TEXAS '74
REMOTE SENSING PROJECT

Paul W. Cook
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

August 1977

TABLE OF CONTENTS

Texas '74 Remote Sensing Project

By

Paul W. Cook

I. INTRODUCTION

Our research group chose a contiguous group of thirteen counties in Texas during the spring of 1974 for a test of the crop acreage prediction capabilities of the LANDSAT I satellite (at that time it was known as ERTS or Earth Resources and Technology Satellite). Specifically, the counties chosen were as follows: Bastrop, Bell, Burleson, Coryell, Falls, Fayette, Lee, Limestone, McLennon, Milam, Travis, Washington and Williamson counties.

Much of the land in the above 13 counties is primarily used for pasture. Both sorghum and cotton are prevalent in each of these counties. Peanuts, corn, wheat and oats are grown also. SRS crop estimates indicate, however, that no crop is grown on more than 10% of the total county acreage. In fact, most of the crops, other than cotton and sorghum, are grown on only 2 or 3% of the total county area. This means that most of the land in this area is used for pasture and/or hay crops.

The ground information which was obtained affirmed this allocation of crop acreages. Most fields were of pasture land with only small areas devoted to the major crops of the county. This resulted in a shortage of training information for cotton and sorghum.

Another difficulty encountered was obtaining cloud-free imagery. Throughout the growing season, the only usable imagery obtained was for June 27 and June 28, 1974. Because of this lack of multiple time imagery, it was not possible to examine what advantages may be obtained by using multiple time periods in improving the crop estimates.

The majority of the ground data was obtained for the June 27 pass (specifically, LANDSAT Scene 1704-16231). Of the 64 segments for which there was information, 42 were on this scene. The remaining 22 segments were primarily pasture land with very few other crops.

Because of the lack of training data, the project was limited to six counties: Bastrop, Burleson, Fayette, Lee, Milam and Washington. These counties were wholly contained within the scene to be analyzed.

Location of the individual segments (square mile sample areas which are used in our operational probability surveys) proved to be a difficult task. Only 10 segments were sufficiently well located to be used in training the statistical classifier (Gausian multivariate normal assumption). Lack of contrast between fields and other vegetation within the areas containing the segments was the primary cause of our not being able to locate all the areas accurately. Fortunately, however, another 10 segments contained additional ground information that was useful for testing the classifier accuracy. Only the global scene calibration was used to locate these segments, but the accuracy of classification for these areas indicated that they were located sufficiently well for use as test fields.

This lack of a large number of fields available as training data forced a reduction in the scope of the project. Only one county was selected to be analyzed based on the best set of statistics obtained from what training data was available. We chose Milam county since it contained the highest percentage of cotton and sorghum in the six-county area. Also, the location of the 10 training segments seemed favorable for obtaining data representative of the crops in Milam county.

Thus, the 13 county project became a study in making use of only minimal information under adverse conditions. Not only was the number of fields available for training quite small, but the contrast between fields was so minimal that it prevented any visual determination of their correct location. This meant that only the correspondence of reported to digitized acreage could be used for determining what fields to use as training or test data.

The results of the study seem to indicate that more data for training and test is a necessity to make accurate acreage estimates. Small quantities of data cause instability in the estimates of the means and the variance-covariance matrices of the clusters for the various crops. Thus, small changes in what data is clustered and how the clusters are combined can produce radical alterations in the classification accuracy as evidenced by the overall classification accuracies varying from 40 to 75% in this study. (See Appendix D)

Various methods were used to cluster the training data for purposes of classification. Combining all the grasses into pasture categories and then clustering this group while cotton and sorghum were each clustered separately proved to be the best grouping. When only one group was used for pasture, this group showed too great an overlap with the sorghum and cotton groups for accurate classification results.

Since the data base was small, it was difficult to evaluate the effects of classification accuracy on the acreage estimate itself. Conceptually at least, higher accuracy rates would seem to improve the final estimate. Higher accuracy did seem to give better estimates for the direct acreage expansion method, but the data did not allow us to determine whether greater accuracy would improve the estimate while using ratio or regression estimation.

II. REGISTRATION DEVELOPMENT

Before the satellite data can be classified, appropriate training information must be located. The segments in the June Enumerative Survey were meant to provide random samples of data representing the various crops which occur in a given area or, in our case, within a LANDSAT scene.

Locating the segments within the LANDSAT CCT (Computer Compatible Tape) data proved to be a formidable task until means were developed to utilize mathematical equations (called affine transformations) for relating the LANDSAT coordinates and map base latitude and longitude. This process of correlating the two sources of information is called registration.

First, registration points (i.e., points of high contrast in the imagery and map base) are located (see Appendix A). Latitude and longitude values for such points are then obtained from a USGS 7 1/2' quadrangle map (where available) to the nearest .0001 degree of latitude and longitude and the appropriate LANDSAT coordinates are also obtained to the nearest .1 pixel. A pixel (or picture element) is the smallest resolution element of the satellite, and is about 79 meters long by 57 meters wide. Therefore, the map accuracy corresponds to about 11 meters on the earth's surface, whereas the grey scale gives about eight meters placement accuracy for the LANDSAT data.

During the early experimentation with the techniques for registering LANDSAT data to map base, only small 400 by 400 pixel areas were used for locating registration points (typically, about 10 were used in each area). A set of linear affine transformations related the map coordinates to the satellite coordinates (see Appendix B). This method allowed us to obtain maximum residual errors of 150 meters on the ground.

Next we did a second step of registration for each segment by locating six more registration points near each segment. These points were again checked with the larger area registration points to be certain that they were accurately located as well. These six points whose locations had been accurately determined were then used to locate the field boundary locations for each segment so registered. For this second step of registration maximum residual errors typically ranged in the 30-60 meter range.

Two different systems for improving our initial methods of registration are now available. First, there is a program developed by Ed Schlosser of Lockheed (Johnson Space Center). It provides for full frame registration with the use of only 16-20 control points well scattered about the scene. Accuracies for an entire scene vary, but may often be on the order of 100 meters root mean square error or less (See Appendix C). This program is available as DAM-COEF, coded as a SAS (Statistical Analysis System) procedure.

The other registration program is by Walt Donovan of the Center for Advanced Computation at the University of Illinois. Donovan's program uses a second order full term polynomial for the global calibration with residual error on the order of 300-400 meters (i.e., 4-5 pixels). A second step can then be either a local linear registration or (newly developed since this Texas project was completed) a movement of the entire segment file by fractional row and column pixel increments.

III. DIGITIZATION

After a LANDSAT scene is registered to map base, those LANDSAT pixels which correspond to the appropriate areas of interest must be extracted. Of course, appropriate data for training our classifier is required. The field boundaries must be accurately located in order to obtain the correct pixels for training the classifier. Another requirement is to locate those areas for which we wish to make acreage estimates.

Because the satellite does not produce its data relative to an exact map scale, the LANDSAT data does not directly overlay onto maps. Consequently, a method of making the boundary data available in computer compatible form whereby the corresponding pixels for each field (or any other area) could be extracted for analysis purposes as accurately and easily as possible was needed.

A digitizer can give the location (i.e. an (x,y)-reading) of any point on a map (or photo or drawing, etc.) to either .01" or .001" (depending on how accurately the tablet and controller are designed to perform). It does this by measuring the time it takes a current to travel from a point on the tablet back to the controller, an electronic timing device.

An interactive analysis system (EDITOR) programmed on the computers (PDP-10's) operated by Bolt, Beranek and Newman (BBN) at Boston, Massachusetts was used to analyze the data. Walt Donovan and Martin Ozga of the Center for Advanced Computation (CAC) at the University of Illinois provided the programming expertise in developing software. Many improvements and modifications of their programs were made as a result of the analysis of the Texas data.

An Altek digitizer with .001" accuracy is directly interfaced into the computer network so that processing of the digitized data may be done immediately. Only individual data points defining closed boundaries are selected to create the files corresponding to field, segment, or strata boundaries. Consequently, individual points for each area to be digitized must be marked. For regular boundaries, only the vertices need be marked, but for irregular edges, one must indicate enough points to approximate any curved lines.

During digitization, calculations are made to determine the acreage of each enclosed area digitized. This allows for a check against reported acreage. As an example, when digitizing a segment each field is digitized separately and stored onto a common file for that particular segment. During this process, no point is recorded more than once. Thereby, no common boundaries may overlap and common edges are preserved. A checking program may be run as well to make sure no areas have been digitized incorrectly.

Each digitized file is first converted to geographic latitude and longitude values by means of calibration points for which latitude and longitude are known. This process prepares the file for use with the satellite-to-map-base registration coefficients. In this way, one obtains the location of the field boundaries (or in a more general sense, for land-use boundaries on a county or state level) by which the appropriate LANDSAT pixels may be examined.

Digitizing a segment creates what we call a segment file. Included in this process is the calibration to latitude and longitude values. Creation of a mask file is done by using a calibration or registration file of coefficients to transform the original file into LANDSAT coordinates. The mask file is then used to extract the appropriate LANDSAT data in a process called packing a window. A packed window file contains the LANDSAT data tagged appropriately as to whether each pixel is within a field and has a cover type, on a boundary between fields, or background (i.e., has an unknown cover type). See reference 2 for more details.

This process of packing a window allows the selection, in various ways, of data to be used in developing the training statistics for each cover type. Either a specific cover type may be chosen or specific fields from selected segment mask files may be chosen. Also, specific fields may be denoted as training or test fields so that selection only on that basis may be made as well. Boundary pixels which overlap fields may be either included or excluded as desired.

IV. DATA ANALYSIS

In Texas, much of the data was not usable for determining the crop classes. Inaccurate field boundaries (as indicated by rather large discrepancies between reported and digitized acreage), fields of burnt or abandoned crops, and pasture land of uncertain content made selection of training and test fields a difficult task.

Because these effects could not be easily removed from the field data, we chose to eliminate problem fields from analysis and concentrate on those characteristics most closely matched to the crop information of interest. This subset of fields was used to obtain the statistics for classifying the remainder of the scene. Also, only those segments for which a local calibration file could be made were used for training data so as to insure that the data was accurately located.

All twenty-one available segments were located by means of the full second order linear equation provided by Donovan. Unfortunately, the lack of adequate contrast between fields did not permit the determination of how accurately the fields were located. This made it necessary to perform a second step of registration, which would assure that each field was located to an accuracy of about 30-60 meters (i.e., less than a pixel's width).

The lack of adequate contrast between fields also made the location of points for use in local registration difficult. Usually, small woods and rivers proved to be the best areas for the determination of the calibration points. For regularly shaped areas, the centers of woods were best, whereas for irregularly shaped areas or twisting rivers, the sharp edges of the bends worked well. In each case, an area no larger than 100 pixels by 100 pixels was used to ensure that the linear equations remained appropriate.

Of the twenty-one available segments, ten were successfully registered locally. Ten other segments seemed to be reasonably close for use in testing, while one segment defied all attempts to determine a correct location for it. Consequently, only the ten locally registered segments were used for obtaining the crop signatures. The remaining ten segments were then used to obtain part of the test fields. Also, the original ten segments' fields were included in the test data in order to provide sufficient test data since the ten locally registered segments contained about two-thirds of the available cotton and sorghum pixels.

The list of fields with boundary pixels on page 9 points out how little data was available for the project. The first

Segment Fields with Boundary Pixels

Segment	Sorghum		Cotton		Pasture1/	
	Pixels	Acres	Pixels	Acres	Pixels	Acres
1					204	390
2	10	79	285	350	8	15
3						
4	27	50			147	196
5	2	25	1	25	29	50
6	64	161	4	60	95	128
7	0	34			126	218
8	73	121	212	314.5	28	136.4
9			2	10	210	234
10					291	406.2
11	1	54			447	531
12	12	210			86	103
13	64	196.8	21	65	89	212.5
14					352	436
15					411	605
16	5	39	4	30	52	67
17					109	326
18	6	39			157	284
19					248	286.5
20	0	5	41	105	388	470
21					240	259
TOTAL	264	1013.8	570	959.5	3717	5353.6

1/ Pasture with trees; stubble; sudan stubble; Johnson grass; hay; wheat stubble; oats; oat stubble; idle; coastal pasture; pasture eleven segments listed (segment 3 could not be adequately located) were the ones which made up the training data since they had most of the cotton and sorghum data and also were locally registered. The remaining ten segments contained primarily pasture test data with only a few pixels of sorghum and cotton.

During the course of this project, Donovan developed software to drive a ZETA (x,y)-plotter to make plots of the digitized segments both to map and LANDSAT scale. This allowed further checks on the locational accuracy of each segment as well as shifting of each segment to match field boundary outlines (where such outlines were discernable).

This development of plotter software has allowed a check of the presently available registration methods. Plots of the field outlines were made for each segment by using the second-order polynomial, the linear local fit and DAM-COEF to compare the results of each method.

The actual location could not be compared for each method because photo-interpretation techniques were not satisfactory to determine the location for each segment. However, the orientation and size of the plotted fields could be compared. If the local fit is considered as correct, then DAM-COEF was closer to the way in which the local fit showed the segment. In all cases, the second-order fit showed the fields to be tilted differently and to be somewhat smaller than in the case of the linear fit.

Although the best method seems to be the linear local fit, great care must be taken in performing the local calibration step. Points chosen as local registration points must be checked with the global calibration point file to be sure that they are properly chosen. These points will generally be outliers as far as the global fit is concerned, but there should be less than a two pixel error for them to be acceptable. The rms error for this local fit should be less than .5 pixel in either line or column while the maximum error should be less than .6 pixel. This will allow quite accurate location of the field information within each segment.

After the ten segments had been located, a first attempt at unsupervised data analysis was made. Fields for which the ground data was deemed satisfactory (i.e., the reported acreage and digitized acreage checked within 10% and the crop or use information agreed) were chosen from the 10 locally registered segments and packed into one large file. No boundary pixels were included.

The cover types included in the fields chosen were as follows: corn, sorghum, permanent pasture, cotton, pasture, hay-grazer, coastal pasture, idle land, oats, hay, Johnson grass, sudan stubble, cucumbers, milo, stubble, pasture-with-trees, sorghum stubble, other vegetables, plowed land, and cotton. Since some of these categories contained very little training data and were very similar spectrally, it seemed apparent that not all of these groups would be readily separable.

EDITOR uses a clustering routine based on the ISODATA algorithm in which the four-dimensonal space is partitioned into clusters or groups based upon a Euclidean distance criterion for separation of clusters. For the first attempt at making this separation all the data available within the ten registered segments was used to set up an appropriate number of categories (twenty) to correspond with the number of different crops represented. This clustering of the LANDSAT data corresponding to these selected fields was a first attempt at determining what spectral groups occured naturally in the scene.

Based on the initial unsupervised clusters, no apparent separation of crop types seemed possible. The cluster groups' means and their variance-covariance matrices when used to classify the crop groups picked up most of the cotton and sorghum as though they were pasture groups. Consequently, it was decided that for this data set the clustering technique available was unable to separate known crop types without supervision.

When only cotton acreage was examined it was noted that certain fields belonged almost entirely to one of two spectrally separable groups based on clustering cotton into seven groups. Examination of all the ground information indicated that the cotton fields contained two quite distinct types of cotton. One variety is quite short - usually about 16" - 20" tall during the June visit whereas the other is quite tall - usually 30" - 36" tall at this time. Also, the two varieties of cotton require quite different methods of harvest. The short variety is harvested with a stripping machine whereas the taller variety requires a picking machine.

Although detection of differences within the same crop type had not been the purpose of this project, the forethought of providing supplementary ground information forms (as developed by Fred Warren) had made it possible to determine that clustering had shown the existence of two spectrally different cotton groups.

To determine if the forming of additional clusters within each crop category would reduce the overlap of categories, the cotton pixels were clustered into 16 categories, sorghum into 10 categories, and the pasture group into 8 categories. For each crop the statistics file was run through a cluster combining program (developed by Bob Ray of CAC). Based on the results from running this program the cluster classes were combined as follows: the cotton clusters were combined into six groups, the sorghum clusters were combined into six groups and the pasture clusters were combined into eight groups. These 20 groups' means and their variance-covariance matrices were combined into one statistics file for later classification tests.

Cluster groups were only combined when they showed little informational loss and they were not of the same category type. When a group of fewer than eighty (80) points could not be combined into another group, it was deleted.

Even with this large number of clusters for each group, considerable overlap between clusters was noticeable as indicated by the low rates of classification accuracy and the frequent classification of cotton and sorghum into the pasture classes. Also, the cluster combining program indicated that little informational loss would occur from combining some of the cotton and sorghum clusters with those from the pasture clusters.

Additional classifications of the selected fields were done using various numbers of clusters applied to all the pixels whose locations were accurately known. The accuracy of classification results varied considerably from one group of clusters to another. In no case did the classification accuracy exceed 80%, though low accuracies of nearly 40% did occur. This instability of the cluster group statistics' classification accuracy seems to be a further indication that crop types for this particular LANDSAT scene are not distinct spectrally.

Based on ten groups of clusters, the statistics were determined and classification of Milam County on the ILLIAC IV at Moffett Field, California was undertaken.

An attempt was made to improve this classification by applying weights to each category based on crop acreage estimates made by SRS for 1972. Assuming each pixel to represent 1.14 acres, the following acreage estimates were obtained (compared to 1974 SRS estimates).

	1974 SRS ESTIMATES	ACRES EQUAL PRIORS	UNEQUAL PRIORS
Sorghum	57,700	45,501	7,731
Cotton	31,500	74,303	15,240
Pasture/Stubble	N/A	536,840	632,416
Water*	959	1,093	1,087

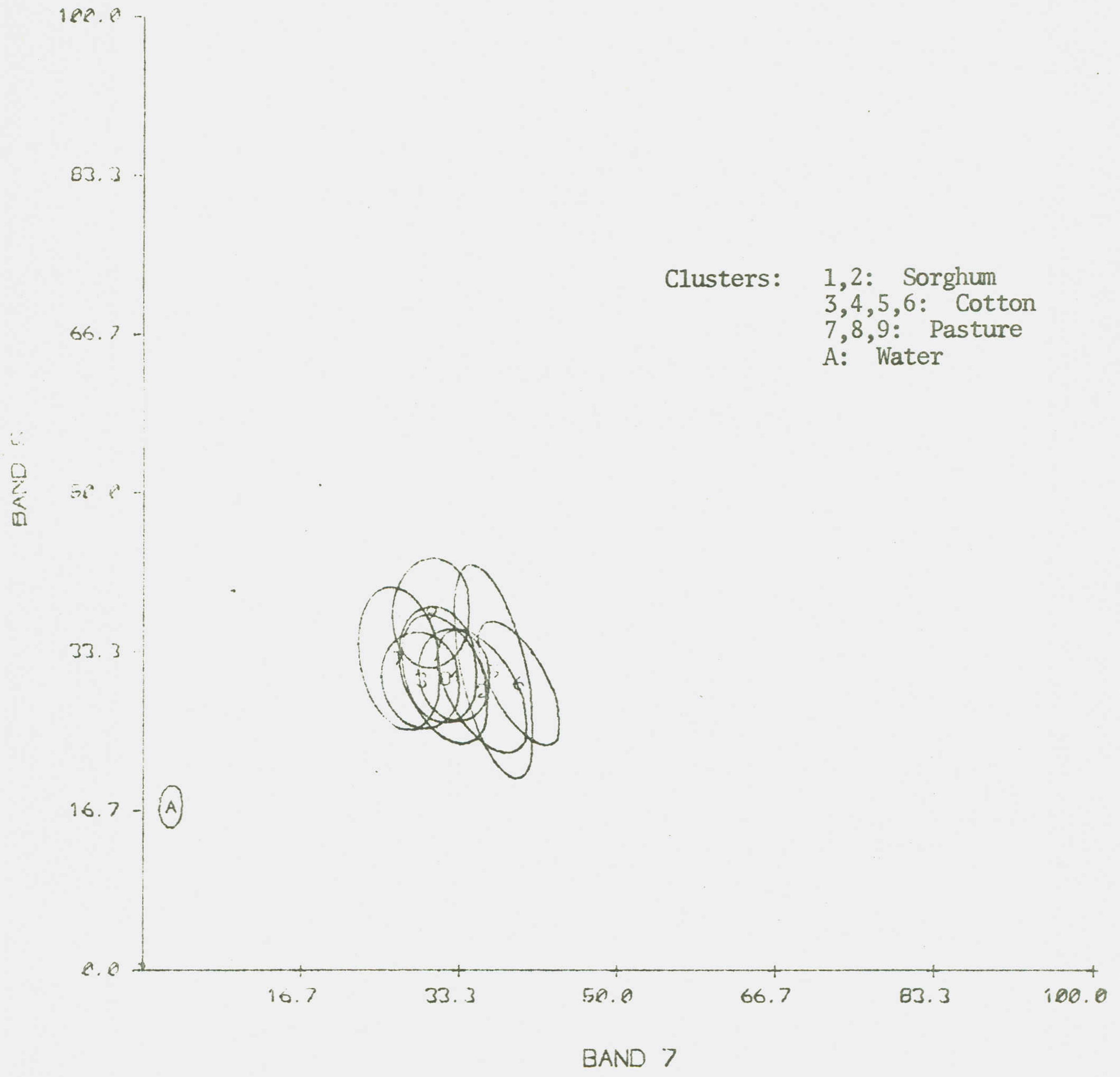
*Data from Texas Water Resources Board

A computer program developed by D. Christodoulidis of Ohio State University further indicates why such large changes in acreage may occur. The circular Fourier series plot of the mean vectors and variance vectors show little differences between cluster groups except in the case of water. This confirms the high separability of water from the agricultural crops for this LANDSAT scene. Consequently, in the discriminant space one would find the water vector separate from the other groups. However, the cotton, sorghum and pasture/stubble vectors show a great deal of overlap. (See Appendix D.) Further evidence of the overlap between cluster groups is shown by a more recent program developed by Bob Starbuck of our group. It shows rather extreme overlap among groups within the two dimensional vector space of bands 5 and 7. (See page 14 to examine this plots.) This presence of similarity or overlap between cluster groups would explain why prior probabilities applied to the cluster groups would yield widely varying acreage estimates.

Appendix E contains the regressions and plots of data from the usable fields within the twenty-one segments that contained the most reliable ground information. Unfortunately these plots of reported acreage vs. classified pixels within each field (including boundary pixels) do not indicate that a linear estimation technique would be valid for the analyzed data.

90.0% CONCENTRATION ELLIPSES FOR BANDS 7 AND 5

TEN CLUSTERS FOR MILAM COUNTY



(USDA-SRS) MILAM.CF;1,15-Jun-76 15:32:56

V. CONCLUSION

This study of Milam county, Texas has pointed out a number of problems which are present in analysis of the LANDSAT data using our presently available ground data. Solutions to some of these procedural problems were developed during this study.

Because of the difficulties in determining the accuracy of field sizes and crop types for certain segments, our present projects have included greatly improved checking procedures for field acreages and crop types. Data is collected in the same manner, but with more emphasis on the accuracy of field placement, of reported acreage from the farmer, and digitized acreage from aerial photography (either infra-red or ASCS acquired enlargements). Our review procedures have greatly improved the accuracy of field location and sizes and thereby allow the use of regression estimation techniques with full segment data.

Accurate location of training data is the next requirement for use of the digital LANDSAT data. Registration techniques were developed as part of this study that now allow the location of field boundaries to within 40-60 meters of their actual location. This means that the data chosen as training or test is definitely of a given crop type. Consequently, full use of the data in making regression estimates of crop acreages is now possible.

Two techniques of analysis used in this project need to be tried on larger data sets to see if they contribute to overall accuracy and efficiency. Clustering the various crops into subgroups determined by the information loss program provided by Bob Ray seemed to improve the overall accuracy as well as the field-by-field accuracy of classification. The overlap between cotton and sorghum seemed to be somewhat reduced by this method of cluster combining. Indeed, two varieties of cotton could be separated this way. The tall (or picker) cotton seems to be quite separable from both the short (or stripper) cotton as well as from sorghum.

Use of equal priors with the cluster groups seemed to give the best accuracy results. Also, the estimates using equal priors were much more reasonable.

The great spectral confusion of the various crops for this scene does point to the importance of knowing or having imagery for the time at which the crops are most separatable spectually. Quite possibly for another time of the year, the crops could be separated much more easily. In fact, with many crops having different growing and harvesting times, multi-temporal imagery may make a more accurate crop classification possible.

However, one must also acknowledge the limits of the sensor to differentiate among the various crop types. Such similar crops as oats, hay and Johnson grass did not seem to be distinguishable either from one another or from the pasture and idle land categories. This finding does not seem unreasonable since the pasture category is probably a mixture of all these grasses as well as some others not mentioned. Also, all these categories look similar when viewed in infra-red photography.

APPENDIX A

Some Notes On Registration Points

1. Generally only bands 5 and 7 need be used for locating registration points. Band 7 is particularly useful for water bodies and land-water interfaces. In band 5 areas of vegetation, such as trees, and interstate highways are particularly easy to find.

2. Areas where large variations may occur during wet or dry weather conditions should not be used. For example, bends in rivers where the map indicates the sides of the river have only very gentle rise should not be used. However, steep sides of a river or lake will usually provide a stable relationship between land and water and thereby make more promising registration points.

3. When using the center of a lake or wooded area be sure that it is relatively small, i.e., no more than 30 pixels in area. However, for irregularly shaped lakes or vegetation areas, you may use points within the centers of areas jutting out from its central mass.

4. Choose points such that the point may be defined as a single pixel. This simplifies location of the correct pixel and makes it easier to enter the correct LANDSAT coordinates into the program.

5. When choosing road intersections in band 5, be sure to choose only narrow roads that form a distinct intersection point. Large road intersections generally are difficult to determine a precise meeting point.

APPENDIX B

The Affine Transformation

Ralph Bernstein (See Reference 9) notes the various components of error for the MSS as being due to the following:

Platform Effects

Altitude: Scale distortions occur along scan and vary with time. Magnitude of correction - about 1.5 km.

Attitude: The three main components of attitude are pitch, roll and yaw. Yaw and pitch changes affect the data spacing along the flight path, while roll changes it along the scan line. Magnitudes vary, but pitch is about 12 km, roll 12 km, and yaw 2.46 km.

Velocity: Changes in velocity may produce 1.5 km of along-track scale distortion.

Scene Effects

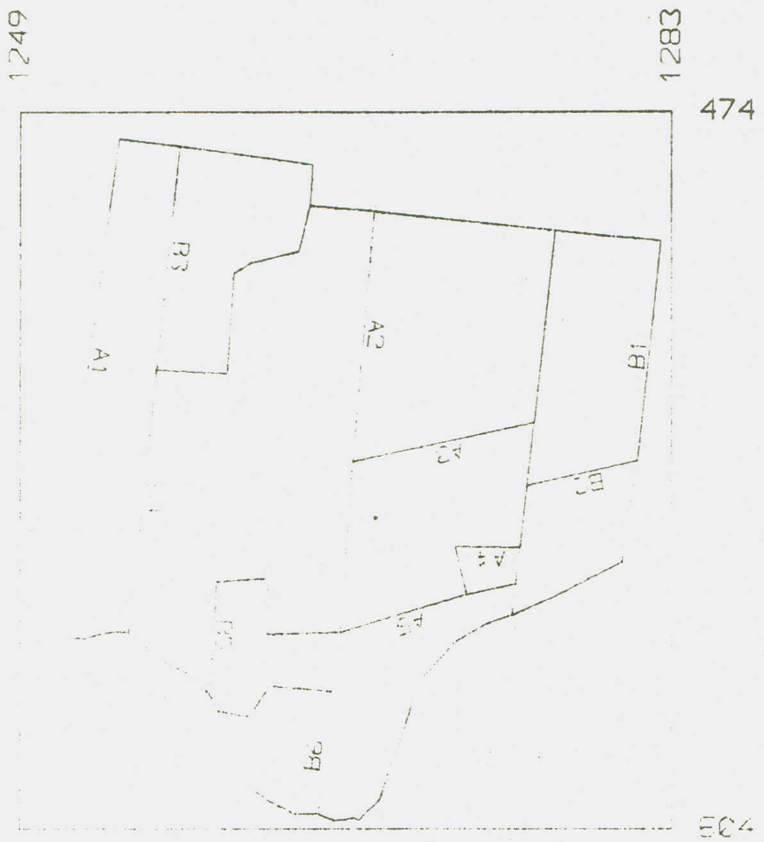
Earth Rotation: While the MSS mirror completes each scan, the rotating earth below it causes along-scan distortion of about 13.3 km.

These effects require the indicated corrections when we consider the entire 180 km by 180 km image. However, for only small areas of the MSS image, such effects are reduced and a simple first oder polynomial (or affine transformation) can adequately model these various platform and scene effects. The equation is of the following form:

$$\text{LANDSAT line} = a \cdot \text{latitude} + b \cdot \text{longitude} + e$$
$$\text{LANDSAT column} = a' \cdot \text{latitude} + b' \cdot \text{longitude} + e'$$

Location of the registration points by latitude and longitude allows estimation of the parameters of the equation and thereby prediction of the location of geographic coordinates in terms of the scanner coordinates. By taking the inverse of the equations, map coordinates of the indicated MSS coordinates of interest may be predicted. Thus, every geographic point is cross-referenced with its corresponding scanner location.

```
THE ESTIMATING EQUATIONS ARE:

     LAT =  0.3950044E+02 + (-0.6937944E-03 * ROW) + (-0.1009679E-03 * COL)
     LON =  0.8907613E+02 + ( 0.2393281E-03 * ROW) + (-0.6464599E-03 * COL)

     ROW =  0.3540331E+05 + (-0.1367606E+04 * LAT) + ( 0.2136158E+03 * LON)
     COL =  0.1508894E+06 + (-0.5062688E+03 * LAT) + (-0.1467729E+04 * LON)

R-SQUARES:

        LAT       LON       ROW       COL
     ----------------------------------------
     .999959   .999957   .999965   .999949

           OBSERVATIONS                          RESIDUALS
  ROW     COL     LAT       LON            ROW     COL      LAT         LON
---------------------------------------------------------------------------------

2209.55  335.55 38.22913  89.37992        -4.35  -14.22  -0.00446    -0.00810
2041.21  592.83 38.32126  89.17812        -3.59   -6.49  -0.00315    -0.00329
1740.20  718.39 38.51826  89.02606        -2.70   -3.88  -0.00226    -0.00182
2143.72  958.07 38.21523  88.96441        -0.43   -8.59  -0.00118    -0.00542
1531.68  338.99 38.70022  89.22584        -5.04    1.57  -0.00333     0.00227
1479.85  251.92 38.74561  89.27045        -4.32    2.95  -0.00269     0.00300
 270.41 1278.12 39.49171  88.30883        12.03   -4.52   0.00793    -0.00577
 330.49 1194.34 39.45187  88.39096         0.07   12.08   0.00131     0.00782
2137.91 1912.12 38.12562  88.35146         2.15    0.44   0.00151    -0.00023
 363.92 2038.46 39.34312  87.84888         0.58    5.52   0.00099     0.00343
 184.27 2414.42 39.42772  87.56016        -1.70    0.54  -0.00110     0.00075
 137.91 2543.98 39.44119  87.43627        -3.17   -4.91  -0.00267    -0.00243
 129.78 2738.86 39.43060  87.33326        -3.78   -6.58  -0.00327    -0.00337
1411.20 1490.80 38.67095  88.46188        -2.36   17.28   0.00011     0.01175
 251.67 3016.70 39.31719  87.18145        -4.56   -8.97  -0.00405    -0.00474
 366.38 3092.44 39.23127  87.15800        -2.35  -11.15  -0.00274    -0.00668
1862.24 3053.13 38.19596  87.54996        -6.11    0.68  -0.00420     0.00187
 687.98 2830.52 39.03779  87.40938         0.95   -2.07   0.00046    -0.00159
 471.24 2777.13 39.19261  87.39173        -0.29   -2.99  -0.00049    -0.00188
1166.09 2621.40 38.72694  87.65863         0.69   -2.73   0.00020    -0.00195
1162.31 2836.18 38.70723  87.51924        -0.26   -2.52  -0.00044    -0.00159
1417.06 2903.53 38.52334  87.53917        -1.26    0.98  -0.00079     0.00091
1607.48 2660.95 38.41410  87.74376        -3.94    3.39  -0.00241     0.00311
1877.86 2383.29 38.25753  87.98696         0.36    3.41   0.00057     0.00210
1939.52 2552.75 38.20030  87.69145         4.16    3.71   0.00323     0.00138
1939.37 2552.07 38.20384  87.88903         9.36    1.27   0.00660    -0.00144
1960.25 1735.45 38.26387  88.41864        -0.80   -7.63  -0.00134    -0.00474
1952.45 1147.56 38.33448  88.79339         7.92   -9.75   0.00450    -0.00817
1148.52  420.93 38.96086  89.08478        -1.61    8.43  -0.00024     0.00589
1361.65 1680.11 38.68864  88.31739         3.15    3.47   0.00254     0.00150
 747.44 2605.35 39.02040  87.56938         2.44   -1.21   0.00158    -0.00138
1669.42 1865.60 38.45839  88.25760         8.80    4.63   0.00657     0.00089
 438.39 1573.09 39.34013  88.16915         2.54    8.70   0.00267     0.00504
2039.93 1728.57 38.21196  88.44722         1.78    1.15   0.00134     0.00032
 845.42 2778.20 38.93438  87.48202         1.45   -0.13   0.00099    -0.00045
1712.18 2590.33 38.34863  87.81510        -4.02    4.32  -0.00237     0.00374
1833.54 3017.53 38.22383  87.56783        -0.52    5.42   0.00016     0.00359
2075.79 3147.97 38.04085  87.54130        -2.85    4.29  -0.00158     0.00341
1469.70 2720.10 38.50765  87.67169         1.61    4.12   0.00152     0.00225
```



```
SEGMENT 4435. PART 1 OF 1,   11 FIELDS  69 EDGES  59 VERTICES
TOTAL AREA        706.8 ACRES, SCALE IS 1.  24075
PLOT FILE <USDA-SRS>4435.SEG;8,11-MAR-76 13.36.25
USER IS   BILL CALLIS
CAL  FILE <USDA-SRS>4435.CAL;1, 4-FEB-76 13.10.06
```

APPENDIX C

The DAM-COEF procedure (as developed by Ed Schlosser, Lockheed Corp. at the Johnson Space Center, Texas) provides a reasonably accurate means of registering LANDSAT data to map base without requiring that the LANDSAT data be altered--either geometrically or radiometrically. Accuracies for the location of predicted areas seem to be on the order of 100 meters when a control point network is carefully chosen. A diagram which spatially lists the control points chosen greatly facilitates the choice of points to ensure that the scene is covered by a well-scattered network of points.

By modeling the irregularities of the sensor's operation and its platform's location, Schlosser has made it possible to use a first order affine transformation to relate the LANDSAT data to map coordinates. Major components in the model are a consideration of the sinusoidal mirror velocity profile (affecting the pixel size across scan), the satellite's altitude (scale distortions), and its pitch and roll. Also, to further make the linear model appropriate, all latitude and longitude values are converted to the Universal Transverse Mercator (UTM) projection.

A further advantage is that DAM-COEF requires approximately twenty points to register an entire scene to within about 100 meters accuracy. This represents a significant time savings over use of higher order polynomials which require upwards of 100 points to ensure sufficiently accurate location for our work. Its major disadvantage, at least for now, is its inability to allow for multitemporal data analysis in the same manner which is presently available, i.e. making the pixels of two or more scenes overlay so as to form an eight channel (or more) tape with each pixel having data for two or more dates available for it.

APPENDIX D

The next two pages give two examples of classification accuracy based on only the training data using different methods of obtaining the classification statistics. As can be noted, considerable differences were evident.

Following these pages are the results of the final ten cluster groups used. Accuracies of these groups used in classification are given for both the training and test data sets. Also, charts showing the similarities among the different clusters and the overlap between cluster groups are displayed.

```
             CATEGORIZED TO:
             FROM:
                  :          :   SORGHUM STUBBLE:          :     :
                  :          :PASTURE WITH TREES:          :     :
                  :          :           STUBBLE:          :     :
                  :          :     SUDAN STUBBLE:          :     :
                  :          :     JOHNSON GRASS:          :     :
                  :          :               HAY:          :     :
                  :          :              OATS:          :     :
                  :          :              IDLE:          :     :
                  :      MILO:   COASTAL PASTURE:          :     :
                  :HAY GRAZER:           PASTURE:COTTON LOW:     :
                  :   SORGHUM: PERMANENT PASTURE:    COTTON:TOTAL:
              MILO:          :                  :          :     :
        HAY GRAZER:          :                  :          :     :
           SORGHUM:       154:                67:       135:  356:
   SORGHUM STUBBLE:          :                  :          :     :
PASTURE WITH TREES:          :                  :          :     :
           STUBBLE:          :                  :          :     :
     SUDAN STUBBLE:          :                  :          :     :
     JOHNSON GRASS:          :                  :          :     :
               HAY:          :                  :          :     :
              OATS:          :                  :          :     :
              IDLE:          :                  :          :     :
   COASTAL PASTURE:          :                  :          :     :
           PASTURE:          :                  :          :     :
 PERMANENT PASTURE:       104:               478:       338:  920:
        COTTON LOW:          :                  :          :     :
            COTTON:        76:                60:       534:  670:
             TOTAL:       334:               605:      1007: 1946:



PER CENT CORRECT BY COVER

                 MILO
           HAY GRAZER
              SORGHUM    43.26%
      SORGHUM STUBBLE
   PASTURE WITH TREES
              STUBBLE
        SUDAN STUBBLE
        JOHNSON GRASS
                  HAY
                 OATS
                 IDLE
      COASTAL PASTURE
              PASTURE
    PERMANENT PASTURE    51.96%
           COTTON LOW
               COTTON    79.70%


OVERALL PERCENT CORRECT=    59.92%
```

```
             CATEGORIZED TO:
             FROM:
                  :          :   SORGHUM STUBBLE:          :     :
                  :          :PASTURE WITH TREES:          :     :
                  :          :           STUBBLE:          :     :
                  :          :     SUDAN STUBBLE:          :     :
                  :          :     JOHNSON GRASS:          :     :
                  :          :               HAY:          :     :
                  :          :              OATS:          :     :
                  :          :              IDLE:          :     :
                  :     MILO:   COASTAL PASTURE:          :     :
                  :HAY GRAZER:           PASTURE:COTTON LOW:     :
                  :   SORGHUM: PERMANENT PASTURE:    COTTON:TOTAL:
              MILO:          :                  :          :     :
        HAY GRAZER:          :                  :          :     :
           SORGHUM:       193:                76:        87:  356:
   SORGHUM STUBBLE:          :                  :          :     :
PASTURE WITH TREES:          :                  :          :     :
           STUBBLE:          :                  :          :     :
     SUDAN STUBBLE:          :                  :          :     :
     JOHNSON GRASS:          :                  :          :     :
               HAY:          :                  :          :     :
              OATS:          :                  :          :     :
              IDLE:          :                  :          :     :
   COASTAL PASTURE:          :                  :          :     :
           PASTURE:          :                  :          :     :
 PERMANENT PASTURE:        90:               743:       135:  968:
        COTTON LOW:          :                  :          :     :
            COTTON:        39:                63:       568:  670:
             TOTAL:       322:               882:       790: 1994:


PER CENT CORRECT BY COVER

                MILO
          HAY GRAZER
             SORGHUM    54.21%
     SORGHUM STUBBLE
  PASTURE WITH TREES
             STUBBLE
       SUDAN STUBBLE
       JOHNSON GRASS
                 HAY
                OATS
                IDLE
     COASTAL PASTURE
             PASTURE
   PERMANENT PASTURE    76.76%
          COTTON LOW
              COTTON    84.78%


OVERALL PERCENT CORRECT=    75.43%

CPU TIME: 6.20  ELAPSED TIME: 3:57.18
NO EXECUTION ERRORS DETECTED
```

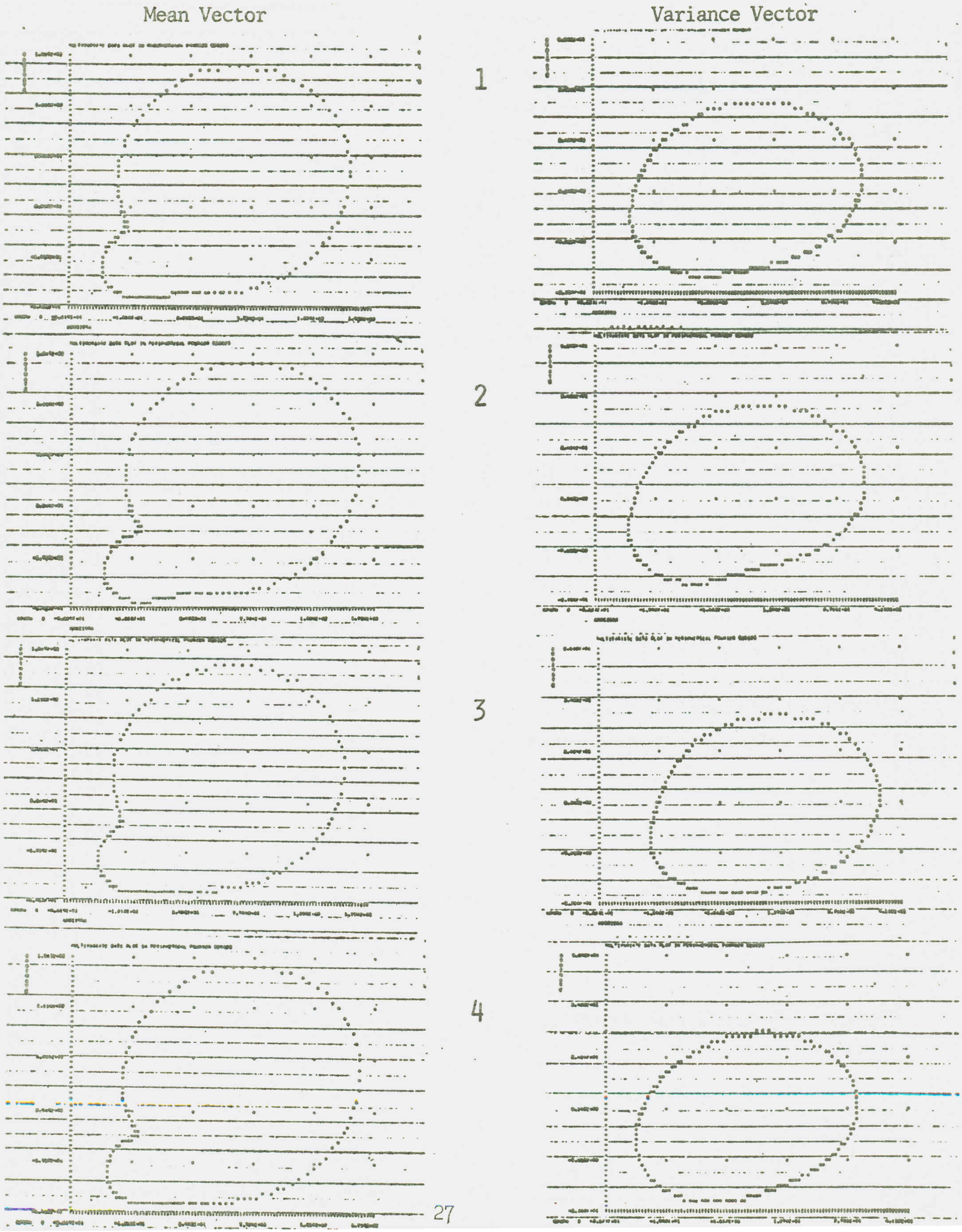
```
#N

CAT#              1       2       3       4       5       6       7       8
# POINTS        135     168     154      62      85     349     244     120
CAT#              9       A
# POINTS        121     135
TOTAL NUMBER OF POINTS=   1593


#M

MEANS:
                              CHANNELS
CAT#              1       2       3       4
   1          37.03   31.95   55.26   31.19
   2          35.27   29.17   61.05   35.90
   3          36.23   30.29   54.82   29.36
   4          36.46   30.84   61.48   32.90
   5          35.71   31.16   66.25   37.08
   6          34.93   29.99   71.15   39.74
   7          36.64   32.57   51.49   27.10
   8          36.28   30.39   57.24   31.92
   9          39.93   37.28   57.82   30.49
   A          27.73   17.08   12.19    2.99


#Z

VARIANCE-COVARIANCE MATRIX, (UPPER TRIANGULAR):
                              CHANNELS
CAT#              1       2       3       4
   1           3.03    3.06    1.29   -0.05
                       8.02    2.05   -0.84
                               4.61    1.48
                                       3.88

   2           2.99    3.56    0.04   -1.81
                       9.05   -0.81   -3.67
                               3.72    2.24
                                       4.99

   3           3.34    2.71    1.27   -0.33
                       5.55    1.51   -0.36
                               8.24    3.50
                                       3.88

   4           2.05    1.07    0.20   -0.78
                       5.02    0.57   -0.31
                               2.18    0.17
                                       3.22

   5           5.38    9.61    0.11   -2.62
                      27.31   -0.01   -5.72
                               3.43    1.04
                                       3.96

   6           2.32    3.15   -1.10   -1.31
                       9.26   -4.35   -4.04
                               7.98    3.09
                                       4.16

   7           4.17    5.23    0.64   -0.58
                      12.11    0.55   -1.44
                               6.96    3.44
                                       4.17

   8           4.56    5.11    0.86   -1.28
                       9.85   -0.65   -2.56
                               5.58    2.37
                                       4.67

   9           2.63    2.48    0.25    0.02
                       7.17    0.46    0.32
                               6.32    3.37
                                       3.80

   A           1.04    0.30    0.17    0.14
                       1.09    0.36    0.10
                               0.92   -0.08
                                       0.40
```

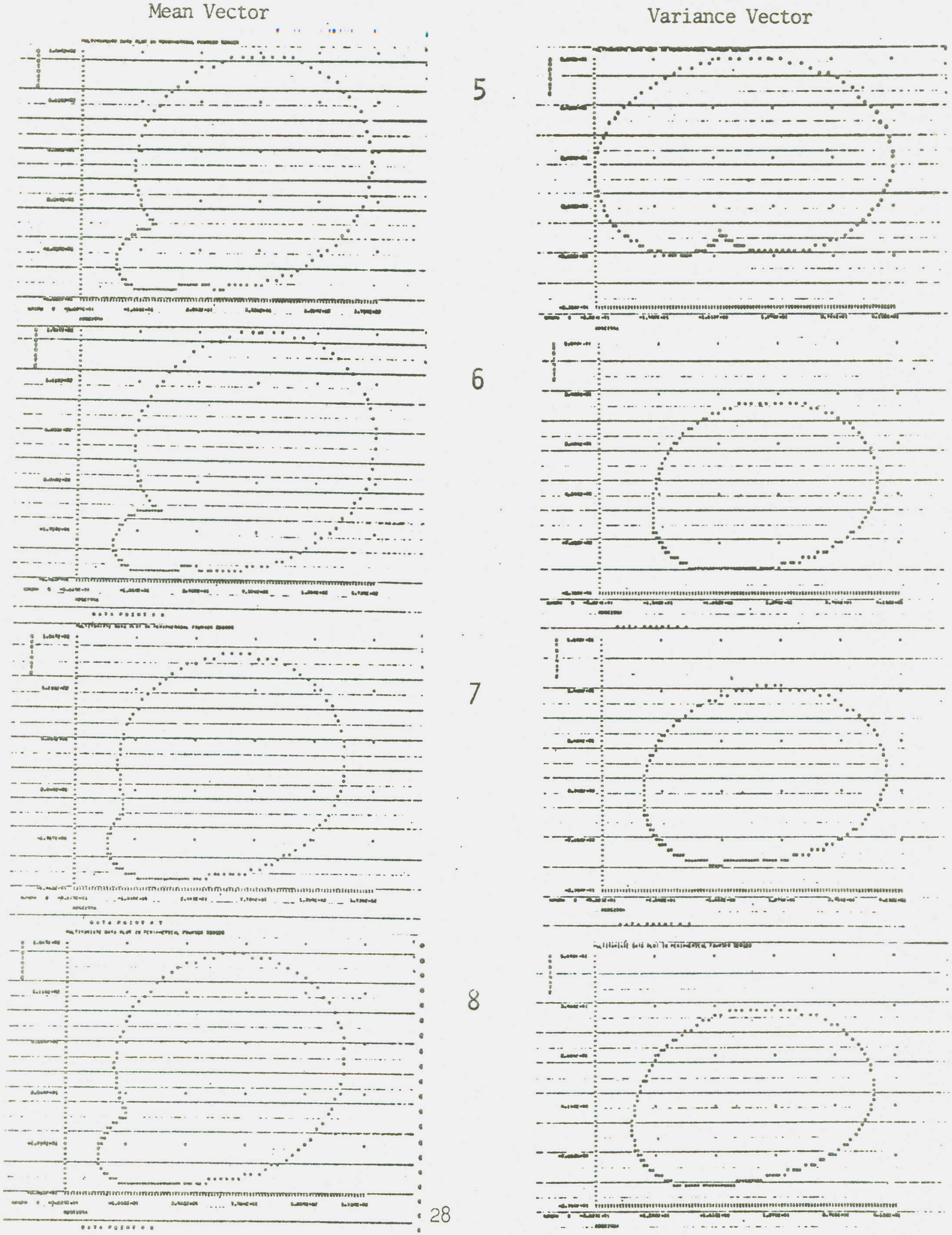
The Ten Clusters Examined in the Following Plots.

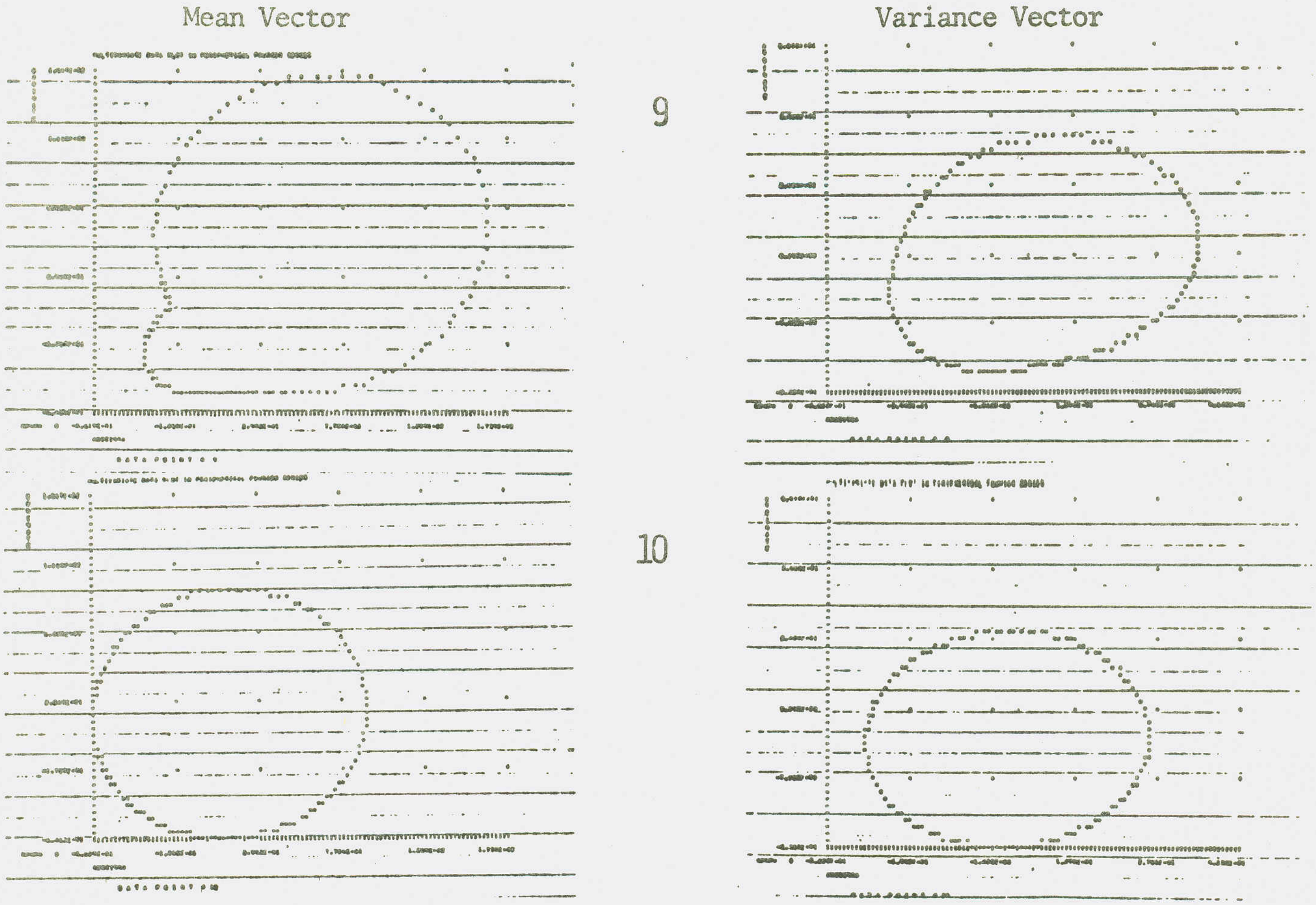
```
             CATEGORIZED TO:
             FROM:
                  :       :      :PASTURE WITH TREES:     :
                  :       :      :           STUBBLE:     :
                  :       :      :     SUDAN STUBBLE:     :
                  :       :      :     JOHNSON GRASS:     :
                  :       :      :               HAY:     :
                  :       :      :     WHEAT STUBBLE:     :
                  :       :      :              OATS:     :
                  :       :      :       OAT STUBBLE:     :
                  :       :      :              IDLE:     :
                  :   MILO:      :   COASTAL PASTURE:     :
                  :SORGHUM:COTTON:           PASTURE:TOTAL:
              MILO:       :      :                  :     :
           SORGHUM:    264:   155:               404:  823:
            COTTON:     63:   570:               181:  814:
PASTURE WITH TREES:       :      :                  :     :
           STUBBLE:       :      :                  :     :
     SUDAN STUBBLE:       :      :                  :     :
     JOHNSON GRASS:       :      :                  :     :
               HAY:       :      :                  :     :
     WHEAT STUBBLE:       :      :                  :     :
              OATS:       :      :                  :     :
       OAT STUBBLE:       :      :                  :     :
              IDLE:       :      :                  :     :
   COASTAL PASTURE:       :      :                  :     :
           PASTURE:    313:   674:              3717: 4704:
             TOTAL:    640:  1399:              4302: 6341:



PER CENT CORRECT BY COVER

                MILO
             SORGHUM    32.08%
              COTTON    70.02%
  PASTURE WITH TREES
             STUBBLE
       SUDAN STUBBLE
       JOHNSON GRASS
                 HAY
       WHEAT STUBBLE
                OATS
         OAT STUBBLE
                IDLE
     COASTAL PASTURE
             PASTURE    79.02%


OVERALL PERCENT CORRECT=    71.77%



ENTER "S" TO SEE PERCENT CORRECT BY SEGMENT  OR "Q" TO QUIT: S
```

CIRCULAR FOURIER SERIES PLOTS

(Four Channel Plots)




Mean Vector
Variance Vector
5
6
7
8



NOTES

1. Plots 1 and 2 are both sorghum plots.
2. Plots 3, 4, 5 and 6 are of cotton. We note that plot 4 is very much similar to those of sorghum.
3. Plots 7, 8 and 9 are of pasture, stubble and oats. Again, some similarities with the previous groups are evident.
4. Water is represented by plot 10. This is clearly different from the agricultural groups.

APPENDIX E

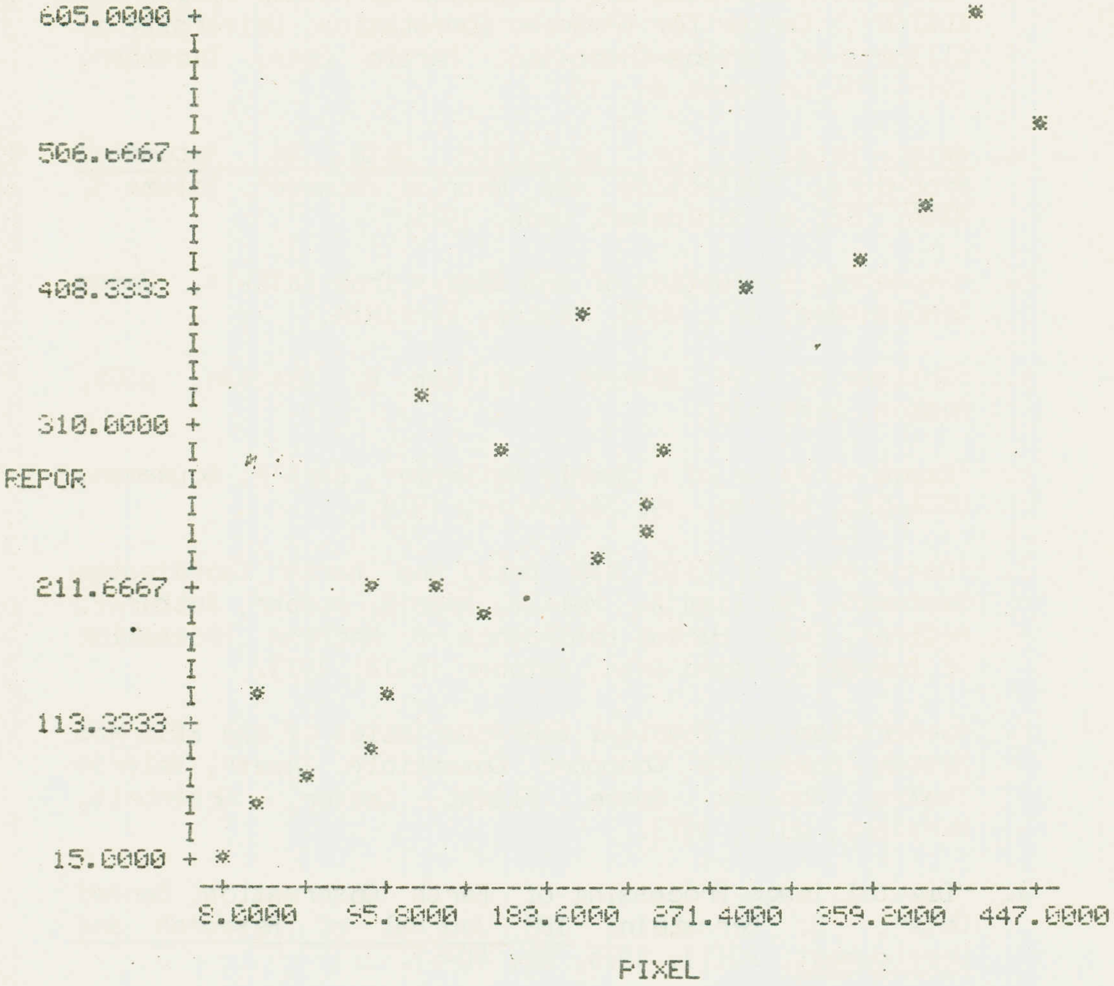
```
THIS PROGRAM PERFORMS A LINEAR REGRESSION OF VARIABLES Y ON X.

NAME OF Y-VARIABLE = REPORTED ACRES+COTTON

NAME OF X-VARIABLE = PIXELS COTTON

IS THE DATA TO BE ENTERED BY KEYBOARD? (Y OR N) Y
```

```
NAME OF X-VARIABLE = PIXELS COTTON

IS THE DATA TO BE ENTERED BY KEYBOARD? (Y OR N) Y
```

```
WOULD YOU LIKE TO SEE A PLOT OF Y VERSUS X? (Y OR N) Y
   350.0000 +                                                            *
            I
            I
            I                                                    *
            I
   293.3333 +
            I
            I
            I
            I
   236.6667 +
            I
            I
            I
            I
   180.0000 +
 REPOR      I
            I
            I
            I
            I
   123.3333 +
            I
            I        *
            I
            I
    66.6667 +     *
            I  *
            I
            I  *
            I *
    10.0000 + *
             -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
             1.0000    57.8000   114.6000   171.4000   228.2000   285.0000

                                  PIXEL


THE REGRESSION EQUATION IS:

     REPORTED ACRES+COTTO =  0.35350E+02 + ( 0.11872E+01 * PIXELS COTTON        )


        ANALYSIS OF VARIANCE TABLE FOR VARIABLE REPORTED ACRES+COTTO
SOURCE OF   DEGREES OF      SUM  OF                MEAN
VARIATION    FREEDOM        SQUARES               SQUARES
                                                                    F-VALUE    P-VALUE
-----------------------------------------------------------------------------------------
REGRESSION      1         123627.182000
```

```
DO YOU WISH TO TEST ANOTHER HYPOTHESIS? (Y OR N) N

DO YOU WISH TO ANALYZE ANOTHER DATA SET? (Y OR N) Y









NAME OF Y-VARIABLE = REPORTED ACRES←SORGHUM

NAME OF X-VARIABLE = PIXELS←SORGHUM

IS THE DATA TO BE ENTERED BY KEYBOARD? (Y OR N) Y


WOULD YOU LIKE TO SEE A PLOT OF Y VERSUS X? (Y OR N) Y
        196.8000 +                                              *
                 I
                 I
                 I
                 I
        164.8333 +
                 I                                              *
                 I
                 I
                 I
        132.8667 +
                 I
                 I                                                   *
                 I
                 I
        100.9000 +
                 I
     REPOR       I
                 I        *
                 I        *
                 I
         68.9333 +
                 I
                 I  *
                 I                   *
                 I
         36.9667 + *  **
                 I
                 I  *
                 I
                 I
          5.0000 + *
                  -+----+----+----+----+----+----+----+----+----+----+-
                 0.0000   14.6000   29.2000   43.8000   58.4000   73.0000

                                           PIXEL



THE REGRESSION EQUATION IS:

     REPORTED ACRES←SORGH =  0.30687E+02 + ( 0.18502E+01 * PIXELS←SORGHUM      )


         ANALYSIS OF VARIANCE TABLE FOR VARIABLE REPORTED ACRES←SORGH

SOURCE OF  DEGREES OF    SUM  OF           MEAN
VARIATION   FREEDOM      SQUARES      SQUARES               F-VALUE  P-VALUE
----------------------------------------------------------------------------
REGRESSION     1       29585.872800     29585.872800        35.6194   0.0002
RESIDUAL       9        7475.508730       830.612083
----------------------------------------------------------------------------

R =  0.893473         RSQUARE =  .798294
```

```
NAME OF Y-VARIABLE = REPORTED ACRES←PASTURE

NAME OF X-VARIABLE = PIXELS←PASTURE

IS THE DATA TO BE ENTERED BY KEYBOARD? (Y OR N) Y

WOULD YOU LIKE TO SEE A PLOT OF Y VERSUS X? (Y OR N) Y
```



```
THE REGRESSION EQUATION IS:
  REPORTED ACRES←PASTU =  0.56174E+02 + ( 0.11380E+01 * PIXELS←PASTURE


        ANALYSIS OF VARIANCE TABLE FOR VARIABLE REPORTED ACRES←PASTU

SOURCE OF   DEGREES OF      SUM  OF            MEAN
VARIATION    FREEDOM        SQUARES           SQUARES            F-VALUE   P-VALUE
-----------------------------------------------------------------------------------
REGRESSION       1       446341.742000     446341.742000         112.4499   0.0000
RESIDUAL        18        71446.508800       3969.250490
-----------------------------------------------------------------------------------

R =  0.928448            RSQUARE =  .862016
```

REFERENCES

1. Crop Identification and Acreage Measurement Utilizing LANDSAT Imagery, USDA-SRS, Donald H. Von Steen and William H. Wigton, March, 1976.

2. "Selection, Sampling and Tabulation of Masked Fields in EDITOR", Center for Advanced Computation, University of Illinois at Urbana-Champaign, Martin Ozga, December, 1976, CAC Document No. 79.

3. Water Mapping for Satellite Data: An Automated Procedure, "Detection and Mapping Package", Volume 3, NASA, JSC, Ed Schlosser, June, 1976.

4. "Geometric Evaluation of MSS Images from ERTS-1", James Schoonmaker, Jr., USGS, Reston, Virginia.

5. "Griding of ERTS Images", William H. Chapman, USGS, Reston, Virginia.

6. "Expected Value of a Sample Estimate", Earl E. Houseman, USDA-SRS, SRS No. 19, September, 1974.

7. "Correlation of ERTS MSS Data and Earth Coordinates Systems", William A. Malila, Ross H. Hieber, Authur P. McCleer, ERIM, Purdue Conference on Machine Processing of Remotely Sensed Data, October 16-18, 1973.

8. "Generation and Physical Characteristics of the ERTS MSS System Connected Computer Compatible Tapes", Valerie Thomas, Goddard Space Flight Center, Greenbelt, Maryland, July, 1973.

9. "Digital Image Processing of Earth Observation Sensed Data", R. Bernstein, IBM Journal of Research and Development, 20(1), 1976, pp. 40-57.

10. LARSYS USER GUIDE, Volumes 1-3, LARS, Purdue Univesity, June, 1973.

11. "Procedures for Boundary Location in LANDSAT MSS Digital Data", Paul W. Cook, USDA, Statistical Reporting Service, In Progress, 1977.