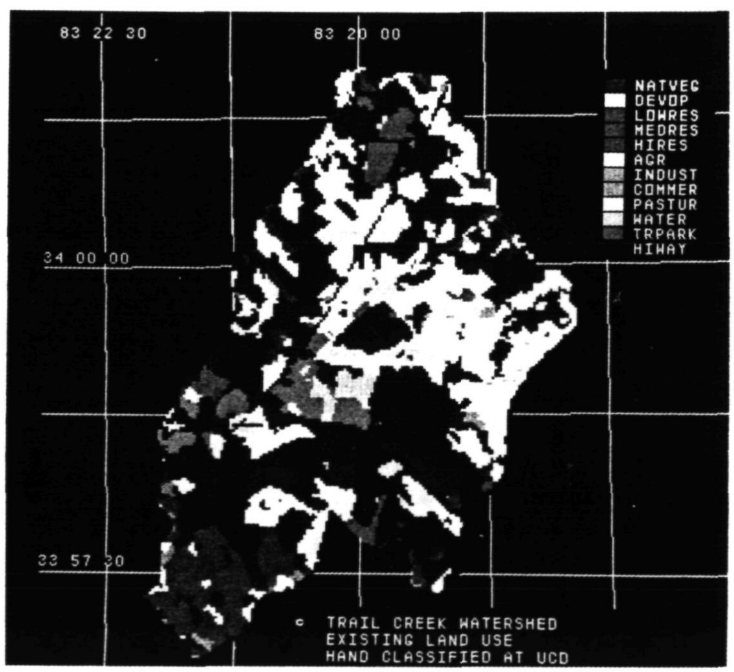


Figure 2. Trail Creek Watershed, Existing Land Use (Ground Truth)

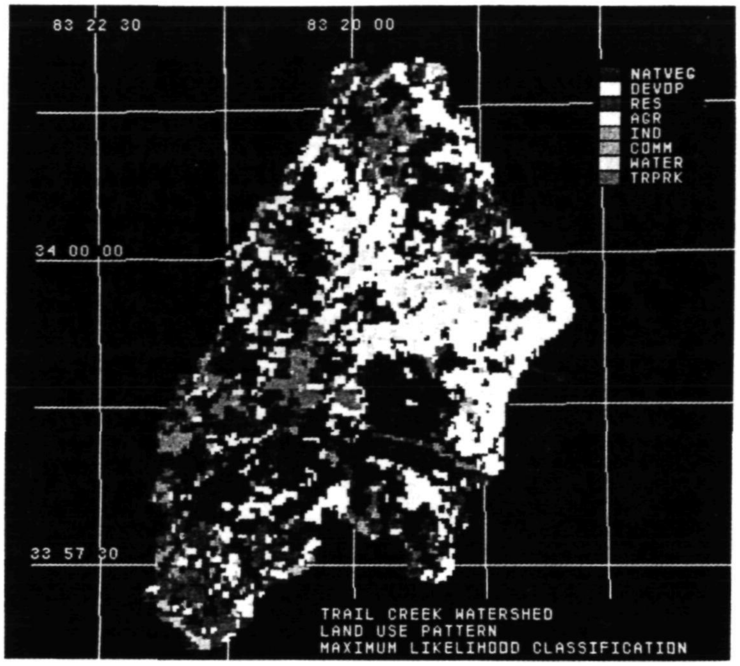


Figure 3. Trail Creek Watershed Land use Pattern, Maximum Likelihood Classifier.

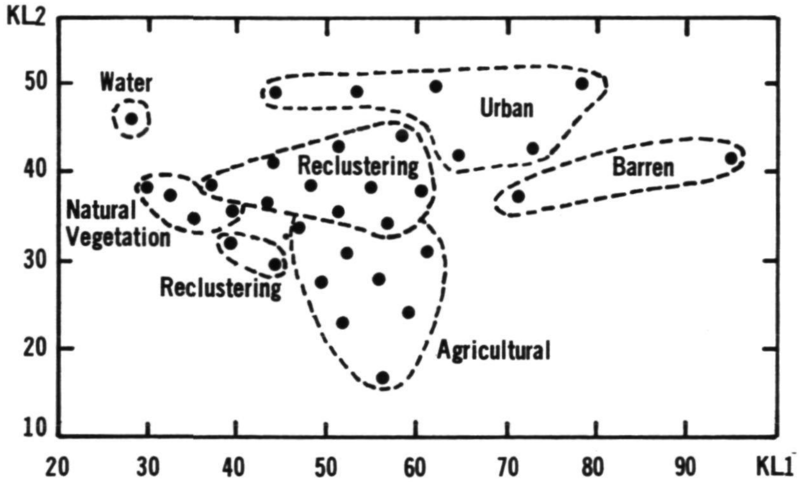


Figure 4. The Cluster Centers and the Initial Decision Made on Each Cluster.

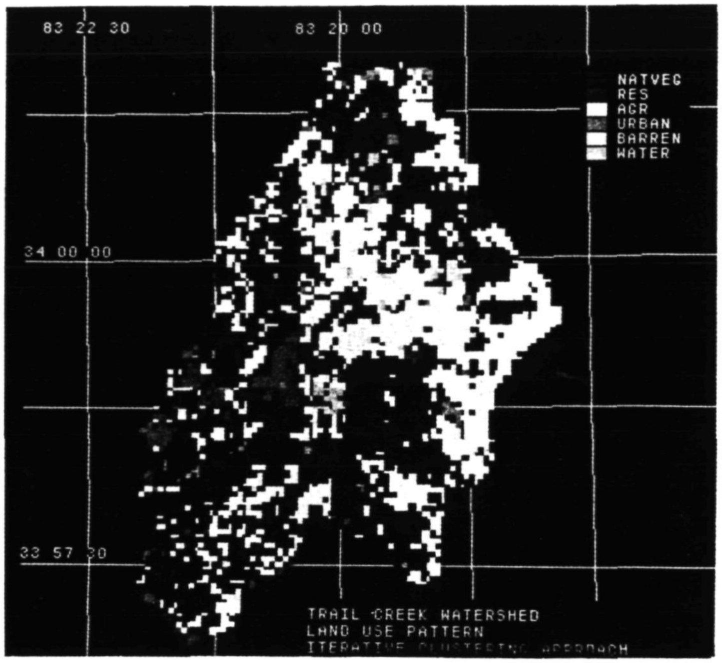


Figure 5. Trail Creek Watershed Land Use Pattern, Operational Procedure.

9. ACKNOWLEDGMENT

The research reported in this paper was supported by the National Aeronautic and Space Administration under NASA Grant NSG 5092 and NASA Grant NGL 05-003-404.