NAS9-14970
T-1314/4
MA-129TA

Final Technical Report

NASA Contract NAS9-14970

June 1, 1976 - May 31, 1977

D. A. Landgrebe, Purdue University
Principal Investigator

J. D. Erickson, NASA/JSC
Technical Monitor

Volume II of III

Submitted by

The Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana

1977

NAS9-14970
T-1314/4
MA-129TA

Final Technical Report

NASA Contract NAS9-14970

June 1, 1976 - May 31, 1977

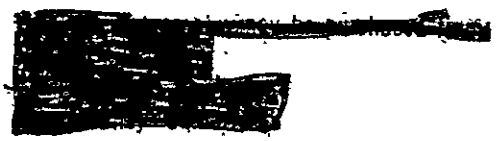
(E77-10186) [APPLICATIONS OF REMOTE SENSING, VOLUME 2] Final Technical Report, 1 Jun. 1976 - 31 May 1977 (Purdue Univ.) 93 p HC A05/MF A01 CSCL 05B G3/43

N77-30550

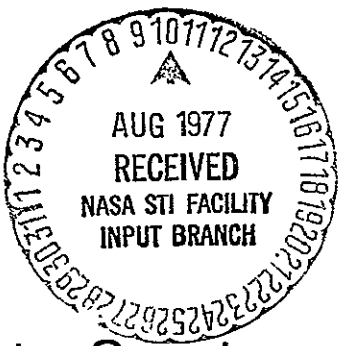
Unclas 00186

D. A. Landgrebe, Purdue University
Principal Investigator

J. D. Erickson, NASA/JSC
Technical Monitor



Volume II of III



Submitted by

The Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana

1977

Table of Contents

RESEARCH TASKS

2.2 Stratification of Scene Characteristics

Task 2.2, Stratification of Scene Characteristics, contains three subtasks: 2.2a. Stratification by Machine Clustering, 2.2b. Digitization and Registration of Ancillary Data, and 2.2c. Crop Inventory Using Full-Frame Classification.

2.2a Stratification by Machine Clustering

The work on the stratification by machine clustering task has progressed in two stages in the past year. First, from June 1976 to January 1977, previous work at LARS was evaluated and a new approach to the scene stratification problem was formulated. Secondly, from January 1977 to May 1977, the new approach was implemented and investigated on a trial basis.

Evaluation of Previous Work

Evaluations of the dynamic stratifications produced in FY76 have suggested that the machine clustering methods proposed in FY76 cannot reliably distinguish when two segments are from different strata. In the stratification experiments, when the partitioning procedure placed two segments in different strata, this assignment was generally correct; however, when segments were assigned to the same stratum by the partitioning procedures, they did not always seem to come from the same stratum. Evidence of this behavior can be seen in Table 1.

The classification results in Table 1 show that the stratification technique was successful in identifying segments which are different. In no case was high classification performance achieved when using training statistics from segments outside the stratum, as shown by the first row and column of the table. For segments identified as members of the same stratum, similar high (approximately 90 percent correct) classification performances were obtained for both local and non-local classifications of several combinations of segments; for example, the non-local recognitions involving Barton-2 and Rice. This indicates that

Table 1. Classification Performances (Wheat vs. Other) for Segments Within and Outside of Strata Determined by Clustering

Strata No.	Source of Training Statistics	Areas Classified*					
		Barton-1	Barton-2	Rice	Ellsworth[1]	Ellis	Stafford
		Overall Percent Correct					
1	Barton-1	83.7	42.9	15.1	69.4	54.1	61.5
2	Barton-2	27.1	96.0	93.8	90.0	56.2	52.5
2	Rice	34.1	92.0	93.4	85.7	47.4	69.1
2	Ellis	63.4	43.4	26.4	60.4	64.8	51.4
2	Stafford	58.2	55.4	42.0	59.9	61.7	89.9

* Landsat scenes 1689-16392 and 1689-16385 acquired June 12, 1974 over Central Kansas.

[1] Ellsworth was not used as a source of training statistics because only wheat field coordinates were available.

these segments are from the same stratum. But, the non-local classification performance was lower than the local classification performance in several other instances, such as the non-local recognitions between Rice and Ellis, indicating that these segments may be from different strata. This would mean that the clustering procedure is grouping the segments into groups of strata which spectrally are too broadly defined.

Additional partitions and classifications of different groups of segments were examined in the evaluation carried out in FY77. (See Tables 2 and 3.) The overall level of classification accuracy is lower in these tables, but conclusions similar to those cited above can be drawn.

A stratification of SRS segments acquired May 24-27, 1974, was evaluated in terms of proportion estimation (Table 4). These results were obtained from the PROCAMS system developed by ERIM.[1] Although Morton and Stevens (S) were placed in the same stratum, training on the Finney Intensive Test Site (from a different stratum) gave an estimate of percent wheat for Stevens (S) which was closer to the actual figure. Similarly, training on Morton ITS (stratum 1) gave better estimates of the percent wheat in Haskell and Kearny (stratum 2) than did training from Finney ITS, which had been placed in stratum 2 by the machine clustering method.

We then concluded that the clustering methods proposed in FY76 did not successfully partition the scene into regions sufficiently homogeneous to assure adequate non-local classification performance. While these results did not rule out the idea of using machine clustering methods for dynamic stratification, they suggested that more knowledge of the nature of stratification is needed if an effective machine method is to

Table 2. Classification Performances (Wheat vs. Other) for Segments Within and Outside of Strata Determined by Clustering

Strata No.	Source of Training Statistics	Areas Classified*					
		Barton	Ellis	Reno	Rush	Russell	Stafford
		Overall Percent Correct					
1	Barton	56.5	37.4	57.5	47.4	69.0	17.5
1	Ellis	47.1	73.1	60.8	76.6	78.8	100.0
1	Reno	51.8	62.6	78.0	61.5	74.0	100.0
1	Rush	33.8	62.6	34.4	70.3	43.5	99.5
1	Russell	56.0	62.6	59.8	51.6	45.3	98.5
2	Stafford	53.7	53.3	39.0	19.8	21.8	47.9

* Landsat scene 1455-16432 acquired October 21, 1973 over central Kansas.

Table 3. Classification Performances (Wheat vs. Other) for Segments Within and Outside of Strata Determined by Clustering

Strata No.	Source of Training Statistics	Areas Classified*				
		Grant	Kearny	Haskell	Stevens S	Stevens N
		Overall Percent Correct				
1	Grant	75.9	63.2	42.6	44.2	42.6
1	Kearny	59.9	70.3	39.4	64.1	51.6
2	Haskell	65.3	48.6	57.4	43.0	49.0
2	Stevens S	41.6	59.6	47.7	63.8	49.5
3	Stevens N	46.9	49.0	44.9	27.3	43.7

* Landsat scene 1691-16501 acquired June 14, 1974 over Southwest Kansas

Table 4. Estimation of Percent Wheat for Segments Within and Outside of Strata Determined by Clustering

Site	Strata No.	Percentage Wheat Training from MORTON	Percentage Wheat Training from FINNEY	Percentage Wheat Actual	Difference from Actual Training from MORTON	Difference from Actual Training from FINNEY
Morton ITS	1	43.1	---	40.36	+2.74	---
Stevens (S)	1	33.2	27.2	20.87	+12.33	+6.33
Finney ITS	2	---	22.5	26.2	---	-3.7
Grant	2	48.6	30.1	32.08	16.52	-1.98
Haskell	2	40.2	9.3	29.69	10.51	-20.39
Kearny	2	32.5	29.8	33.93	-1.43	-4.13

*Data acquired May 24-27, 1974.

*Proportion estimates provided February 2, 1976, by ERIM.

be developed.

Thus a more intensive investigation into the fundamental nature of apparent strata in Landsat imagery was deemed essential to the development of the quantitive measures needed for an effective machine partitioning method. It has been observed that an experienced data analyst develops an ability to visually stratify a scene. The Remote Sensing Research Program at the University of California, Berkeley, has related soil, land use, and long-term precipitation patterns to the visually apparent strata in Landsat imagery and has sought to develop static stratifications based on the physical and agronomic patterns.[2] However, it had not been determined whether the ability to visually stratify a scene is sufficiently precise for achieving the necessary classification accuracy. Nor was it known what scene attributes the analyst utilizes, although it was suspected that they are both spectral and spatial in nature. In addition, it was not clear how these attributes would be quantified and applied in dynamic stratification. An hypothesis was suggested that the strata an analyst determines from imagery are related to classification accuracy. This hypothesis became the basis for the new approach to the scene stratification problem taken at LARS. Essentially the quantitative properties of analyst defined strata were investigated to determine if a quantitative measure could be found to define strata.

Formulation of a New Approach

In seeking support for the above hypothesis, a new approach was formulated. In this approach, partitions defined by three analysts on

two frames of Landsat data covering a portion of the '75-'76 LACIE segments, the static partitions developed by UCB[2], and the partitions used by segment selection in LACIE Phase III were examined. Several image enhancement algorithms were implemented and applied to the two Landsat scenes to see if the enhanced data aided the analysts in defining partitions. Attributes of the partitions were measured, including spectral response and class structure. The partitions were evaluated with respect to these attributes. Due to delays in the receipt of imagery and in obtaining enhanced imagery, less evaluation of the partitions was possible than was projected in the implementation plan. Ideally, classification performance obtained from several training procedures would be used to evaluate the partitions. However, this was not possible due to time constraints.

Implementation of New Approach

Data Selection. The two scenes 2416-16362, acquired March 13, 1976, and 5428-16053, acquired June 20, 1976, were selected according to the following criteria:

1. The full frame Landsat digital data was in-house at LARS in January 1977.
2. The number of LACIE segments in the area was maximized.
3. The date of data acquisition was suitable with respect to the crop calendar.
4. The static partitions developed by UCB and by the USDA for LACIE Phase III were available for the area covered by the scenes.
5. The full frames did not have clouds or large amounts of bad data.

The location of the two frames over Kansas is shown in Figure 1, with scene 5428-16053 being the northernmost frame. The locations of the LACIE segments within each frame are shown in Figures 2 and 3.

Enhancement of Full-Frame Images. Comparison of the "standard" color composite produced with the digital display at LARS, the Landsat false color imagery, and LACIE PFC products indicated major differences among these products. Duplication of JSC Production Film Converter (PFC) false color imagery has been performed with very good success.

In order to generate imagery similar to the PFC imagery, equal range levels are selected covering the data range of $\pm$ 3 standard deviations from the mean in each of the wavelands 4, 5, and 7. These levels for each channel were displayed on the CRT and photographed through blue, green and red filters, respectively. The time of exposure is 12/15 second through the blue and red filters and 4/15 second through the green filter at an f-stop of 8. Although such transparencies have more blue than the PFC product, distinctive soil patterns can be seen and there appears to be good crop condition discriminability.

Three enhancement techniques were applied to the full-frame data and used in the stratification work. These three transformations were: principal components, ratio/magnitude, and tasselled cap.

The principal components transformation was determined for each frame from the covariance matrix of a 4% sample of the frame. The equations for the principal components transformation for each frame are given in Table 5. The first principal component is oriented in the direction of greatest variation in the data; the principal components are orthogonal to

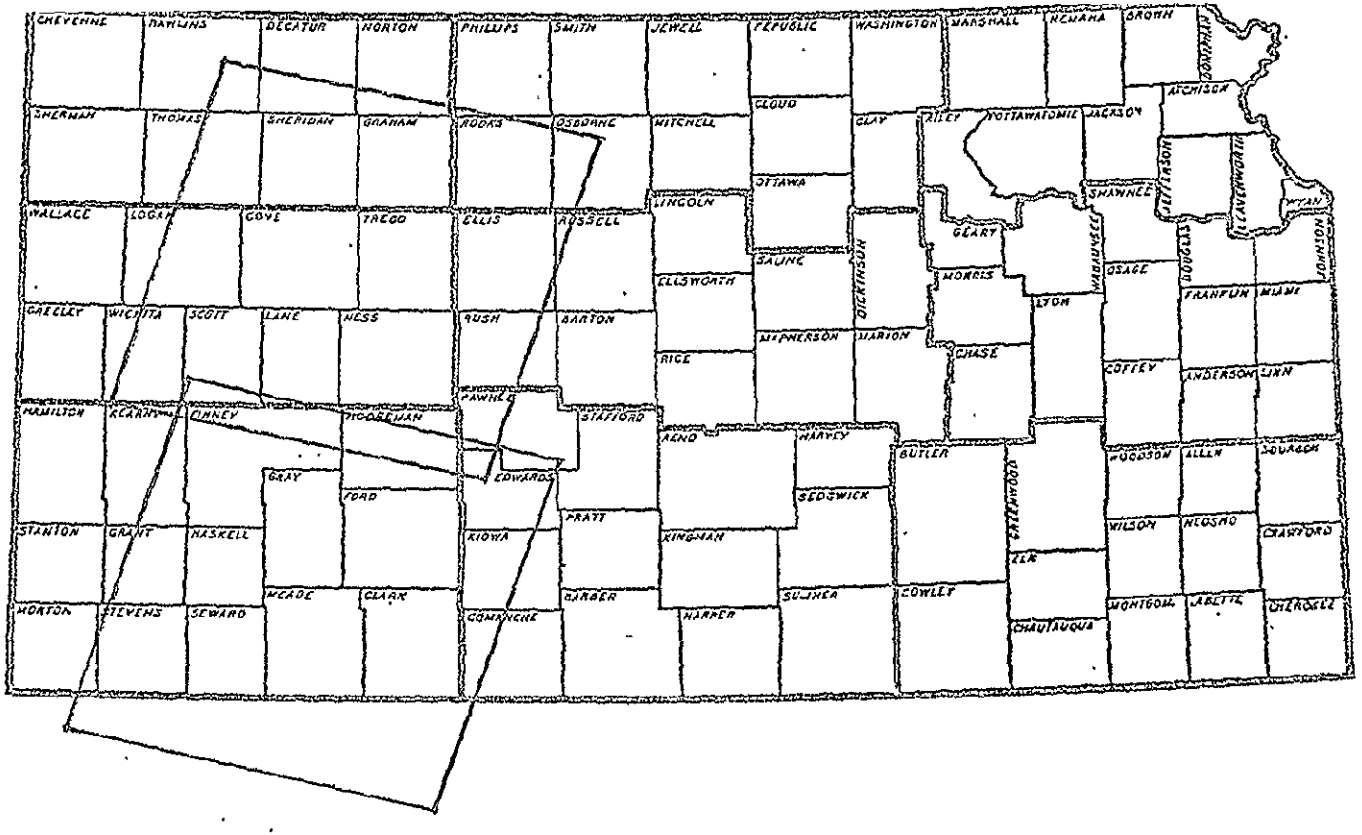


Figure 1. Location of Landsat scenes 2416-16362, March 13, 1976, and 5428-16053, June 20, 1976.

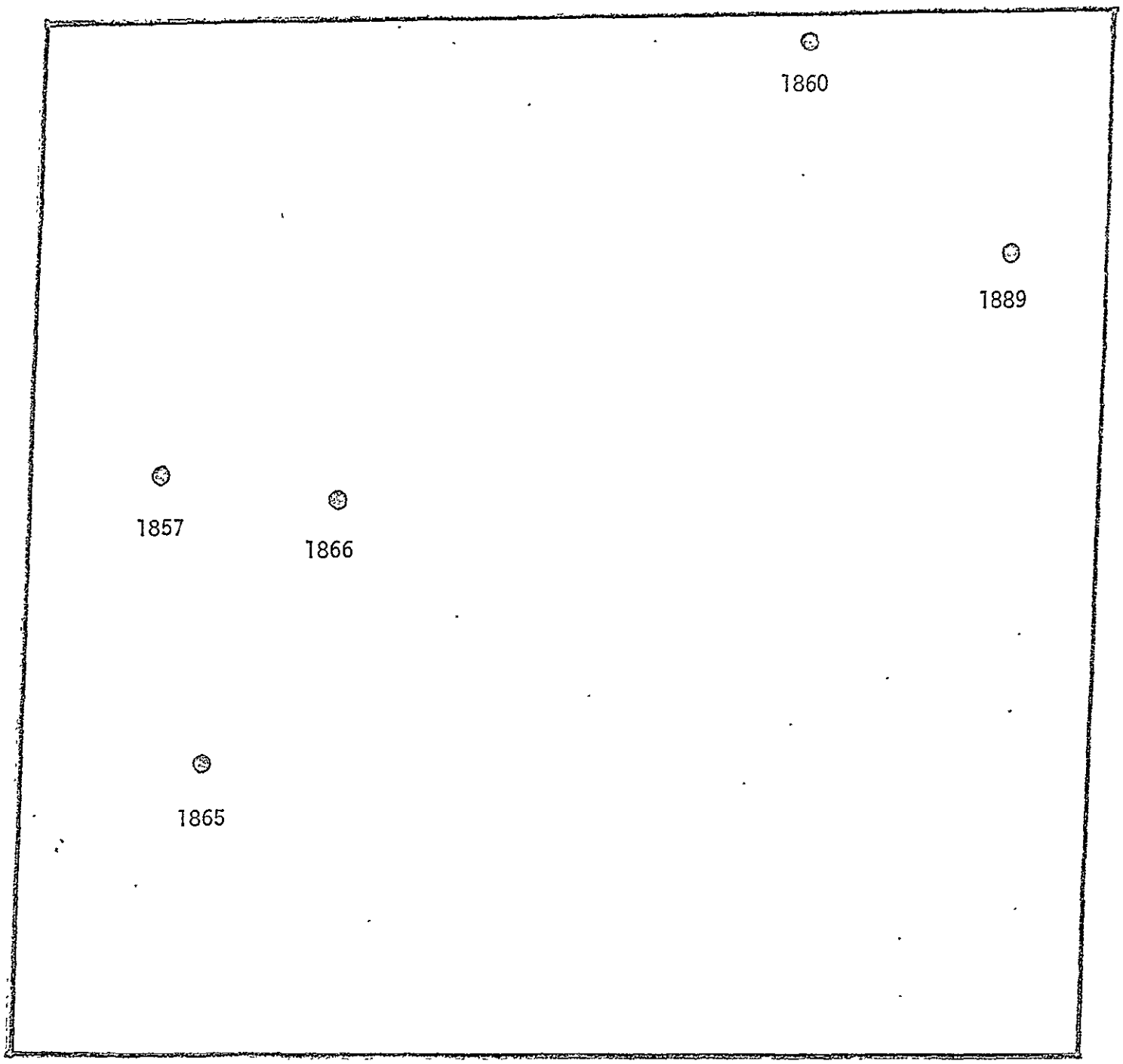


Figure 2. Centers of LACIE segments in Landsat scene 2416-16362, March 13, 1976.

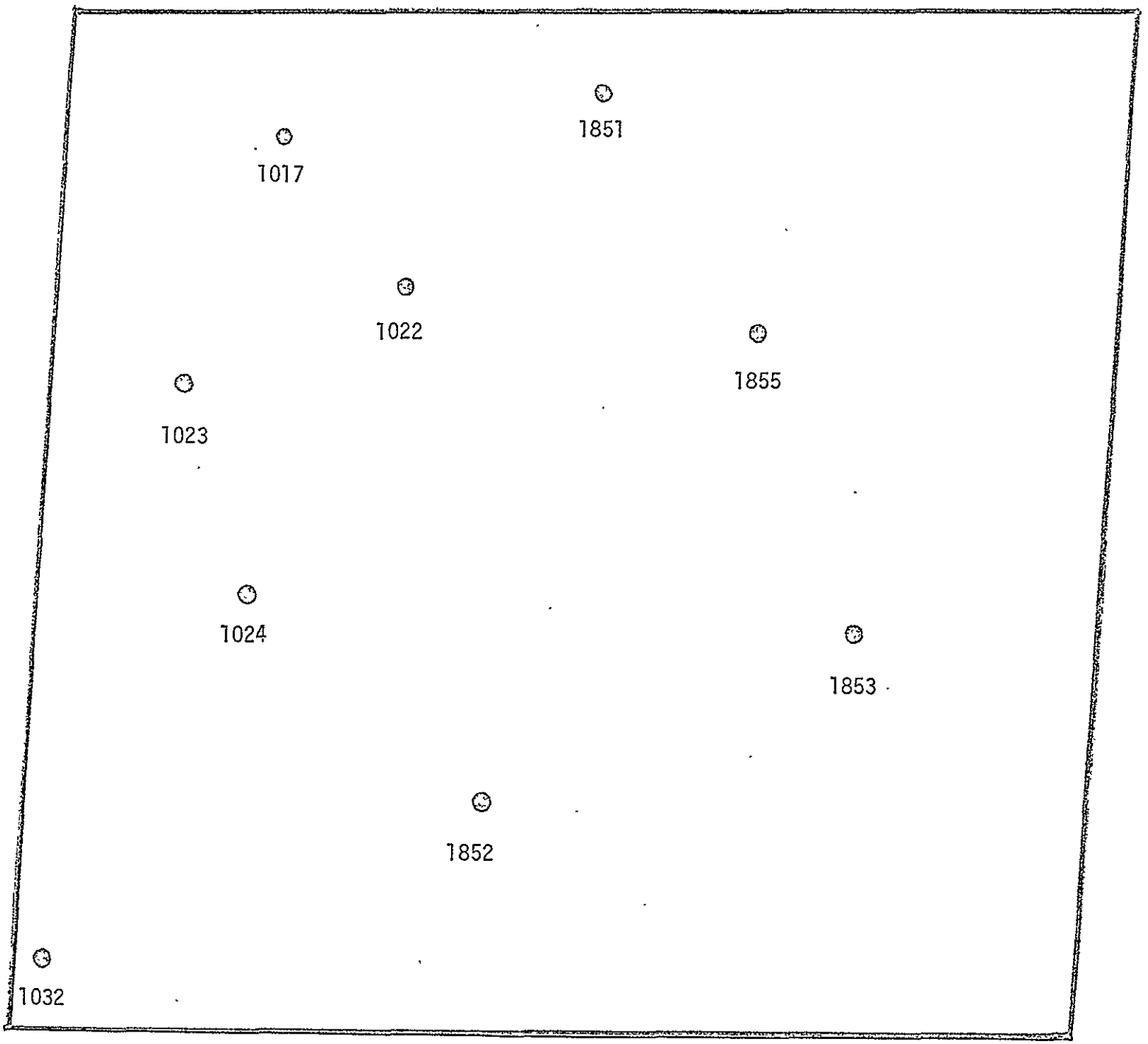


Figure 3. Centers of LACIE segments in Landsat scene 5428-16053, June 20, 1976.

Table 5. Principal components transformation for Landsat scenes 2416-16362 and 5428-16053.

2416-16362, March 13, 1976

$$c_1 = 0.3022x_1 + 0.5901x_2 + 0.6826x_3 + 0.3074x_4 + 64.0$$
$$c_2 = 0.3735x_1 + 0.6340x_2 - 0.5159x_3 - 0.4386x_4 + 64.0$$
$$c_3 = 0.8602x_1 - 0.4955x_2 + 0.0858x_3 - 0.0846x_4 + 64.0$$
$$c_4 = 0.1711x_1 + 0.0651x_2 - 0.5104x_3 + 0.8402x_4 + 64.0$$

5428-16053, June 20, 1976

$$c_1 = 0.4919x_1 + 0.7505x_2 + 0.4243x_3 + 0.1213x_4 + 72.0$$
$$c_2 = -0.1610x_1 - 0.3966x_2 + 0.7397x_3 + 0.5193x_4 + 72.0$$
$$c_3 = 0.7895x_1 - 0.5259x_2 + 0.1005x_3 - 0.3000x_4 + 72.0$$
$$c_4 = 0.3297x_1 - 0.0542x_2 - 0.5126x_3 + 0.7910x_4 + 72.0$$

x_1, x_2, x_3, x_4 are the elements of the original Landsat data, corresponding to wavelands 4, 5, 6, and 7.

each other. The constant vector is chosen so that resulting data values are positive.

The equations for the ratio/magnitude transformation are given in Table 6. The same equations were used for both scene 2416-16362 and 5428-16053. This is a nonlinear transformation which converts sectors in the two dimensional space spanned by the sum of the visible channels and the weighted sum of the infrared channels into rectangles in the ratio/magnitude space. The ratio and magnitude were chosen because they can be related to the spectral response of soil and vegetative cover.

The Tasselled Cap transformation[3] was developed at ERIM and is based on a model of the spectral response of vegetative canopies through their growth cycle. The linear transformation "isolates green development, yellow development, and soil brightness and allows the reduction of the dimension of the feature space."[3] The equations used to transform the two Landsat scenes are given in Table 7.

Partitioning. The Kansas portion of the partitioning of the United States, developed by USDA personnel at JSC and to be used in the sample allocation for Phase III of LACIE, was delivered to the staff of LARS during the LACIE Project Review, January 26-28, 1977. The portions of this partitioning covering the two Landsat scenes chosen for investigation in this study are shown in Figures 4 and 5.

The static stratification of Kansas developed by UCB in FY76 was available in "Application of Photointerpretative Techniques to Wheat Identification, Signature Extension, and Sampling Strategy," by C.M. Hay and R.W. Thomas, Space Sciences Laboratory Series 17, Issue 33. The portions of this stratification covering the two Landsat frames chosen for this study are shown in Figures 6 and 7.

Table 6. Ratio/Magnitude transformation.

$$c_1 = \frac{16(x_1 + x_2)}{(x_3 + 2x_4 + 0.01)}$$

$$c_2 = x_1 + x_2 + x_3 + 2x_4$$

x_1, x_2, x_3, x_4 are the elements of the original Landsat data vector.

Table 7. Tasselled Cap transformation for Landsat scenes 2416-16362 and 5428-16053.

2416-16362, March 13, 1976[4]

$$c_1 = 0.3323x_1 + 0.6032x_2 + 0.6758x_3 + 0.2628x_4 + 75.0$$
$$c_2 = -0.2832x_1 - 0.6601x_2 + 0.5774x_3 + 0.3883x_4 + 75.0$$
$$c_3 = -0.8995x_1 + 0.4283x_2 + 0.0759x_3 - 0.4080x_4 + 75.0$$
$$c_4 = -0.0159x_1 + 0.1307x_2 - 0.4519x_3 + 0.8823x_4 + 75.0$$

5428-16053, June 20, 1976[3]

$$c_1 = 0.433x_1 + 0.632x_2 + 0.586x_3 + 0.264x_4 + 32.0$$
$$c_2 = -0.290x_1 - 0.562x_2 + 0.600x_3 + 0.491x_4 + 32.0$$
$$c_3 = -0.829x_1 + 0.522x_2 - 0.039x_3 + 0.194x_4 + 32.0$$
$$c_4 = 0.223x_1 + 0.012x_2 - 0.543x_3 + 0.810x_4 + 32.0$$

x_1, x_2, x_3, x_4 are the elements of the original Landsat data vector.

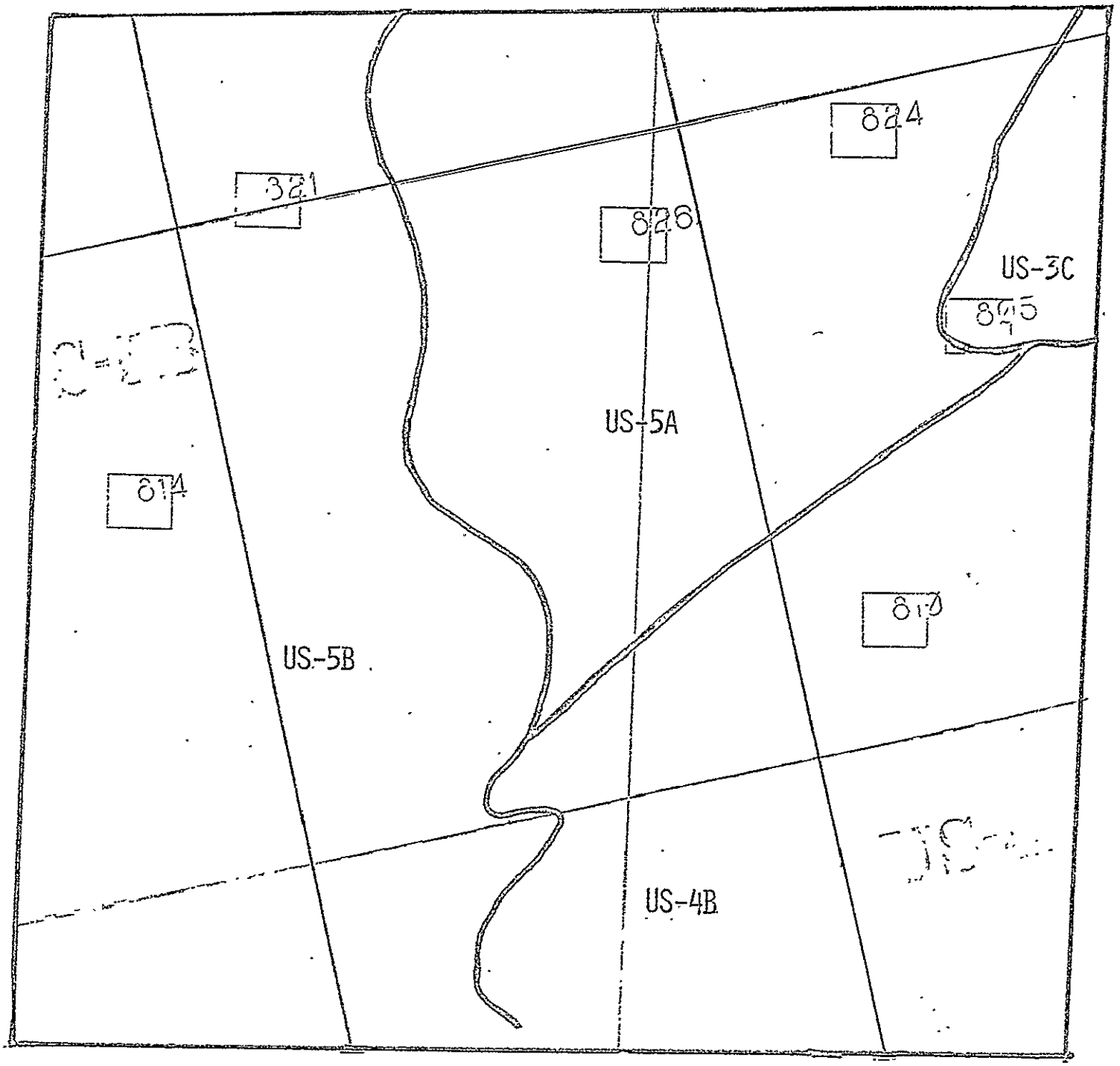


Figure 4. Partitioning of U.S. Great Plains developed for LACIE Phase III segment selection shown for area covered by Landsat scene 2416-16362, March 13, 1976.

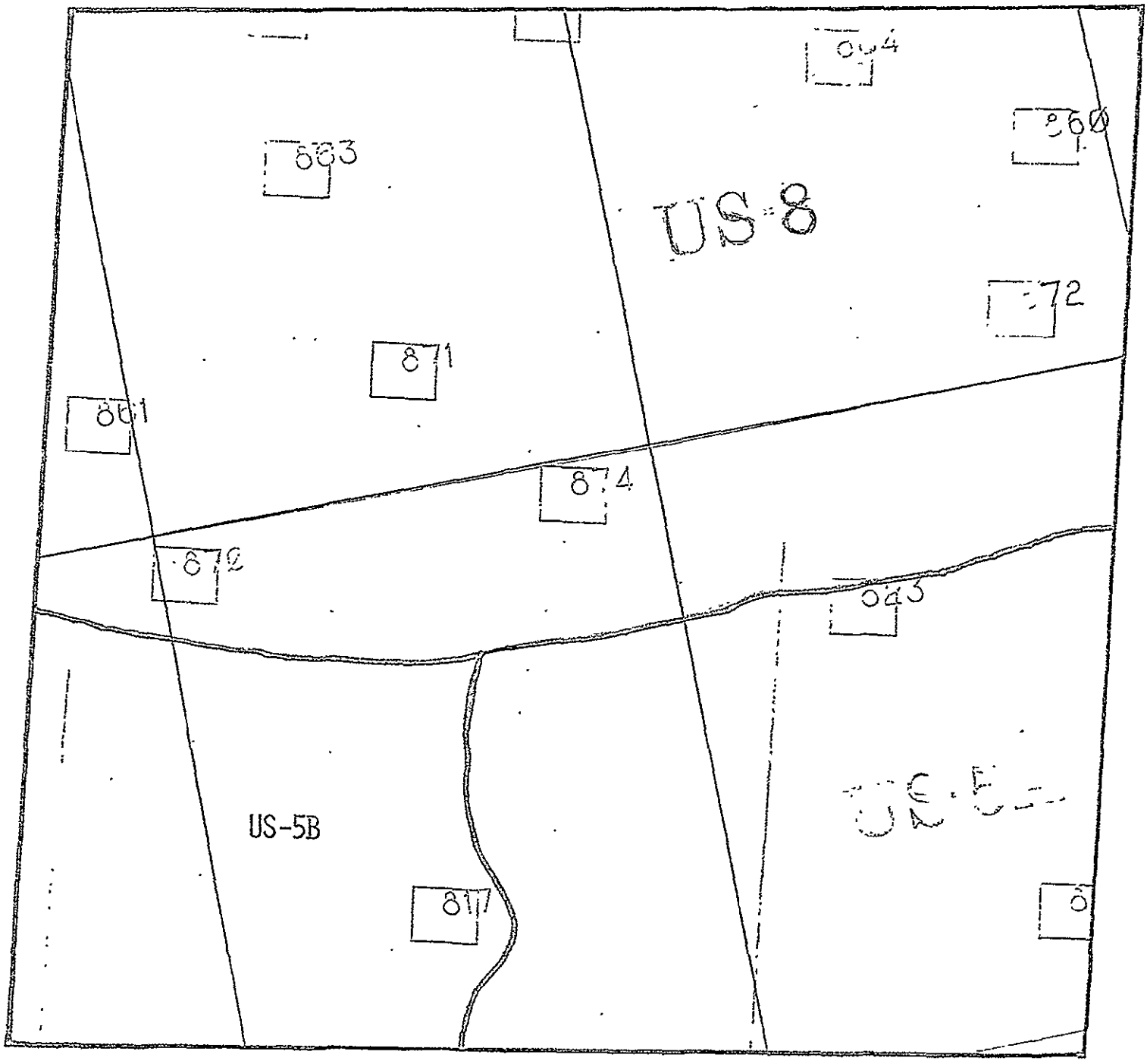


Figure 5. Partitioning of U.S. Great Plains developed for LACIE Phase III segment selection shown for area covered by Landsat scene 5428-16053, June 20, 1976.

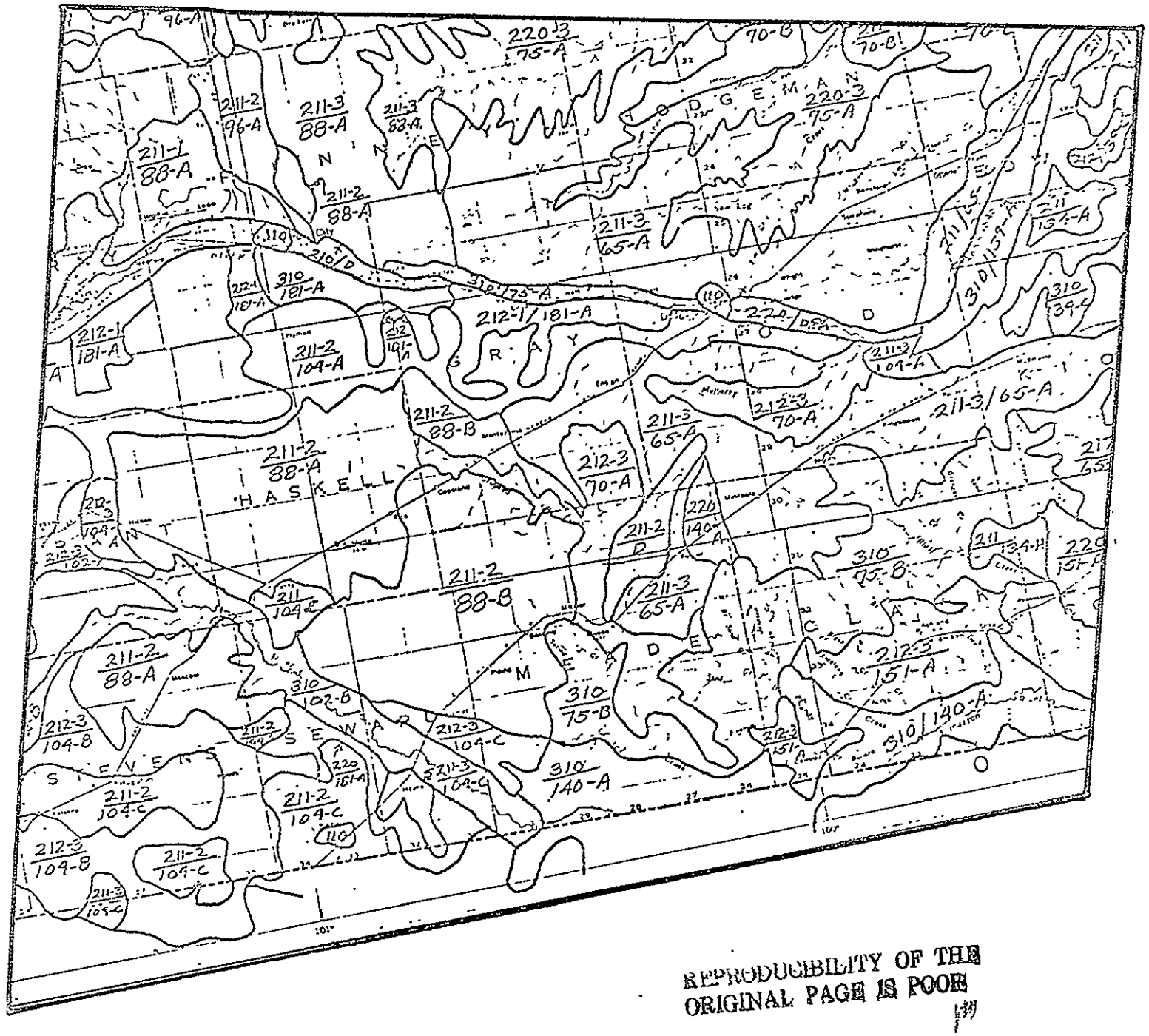


Figure 6. Static partitioning of Kansas developed by UCB[2] shown for area covered by Landsat scene 2416-16362, March 13, 1976.

Figure 7. Static partitioning of Kansas developed by UCB[2] shown for area covered by Landsat scene 5428-16053, June 20, 1976.

The partitioning procedure at LARS involved two major parts: (1) defining the stratification from an original Landsat image or an enhanced image and (2) obtaining the boundaries in digitized form. The first step in defining the partitions was to obtain an eight by ten inch print of a full-frame image taken on the digital display at LARS following the procedure given in the section entitled "Enhancement of Full-frame Images." The analyst covered the print with an acetate overlay and drew the stratification on it. The analysts used three general types of information in partitioning the scene: the overall color of a region, the variability of color in the region, and the spatial homogeneity or texture of colors in the region.

After a scene was partitioned, the strata boundaries had to be converted to digital form so that they could be used to measure attributes of the strata. At this point some unforeseen problems developed. The original intention was to digitize the boundaries of each stratum from the acetate overlay and then correct the change in aspect and the distortion introduced by the digital display. However, the distortion was not of the form originally thought and so was not correctable by the algorithms presently available in-house.

Three alternative procedures were proposed. The first suggestion was that mosaics be formed from partial images from the undistorted portion of the display screen. This would have required between 70 and 80 images and was prohibited by time and monetary constraints. The second alternative proposed was to input the boundaries through a version of the IMAGEDISPLAY processor in LARSYS. This was tried but required a great deal of personnel time. The third alternative was to transfer the partitions to a computer

output with a pixel size of 0.035 by 0.025 inches for a scale of 1:1,000,000. This printout then matched the uncorrected Landsat data and the boundaries were easily digitized. This method worked well, but the delay caused by the need to try alternatives seriously hampered the stratification task as only a few of the original 24 stratifications could be digitized.

Two examples of stratifications developed by analysts at LARS from enhanced images are shown in Figures 8 and 9. Figure 8 shows a partitioning defined on scene 2416-16362, March 13, 1976, with the aid of a Tasselled Cap transformed image while Figure 9 shows a stratification developed on the second scene, 5428-16053, from a principal components enhancement.

Measurement of Attributes. For each stratum in the LACIE Phase III partition and the UCB developed static stratification of Kansas, the overall spectral response was measured by calculating a mean vector and covariance matrix. The overall spectral response of each stratum in those stratifications defined by analysts at LARS for which digitized boundaries had been obtained was also measured. Due to time constraints, the texture and class structure of the strata were not examined.

The '75-'76 LACIE segments which are shown in Figures 2 and 3 provided most of the information on which an evaluation of the stratifications was based. The overall spectral response of each segment and its class structure was calculated. The class structure of a segment was established by clustering four 29 line by 49 column areas, one from each quadrant of the segment. By taking one area from each quadrant, the statistics could later be broken up to correspond to portions of the segment.

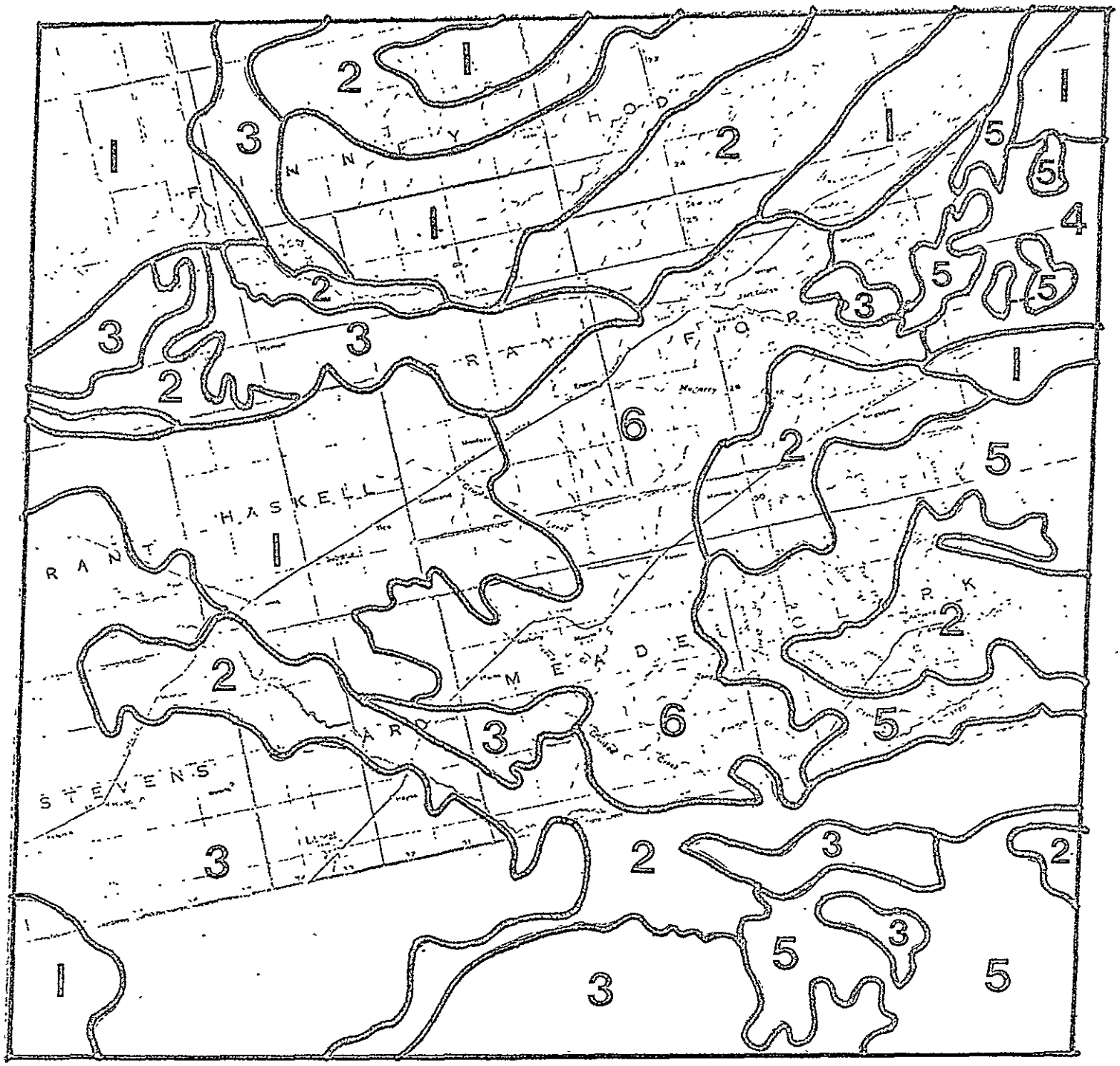


Figure 8. Stratification of Landsat scene 2416-16362, March 13, 1976, based on imagery enhanced by Tasselled Cap transformation.

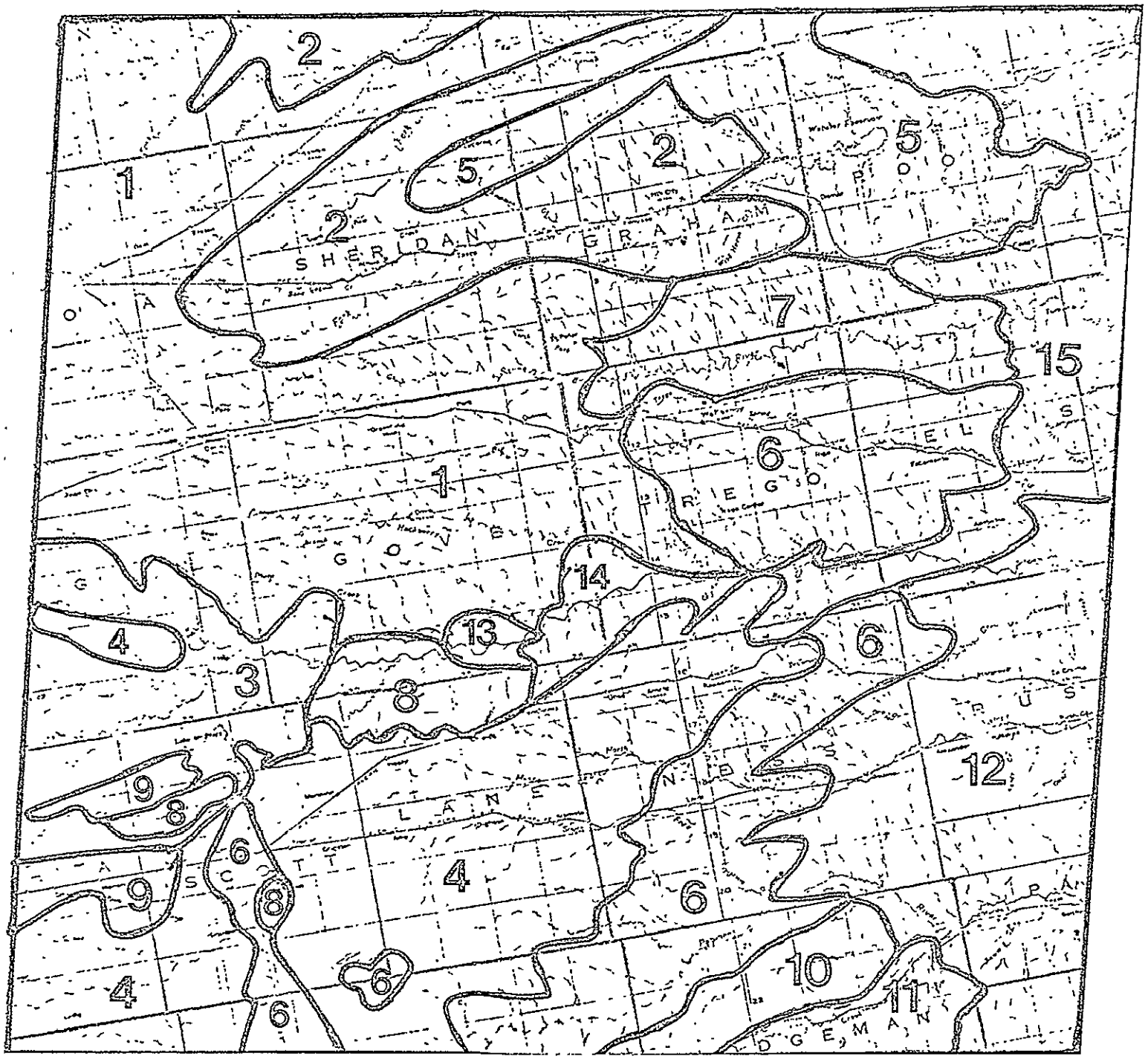


Figure 9. Stratification of Landsat scene 5428-16053, June 20, 1976, based on imagery enhanced by principal components transformation.

The differences between the class structures of two segments were investigated through the use of four measures. Before calculating these attributes, the class structure for each segment was projected into the space spanned by the first three vectors of the Tasselled Cap transformation. The rationale for the use of only the first three components is twofold: (1) the representation of the strata attributes is easier to visualize and interpret in three dimensions, and (2) it has been shown that almost all the information in the four original Landsat bands is represented by the first two principal components, indicating that no significant information in the class structure will be lost in this reduction in dimensionality. The four measures were:

1. The Euclidean distance between the grand means of the clustered data in the two segments. This measure does not use any information about the dispersion of the clusters comprising the class structure of the segments, but does indicate the closeness of the two clouds of data in the Tasselled Cap space.
2. The angle between the vectors specified by the grand means, given by:

$$\arccos \frac{m_1 \circ m_2}{||m_1||\ ||m_2||}$$

where m_1 and m_2 are the respective mean vectors of the two segments, $m_1 \circ m_2$ is the inner product of the vectors, and $||\cdot||$ denotes the length of the vector. This measures the degree to which the class structures of the two segments lie in the same sector of the Tasselled Cap space. If this attribute was small while the Euclidean distance was large,

then one segment has been shifted away from the origin in the same direction that the other segment lies; for example, this could indicate a shift due to haze.

3. The angle between two planes representing the class structures. A plane is constructed for each segment by fitting a plane to the set of cluster means of the segments by least-squares methods. The angle between the planes is the angle between two lines determined by projecting the line joining the grand means into the two planes. If the planes are parallel, this angle is zero. The smaller this angle, the greater the degree to which the planes are parallel.

4. The angle between the normal lines of the planes. This is a more classical mathematical convention for the angle between planes than is attribute 3. It is not affected by the position of the grand means of the segments in the planes, as is attribute 3. A large value for attribute 3 and simultaneously a small value for attribute 4 indicates that the planes are nearly parallel, but the grand means of the segments are far apart; thus, the class structures are different.

In addition to the five spectral attributes, long term climatic data in the form of day degree sums and average precipitation during the growing season were examined for each segment.

Results

The overall spectral response of the strata measured by a mean vector and covariance matrix for each stratum did not show differences among the LACIE Phase III strata using the machine clustering procedures developed in FY76. This was expected since the large strata (as seen in Figures 4 and 5) gave rise to broad normal distributions with a great deal of overlap.

The static stratification of Kansas developed by UCB contains strata which are small in size as can be seen in Figures 6 and 7. The distributions for these strata are not as broad as those based on the LACIE Phase III partitions, but there is still some confusion since strata from different categories are not spectrally distinct.

Due to the problems already mentioned in obtaining digitized boundaries, only two of the stratifications defined by analysts at LARS were measured for overall spectral response. As with the strata developed by UCB, the distributions for the strata were not as broad as those based on the LACIE Phase III partitions developed by USDA personnel at JSC. However, there are also cases where areas the analyst declared as distinct were placed together by the machine clustering of the overall spectral response.

The '75-'76 LACIE segments whose locations are shown in Figures 2 and 3 have been placed in different strata by the various stratification methods. Tables 8 and 9 show the strata in which the LACIE segments were placed.

The segment results will first be discussed for scene 2416-16362, acquired March 13, 1976, which contains segments 1857, 1860, 1865, 1866, and 1889. The strata the segments were placed in are given in Table 8.

Table 8. Partitioning of the '75-'76 LACIE segments in scene 2416-16362, March 13, 1976.

2416-16362, March 13, 1976					
Segment	LACIE	UCB	LARS		
1857	US-5B	211-2/ 88-A	1	1	1
1860	US-5A	211 / 70-B 212-3/ 70-B 220-3/ 75-A	2	2	2
1865	US-5B	211-2/104-C 212-3/104-B	3	3	3
1866	US-5B	211-2/ 88-A	1	4	2
1889	US-3C	211 /139-A	4	5	1
		310 /139-A	5	6	4
		310 /139-C			

The three LARS stratifications were developed by three different analysts using (1) a Tasselled Cap enhancement, (2) a Ratio/Magnitude enhancement and (3) the original Landsat data.

Table 9. Partitioning of the '75-'76 LACIE segments in scene 5428-16053, June 20, 1976.

5428-16053, June 20, 1976 Segment	LACIE	UCB	LARS		
1017	US-8	211 / 93-A 212 / 93-B 220 /102-A	1	1	1
1022	US-8	211-2/ 93-A 220 / 75-G	2	1 2	2
1023	US-8	211-2/ 88-C 212-3/ 93-A	3	1	3
1024	US-8	220 /102-A 310 / 75-I	4	1 3	4
1032	US-5B	211-3/ 88-A 220 /102-A	5	4	5
1851	US-8	211-3/ 70-N 220 / 70-P 310 / 75-G	6	2 5	2
1852	US-5A US-5B	211-3/ 88-A 310 / 75-A	5	4	5
1853	US-5A	211-3/ 70-B 220 / 75-G	2	6	2
1855	US-8	220 / 75-F 310 / 75-F	7	7	6

The three LARS stratifications were developed by three different analysts using (1) and (3) a Tasselled Cap enhancement, and (2) a principal components enhancement.

The LACIE Phase III partitioning and UCB static stratification do have some consistency as the segments labelled US-5B are the only ones containing the UCB land use strata 211-2; however, land use 212-3 appears in both US-5A and US-5B. Only one of LARS analyst-defined stratifications grouped segments 1857 and 1866, which belong to the same stratum in both the LACIE and UCB stratifications.

When the overall spectral response of each segment was examined, segments 1857 and 1889 were grouped together as were segments 1860 and 1866. This is consistent with the third LARS stratification, but not with the LACIE and UCB partitionings. Both the Euclidean distance between the means and the angle between the mean vectors grouped segments 1857, 1866 and 1889 together, which is not consistent with any stratification. Attributes 3 and 4, the two measures of the angle between the planes, grouped segments 1860, 1865, and 1866 together, a group which is somewhat, but not totally contained in the LACIE stratum US-5B.

The climatic data grouped segments 1857, 1866, and part of 1889 together and segments 1860, 1865 and part of 1889 together if only long term average day-degree sums were considered. The precipitation during the growing season grouped segments 1857, 1865, and 1866 together and segments 1860 and 1889 together. The climatic variables were taken in account in the construction of the UCB stratification and so are consistent with it. However, the other stratifications are not completely consistent with the climatic measurements.

The segment results for scene 5428-16053, June 20, 1976, will now be discussed. The strata in which the segments have been placed are

shown in Table 9. As with scene 2416-16362, the LACIE Phase III partitioning and the UCB static stratification are not consistent, although there are some points of agreement. The LARS analyst-defined partitions separate the segments in the LACIE stratum US-8, but group together segments 1032 and 1852 which are from US-5B and US-5A respectively.

When the overall spectral responses of the segments in scene 5428-16053 were examined, segments 1032 and 1853 were separated, with the remaining segments grouped together. This is consistent with the LACIE Phase III partitioning except for segment 1852, which is divided between partitions US-5A and US-5B. This directly opposes the LARS stratifications as segments grouped 1032 and 1853 were with others by the analyst-defined stratifications and many segments considered separate by the LARS stratifications were placed together by the overall spectral response machine clustering.

The Euclidean distance between the means grouped the segments in the following manner: 1032 and 1855; 1017 and 1022; and 1023 and 1851. These do not totally agree with any of the stratifications. The angle between the mean vectors group the segments: 1023, 1851, and 1855; and 1017 and 1022. These groups are within the strata defined by LACIE Phase III but do not include all segments in each stratum.

Both measurements of the angle between the planes gave identical results: segments 1853 and 1855 are grouped together and the remaining segments are grouped together. This is not consistent with previous results.

The climatic data grouped segments 1017, 1023, and part of 1032 together and segments 1022, 1024, part of 1032, 1851, 1852, 1853 and 1855 together if long-term average day degree sums were considered. The precipitation during the growing season grouped segments 1017, 1022, 1023, part of 1852, and 1853 together, segments 1024, 1032 and part of 1852 together and segments 1851 and 1855 together. These climatic variables were taken into account in the construction of the UCB stratifications and so are consistent with it. However the other stratifications are not completely consistent with this information.

Conclusions

Firm conclusions are difficult to draw from the data results obtained in the scene stratification subtask. The stratifications derived from different sources lead to quite different segment partitions.

A rigorous evaluation of the stratifications based on classification performance was not possible due to time constraints. The evaluation of the stratifications based on the spectral attributes and climatic values of the segments is inconclusive as there are many points of agreement and disagreement between the partitionings and the measurements on the segments.

References

1. Nalepka, R. Letter of February 2, 1976.

2. Hay, C.M., and R.W. Thomas. Application of Photointerpretative Techniques to Wheat Identification, Signature Extension, and Sampling Strategy. Space Sciences Laboratory Series 17, Issue 33. University of California at Berkeley. May, 1976.

3. Kauth, R.J., and G.S. Thomas. The Tasselled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. W. Lafayette, Indiana. June, 1976.

4. Trichel, M. Personal Communication. March, 1977.

2.2b Digitization and Registration of Ancillary Data

The ancillary data processing part of the strata project consists of ten tasks, two of which are administrative. The eight technical tasks divide into three general topics: 1.) Ancillary data registration using existing techniques, 2.) Advanced map digitization method research, 3.) Data structure and resource considerations. The first activity was initiated in CY76 and continued in CY77 to complete registration processing and add new variables to the ancillary data base if needed. Extensive manual interaction was required to digitize the ancillary data maps (soil, land use, temperature and precipitation) and the second task was defined to explore more efficient and automatic methods of map digitization. The third category was included to answer basic questions on ancillary data structures and costs. A final report for the year's work on each of the technical tasks is presented next.

Tasks 2, 3 and 4 - Digitization and Registration of Ancillary Data. This task was defined to enable completion of work on registration of the four ancillary variables from CY76 and to add new variables if needed. Completion of the initial registration was achieved in the first quarter and at that point it was decided not to register any additional variables. Topographic data was considered but due to the low relief in the central Kansas test site it was judged that the value of this data would not justify the cost of obtaining it. Additional versions of county temperature and precipitation variables were added; however, this is done very simply once the county boundaries are registered. All these variables were coded for convenience of storage but the actual values may be retrieved during machine analysis by using appropriate calibration codes (LARSYS code 5 or code 6). Table 2.2b-1 is a list of code assignments. The final product is a 13-channel file which con-

Table 2.2b-1 Code assignments and temperature and precipitation values for counties in the Kansas overlay site.

	COUNTY	PRECIPIT		PRECIPIT		MIN TEMP		MIN TEMP		MAX TEMP		MAX TEMP	
1	CHEYENNE	0		0		0		0		0		0	
2	RAWLINS	0		0		0		0		0		0	
3	DECATUR	0		0		0		0		0		0	
4	NORTON	0		0		0		0		0		0	
5	GRAHAM	0		0		0		0		0		0	
6	SHERIDAN	0		0		0		0		0		0	
7	THOMAS	0		0		0		0		0		0	
8	SHERMAN	0		0		0		0		0		0	
9	WALLACE	0		0		0		0		0		0	
10	LOGAN	0		0		0		0		0		0	
11	GOVE	0		0		0		0		0		0	
12	TREGO	44	11.05	61	6.87	0		0		0		0	
13	NESS	41	10.33	56	6.23	69	314.	92	185.	128	577.	155	310.
14	LANE	0		0		0		0		0		0	
15	SCOTT	0		0		0		0		0		0	
16	WICHITA	0		0		0		0		0		0	
17	GREELEY	0		0		0		0		0		0	
18	HAMILTON	0		0		0		0		0		0	
19	KEARNY	0		0		0		0		0		0	
20	FINNEY	0		0		0		0		0		0	
21	HODGEMAN	0		0		0		0		0		0	
22	FORD	0		0		0		0		0		0	
23	GRAY	0		0		0		0		0		0	
24	HASKELL	0		0		0		0		0		0	
25	GRANT	0		0		0		0		0		0	
26	STANTON	0		0		0		0		0		0	
27	MORTON	0		0		0		0		0		0	
28	STEVENS	0		0		0		0		0		0	
29	SEWARD	0		0		0		0		0		0	
30	MEADE	0		0		0		0		0		0	
31	CLARK	0		0		0		0		0		0	
32	COMMANCHE	0		0		0		0		0		0	
33	BARBER	0		0		0		0		0		0	
34	HARPER	0		0		0		0		0		0	
35	SUMNER	0		0		0		0		0		0	
36	SEDGWICK	0		0		0		0		0		0	
37	KINGMAN	0		0		0		0		0		0	
38	PRATT	0		0		0		0		0		0	
39	KIOWA	0		0		0		0		0		0	
40	EDWARDS	0		0		0		0		0		0	
41	PAWNEE	53	13.34	58	6.52	75	338.	97	194.	126	571.	153	306.
42	STAFFORD	66	16.74	95	10.63	76	346.	98	196.	124	558.	150	300.
43	RENO	109	27.41	172	19.33	77	347.	97	195.	124	560.	149	298.
44	HARVEY	0		0		0		0		0		0	
45	MARION	99	24.99	147	16.4	74	335.	96	193.	120	540.5	144	289.
46	MCPERSON	96	24.05	133	14.88	75	341.	94	189.	122	552.	147	295.
47	RICE	71	17.88	87	9.76	75	339.	96	192.	121	546.	146	292.
48	BARTON	69	17.49	90	10.04	78	353.	100	201.	125	564.	152	304.
49	RUSH	56	14.22	70	7.8	70	318.	92	185.	124	559.	150	301.
50	ELLIS	54	13.74	71	7.98	66	297.	89	178.	118	534.	144	288.
51	RUSSELL	67	16.91	82	9.22	71	320.5	93	186.	117	530.	143	287.
52	LINCOLN	69	17.44	85	9.54	84	378.	120	240.	122	552.	148	296.
53	ELLSWORTH	91	22.83	133	14.81	70	319.	92	185.5	122	549.	146	293.5
54	SALINE	104	26.16	153	17.04	75	340.	97	194.	118	532.	143	286.
55	DICKINSON	97	24.42	128	14.25	72	328.	94	189.5	120	544.	145	291.
56	CLAY	58	14.61	86	9.63	74	333.	95	191.	123	554.	148	297.
57	OTTAWA	82	20.79	106	11.88	76	342.	97	195.	122	551.	149	299.
58	CLOUD	72	18.06	79	8.79	71	322.	93	186.	114	515.	141	282.
59	MITCHELL	64	16.16	65	7.28	68	308.4	90	181.	117	528.9	144	289.6
60	OSBORNE	59	14.81	66	7.35	66	301.	88	177.	123	555.	150	300.
61	ROOKS	54	13.58	70	7.85	66	296.5	88	176.5	118	532.5	145	290.
62	PHILLIPS	57	14.28	78	8.69	63	287.5	86	173.5	118	531.	144	289.5
63	SMITH	56	14.11	66	7.38	70	317.	93	186.	119	539.	148	296.
64	JEWELL	67	16.89	61	6.78	66	299.5	85	179.	110	497.	139	279.
65	REPUBLIC	75	18.99	97	9.74	70	316.	92	185.	114	513.	141	282.
66	WASHINGTON	0		0		0		0		0		0	
67	MARSHALL	0		0		0		0		0		0	
68	NEMAHA	0		0		0		0		0		0	
69	BROWN	0		0		0		0		0		0	
70	DONIPHAN	0		0		0		0		0		0	
71	ATCHISON	0		0		0		0		0		0	
72	WYANDOTTE	0		0		0		0		0		0	
73	LEAVENWORTH	0		0		0		0		0		0	
74	JEFFERSON	0		0		0		0		0		0	
75	JACKSON	0		0		0		0		0		0	
76	POTTAWATOMIE	0		0		0		0		0		0	
77	RILEY	0		0		0		0		0		0	
78	GEARY	0		0		0		0		0		0	
79	WABAUNSEE	0		0		0		0		0		0	
80	SHAWNEE	0		0		0		0		0		0	
81	DOUGLAS	0		0		0		0		0		0	
82	JOHNSON	0		0		0		0		0		0	
83	MIAMI	0		0		0		0		0		0	
84	FRANKLIN	0		0		0		0		0		0	
85	OSAGE	0		0		0		0		0		0	
86	LYON	0		0		0		0		0		0	
87	MORRIS	0		0		0		0		0		0	
88	CHASE	0		0		0		0		0		0	
89	COFFEY	0		0		0		0		0		0	
90	ANDERSON	0		0		0		0		0		0	
91	LINN	0		0		0		0		0		0	
92	BOURBON	0		0		0		0		0		0	
93	ALLEN	0		0		0		0		0		0	
94	WOODSON	0		0		0		0		0		0	
95	GREENWOOD	0		0		0		0		0		0	
96	BUTLER	0		0		0		0		0		0	
97	COWLEY	0		0		0		0		0		0	
98	CHAUTAUQUA	0		0		0		0		0		0	
0A	ELK (99)	0		0		0		0		0		0	
0B	WILSON (100)	0		0		0		0		0		0	
0C	NEOSHO (101)	0		0		0		0		0		0	
0D	CRAWFORD (102)	0		0		0		0		0		0	
0E	CHEROKEE (103)	0		0		0		0		0		0	
0F	LABETTE (104)	0		0		0		0		0		0	
0G	MONTGOMERY (105)	0		0		0		0		0		0	
	UPPER BAND	73.10		74.03		73.10		74.03		73.10		74.03	
	LOWER BAND	74.06		74.06		74.06		74.06		74.06		74.06	
	C1	0.1111111		0.25		2.		4.5		2.		4.5	
	C2	0.0277777		0.0277777		0.5		0.5		0.5		0.5	

sists of (i) the original 4-channel LANDSAT data; (ii) soil, land use and county map; (iii) precipitation; and (iv) temperature variables. Such a file is too large to be stored in a single tape, hence it is subdivided in equal portions in three tapes as shown in Table 2.2b-2. Some users found such subdivision is not handy to use, so the entire overlay is re-divided and reduced to two 4-channel files, as shown in Table 2.2b-3.

The final registered data sets were made available to the strata analysts by the end of the second quarter and tasks 2, 3 and 4 were thus completed at that time.

Tasks 5 and 6 - Define and Implement Advanced Map Digitization Methods. The ancillary data digitization methods used in CY76 required extensive and laborious manual use of a table digitizer to convert a map to digital form. Editing and reformatting of the digitized points is an additional task which is tedious and subject to human error. In the case of complicated polygon types of maps, a lot of detail on the map has to be omitted in order to reduce the burden of editing. A flow diagram of the steps required is presented in Figure 2.2b-1.

An advanced method was investigated in this contract year in which the table digitizer is replaced by a film scanner. Table 2.2b-4 compares the steps for the two methods. The map is required to be photo-reduced to a color transparency, which should be small enough to be mounted on the drum of the scanning device. Digitization is subsequently performed automatically so as to generate red, blue and green digital color separation images of the map; then a three-feature classification based on these three color images is used to recover the original color-coded map polygons. This classification is then available for registration with LANDSAT data. This method is highly automatic and retains most of the details in the original map.

Table 2.2b-2(a): List of the 13 channels of final overlay.

Channel	Comment
1	Original data from channel 1 of 74025100
2	Original data from channel 2 of 74025100
3	Original data from channel 3 of 74025100
4	Original data from channel 4 of 74025100
5	Overlay of the soil map
6	Overlay of the county map
7	Overlay of the land use map
8	Total precipitation: Oct 1973 - June 1974
9	Total precipitation: Mar 1974 - June 1975
10	Mean minimum temperature: Oct 1973 - June 1974
11	Mean minimum temperature: Mar 1974 - June 1974
12	Mean maximum temperature: Oct 1973 - June 1974
13	Mean maximum temperature: Mar 1974 - June 1974

Table 2.2b-2(b): Run numbers and tape/file of the final overlay.

Run No.	Tape/File	Comment
74025131	3405/1	Line 1-800
74025132	3406/1	Line 801-1600
74025133	3407/1	Line 1601-2340

Table 2.2b-3: A compact version of the final overlay.

Run 74025141 on file 1 of tape 3408

Channel	Comment
1	Overlay of soil map
2	Overlay of land use map
3	Total precipitation: Oct 1973 - June 1974
4	Total precipitation: Mar 1974 - June 1975

Run 74025142 on file 1 of tape 3409

Channel	Comment
1	Mean minimum temperature: Oct 1973 - June 1974
2	Mean minimum temperature: Mar 1974 - June 1974
3	Mean maximum temperature: Oct 1973 - June 1974
4	Mean maximum temperature: Mar 1974 - June 1974

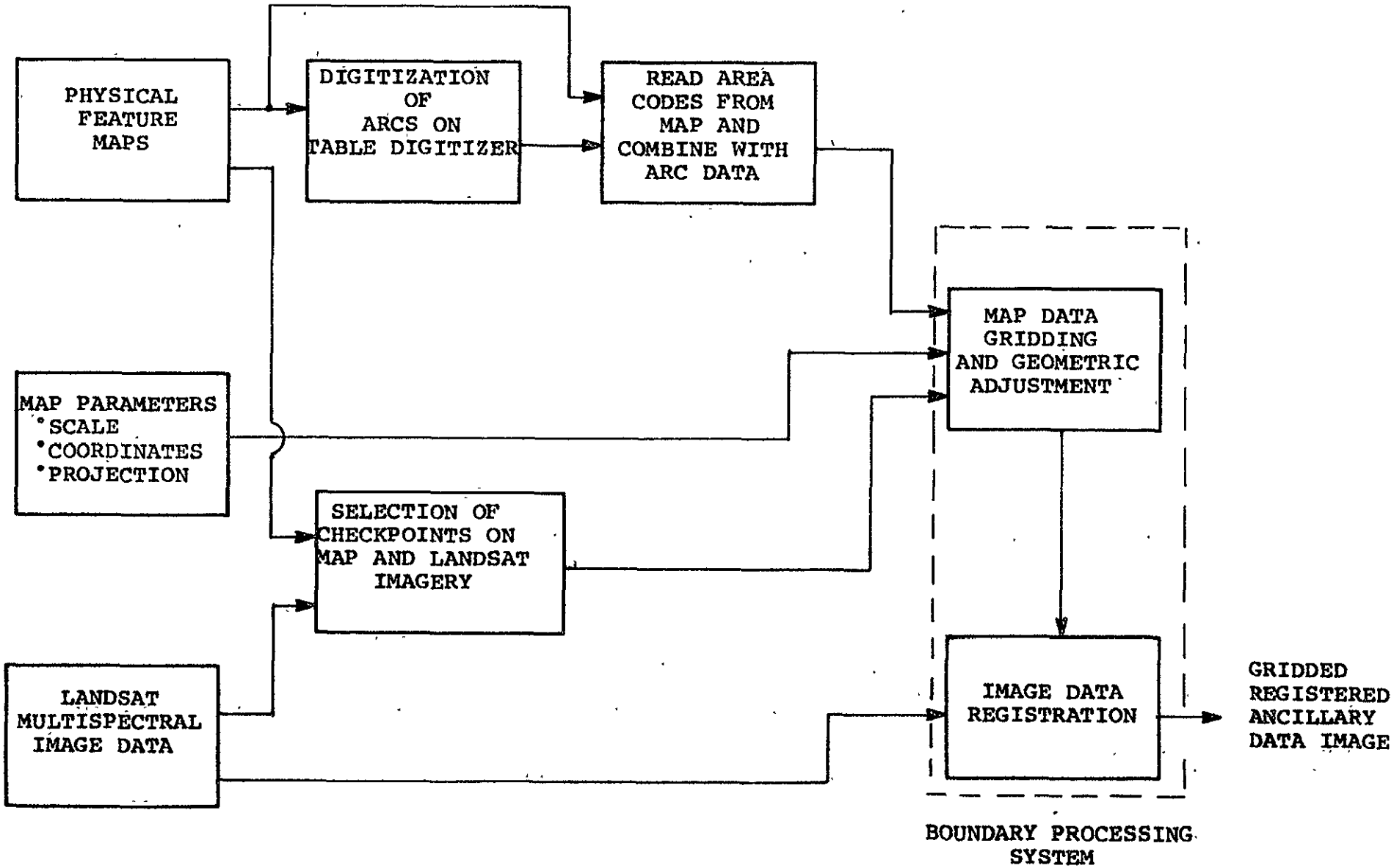


Figure 2.2b-1 Flow diagram of digitization procedure using a table digitizer.

Table 2.2b-4: Comparison of table digitization and scanner digitization.

Table Digitizer	Scanner
1) Spatial Resolution: limited by the resolution of the table digitizer.	1) Limited by the size of aperture of the scanner system.
2) Curve Resolution: limited by human eyes. Previous experience shows that curve can be traced by taking short intervals as 0.1 inch.	2) Limited by size of aperture and degree of concaveness of the curve. Since the aperture can be as small as .100 μm, so it yields far better resolution.
3) Check point accuracy: limited by table digitizer, same as (1) and (2).	3) Limited by size of aperture same as (1) and (2).
4) Data Acquisition Preparation: Need to label arcs before digitizing (high risk of error).	4) Need to photo-reduce the physical map to the size of the drum of the scanner system (low risk of error).
5) Editing: Need expensive re-plot and laborious manual checking.	5) Need post-processing or cleaning but can be highly automated.
6) Job time: Depends on size and complexity of the map.	6) Depends on size of the map only.
7) Efficiency of mass production: May not gain significant improvement.	7) May gain significant improvement.

The map which required the greatest amount of effort for digitization was the land use map of Kansas. It contains a large number of complex polygons in twelve classes (colors). It was this map that initiated the search for an improved method of digitizing complex polygon map types. A black-and-white reproduction of the map is presented in Figure 2.2b-2. This map was used as the test case for the study.

Map Photography. How well the color-coded features on the map can be recovered from its color images depends on how well the classification can be performed. Classification performance is strongly influenced by the quality of the color images, which are ultimately determined by the photo-quality of the transparency. Experimental results show that high photo-quality may be achieved by (1) using high resolution fine grain film such as Kodachrome KM-135-20 and (2) photographing the map under strong uniform sunlight.

It is reasonable to deduce that grain noise may increase variance of the colors, and poor lighting such as inadequate photoflood may reduce contrast of the colors; thus the colors may not fully occupy the available dynamic range; all of these ultimately affect the quality of classification. This is experimentally verified and its effects are reflected in the spectral plots of the colors. Figure 2.2b-3 and Figure 2.2b-4 show the color class coincident spectral plots of the land use map of Kansas. It can be seen that the color classes in Figure 2.2b-3 tend to crowd between values 72 and 144, but the colors in Figure 2.2b-4 tend to spread out between values of 36 and 216. This is due to the strong sunlight which gives excellent color contrast. By comparing the blue channel (Channel 2) of both figures, it can be seen that variances of the colors is larger in Figure 2.2b-3 than in Figure 2.2b-4, where Figure 2.2b-4 corresponds to the fine grain film. A flow diagram of the scanner method of map digitizing is presented in Figure 2.2b-5.

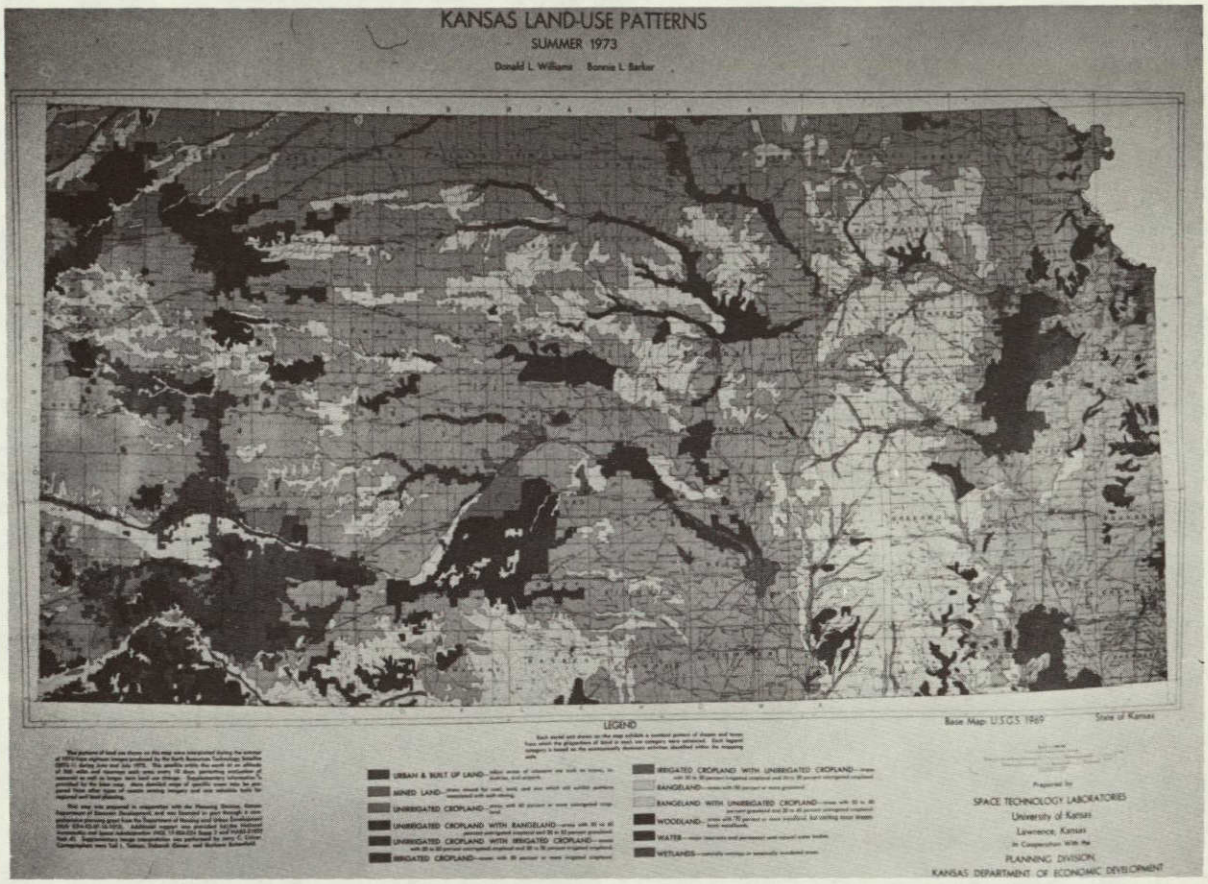


Figure 2.2b-2 Kansas land use map used to test automatic map digitization method.

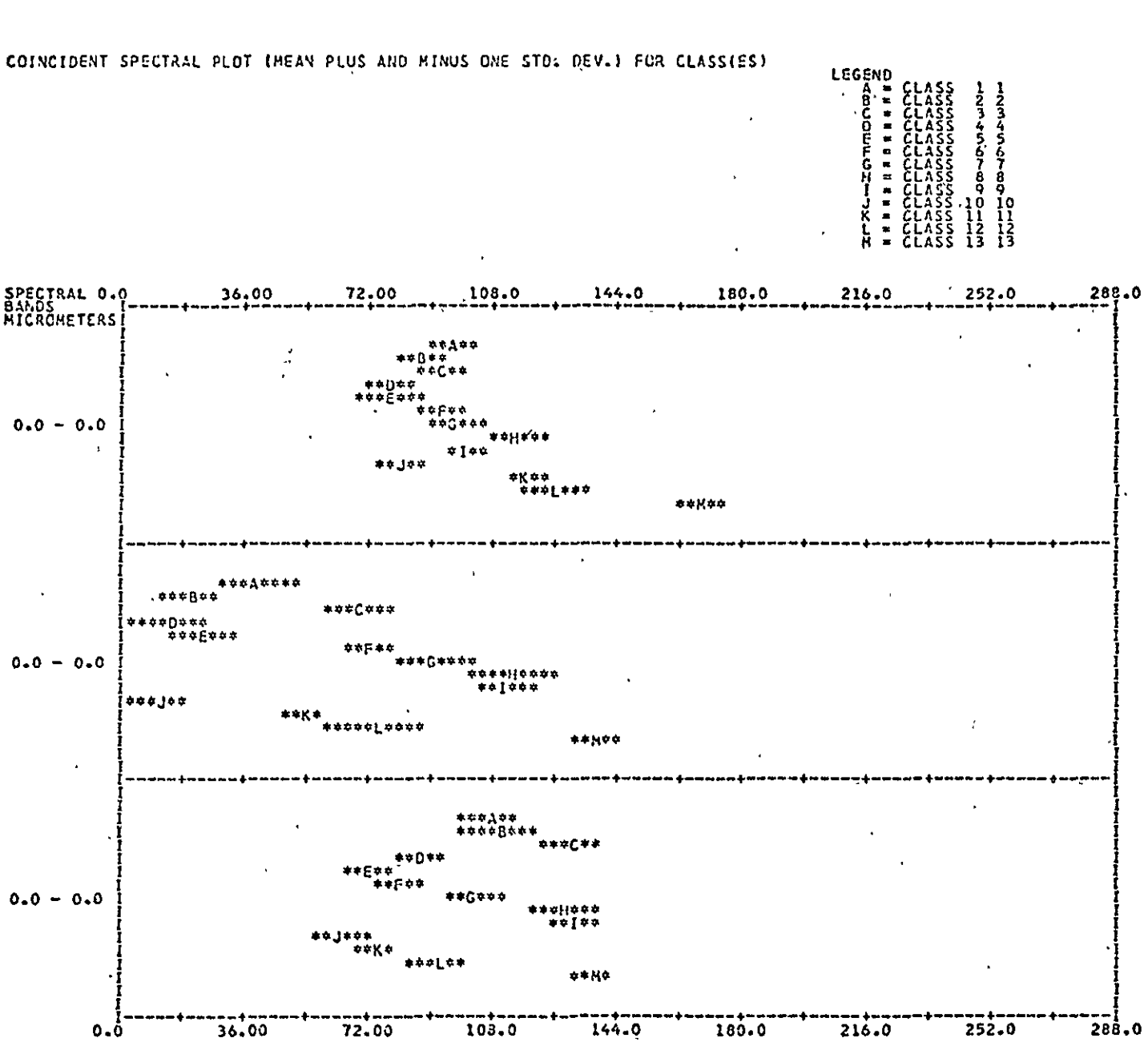


Figure 2.2b-3 Coincident spectral plot of color classes from Kansas land use map photographed under photoflood lamps using coarser grain film.

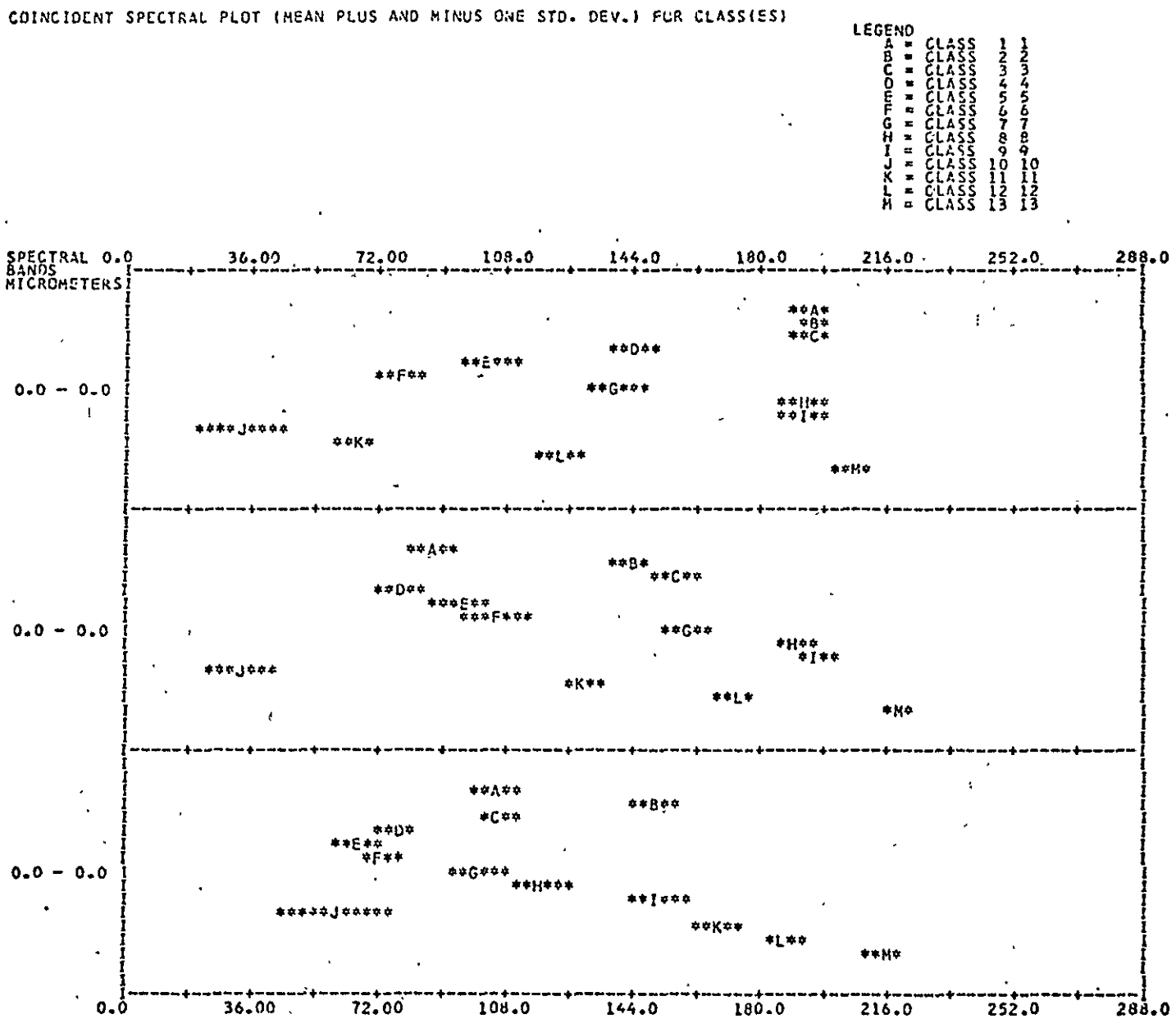


Figure 2.2b-4 Coincident spectral plot of color classes from Kansas land use map photographed under bright sunlight using fine grain film.

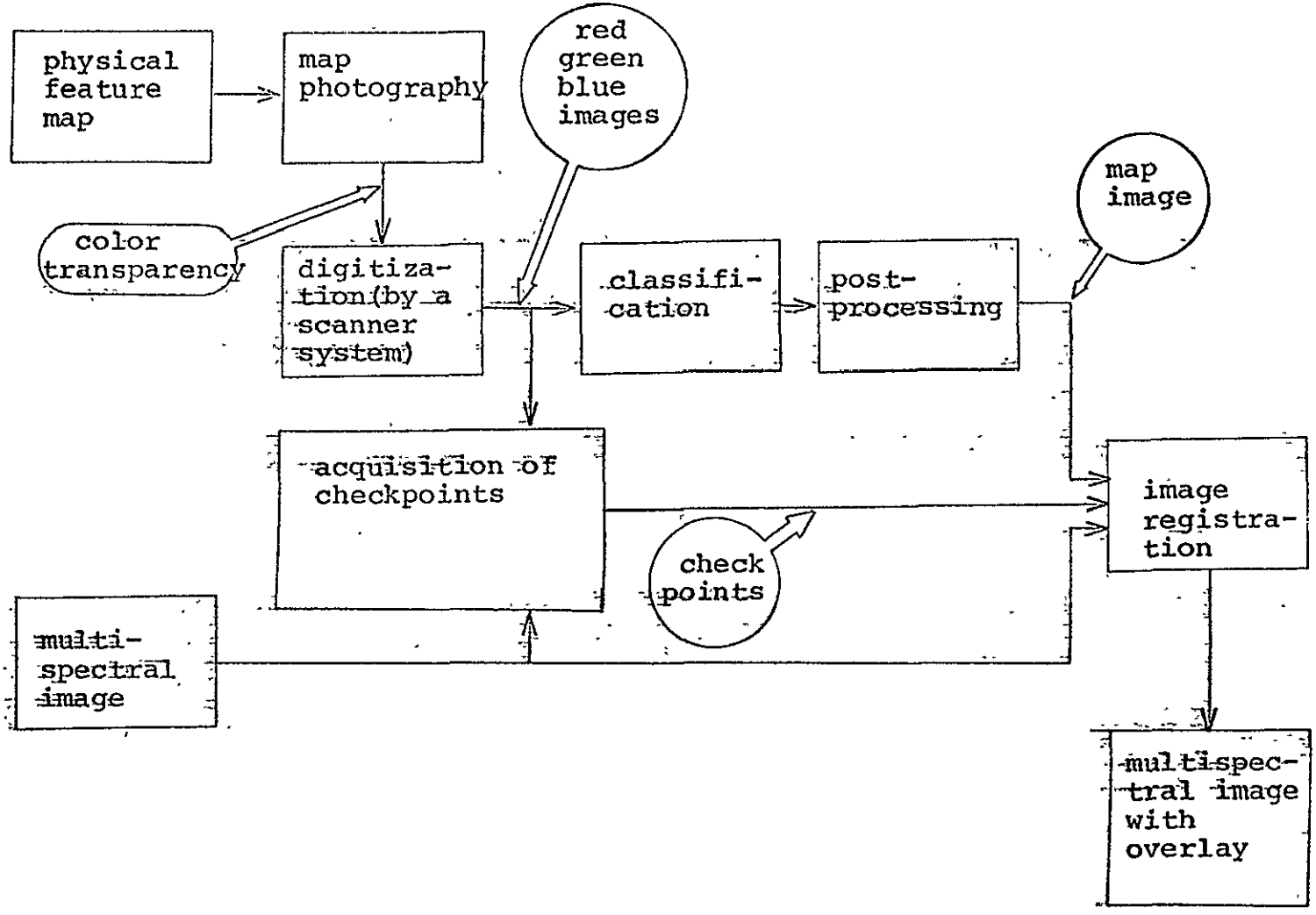


Figure 2.2b-5 Flow diagram for steps used in the scanner method of map digitization.

Resolution. Another objective of advanced digitizing method investigation is to retain as much detail of the map as possible. To do this the digitization rate of the scanner has to be increased to bring about high effective spatial resolution. Three factors influencing effective spatial resolution are:

1. aperture size of the scanner and sampling rate,
2. ratio of photo-reduction,
3. size of the printing dots on the map.

Although there is a minimum aperture size (100 μm, 1 μm = 10^{-6} meters) of the scanner used, it is always possible to achieve a desired effective spatial resolution by appropriate change of the photo-reduction ratio.* However, spatial resolution cannot be increased without limit beyond the dot density of the map (about 100 dots per inch, with respect to the map). If this happens, the scanner sees one dot or part of a dot at a time and records only the color of the dot, which is unfortunately not the true color representation of the feature on the map. Calculations show that the scanner has to "see" three or more dots in order to record a "reasonably true" color of map features. The variance of the recorded color depends on the number of dots that are averaged.

A method to improve chances of accurate recovery of the features without sacrificing effective spatial resolution is to increase sampling rate but keep the size of the scanner aperture fixed. This approach essentially tries to provide more data for classification and post-processing. It may have a better chance of cleaning the misclassified pixels.

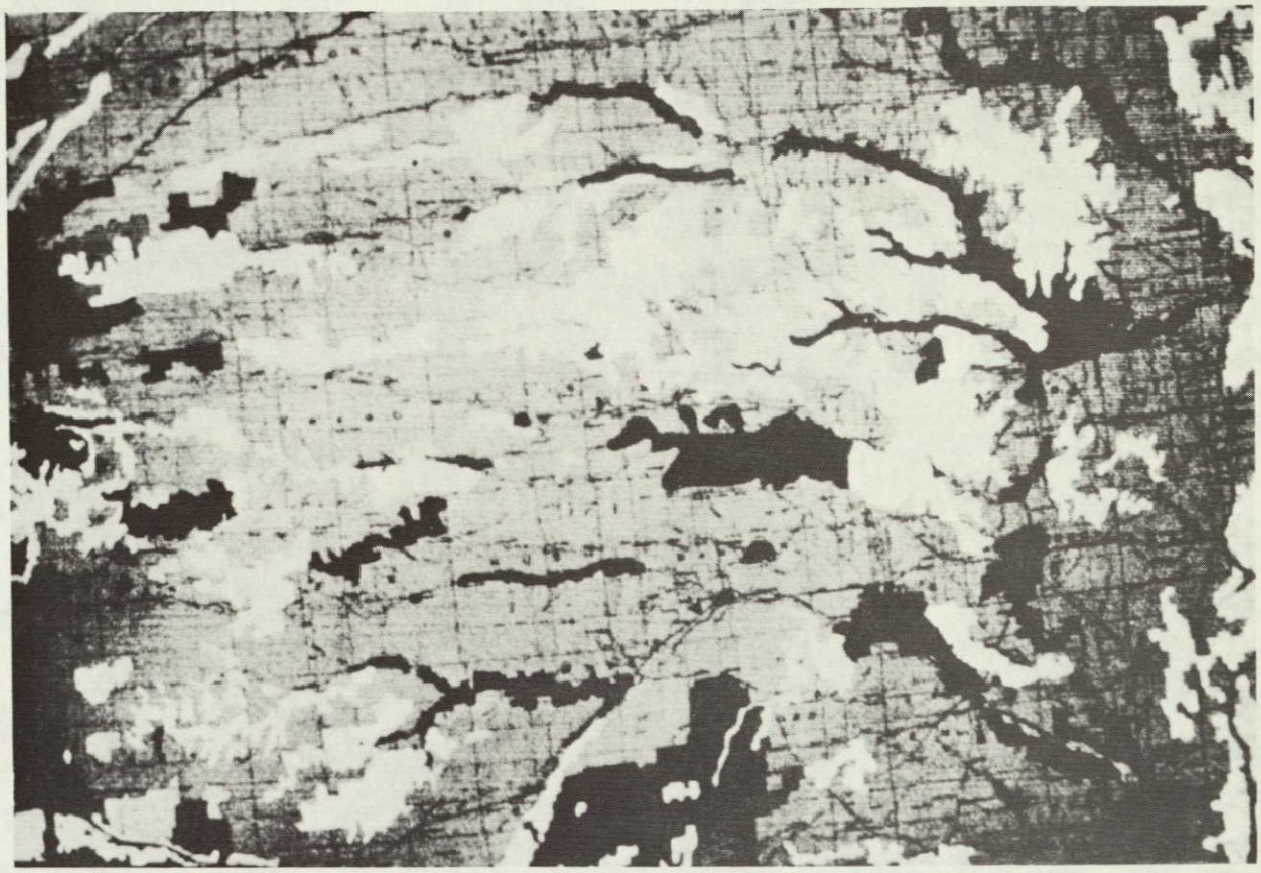
---

*An Optronics International Corporation Model P-1000

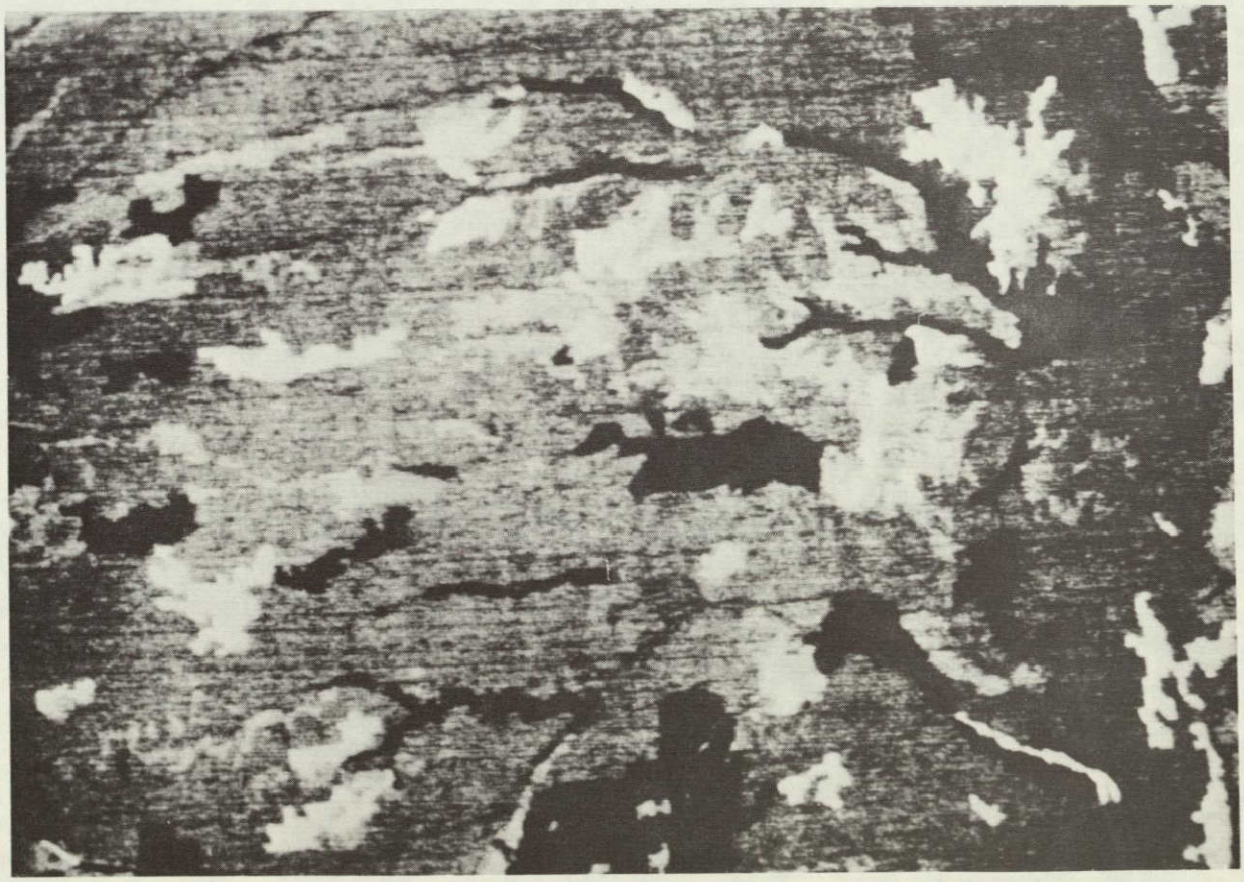
Scanner digitization of the land use map. The central portion of the land use map of Kansas was photographed under bright uniform sunlight on a 35-mm transparency (Kodachrome KM-135-20). The size of the aperture of the scanner was set to 100 microns (100×10^{-6} meters) and sampling interval of 50 microns. This higher sampling rate is used to provide more data for classification and help the cleaning process. The transparency was scanned three times to produce red, blue and green digital images. These image files were reformatted and registered to produce a 3-channel image file of size 510 lines by 790 columns in LARS multi-spectral tape format. It was given a run number of 74025177. From here on, we may use the LARSYS system to display, compute statistics and perform classification. Figures 2.2b-6, a, b and c contain digital display images of the blue, green and red separation images of the color map.

The above choice of aperture size, area covered and sampling rate is a reasonable compromise between spatial resolution and color variance. By both calculations and actual measurements from images, the effective spatial resolution is 27 points per inch with respect to the map. Such a resolution can produce comparable results to those obtained by table digitization. The color statistics, on the other hand, are excellent: variance of each color in each channel is about 5-10 units; color separation within each channel is about 30-50 units. A separability test showed that the corresponding distances between the colors are very large, hence the classification accuracy should be very high.

Classification results. The LARS per-point maximum likelihood classifier was used to classify the 3-channel image file. A digital display image of the classification is presented in Figure 2.2b-7. The results indicate good recovery of the features on the land use map. There is almost no error in field centers but a lot of misclassifications exist along field boundaries.

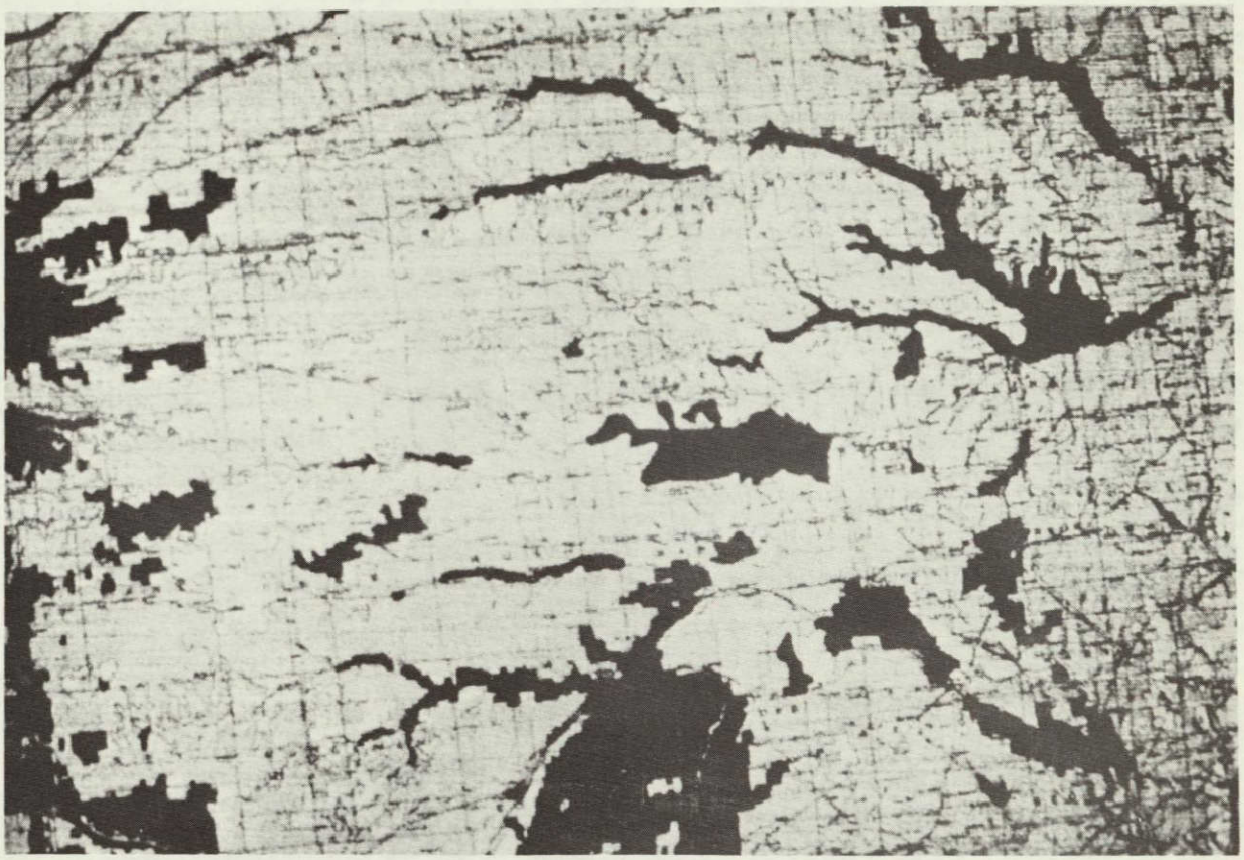


Green separation



Blue separation

Figure 2.2b-6 Color separation images of Kansas land use map.



Red separation

Figure 2.2b-6 Color separation images of Kansas land use map.

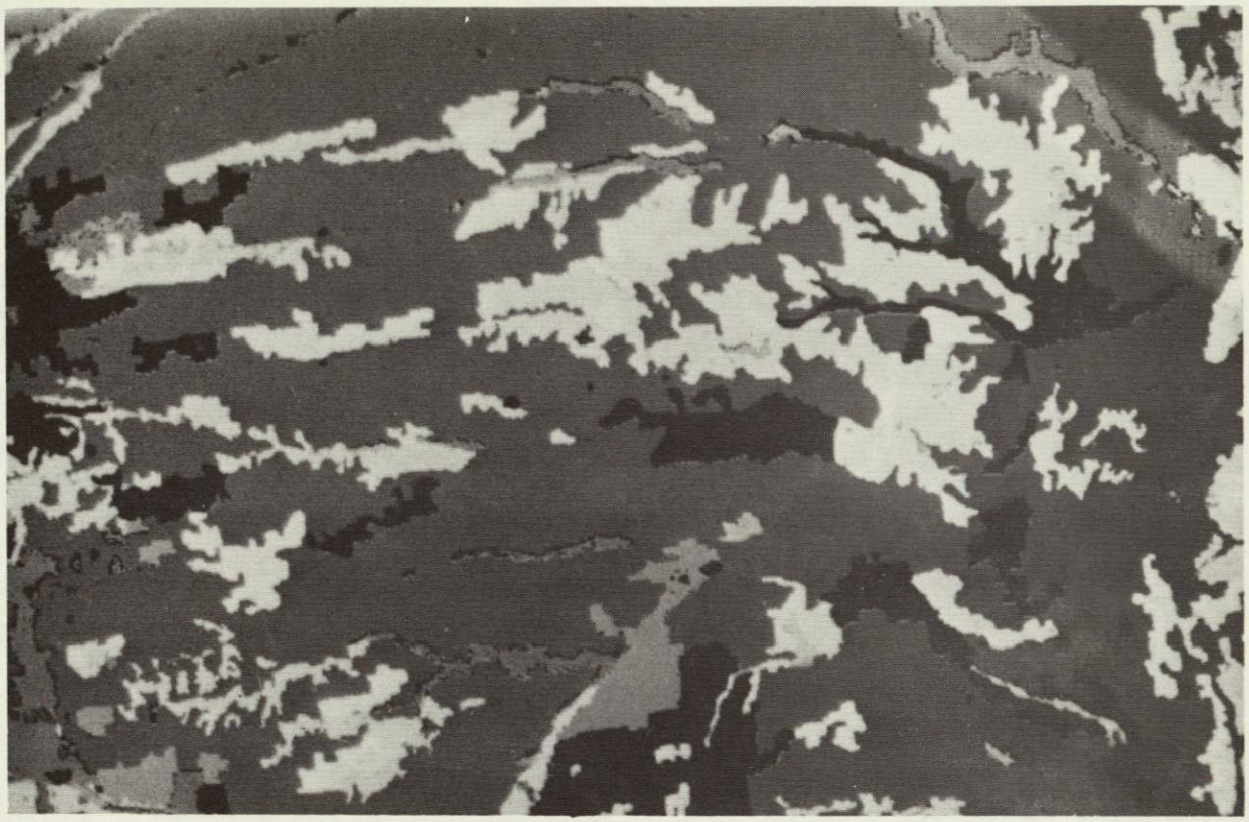


Figure 2.2b-7 LARSYS per-point classification of the three band digitized land use map of Kansas.

Further comparisons with the images show that the error pixels correspond to the lettering and/or marks on the map. These marks are usually "thin" lines and black. During digitization, the scanner "sees" partially black and partially the true colors, hence creating some intermediate color classes. To reduce misclassifications, one may either correctly classify the intermediate classes or correct them afterwards. The first approach suggests using a more powerful classifier. The second approach suggests post-processing. Experiments showed that post-cleaning is more effective, and it was studied in more detail.

Use of an Edge-Detector Post-Processing Approach. An alternate approach to improve boundary recovery is to correct misclassifications after the maximum likelihood classification. A problem associated with this type of approach is that merely looking at one pixel at a time is not sufficient to tell that it is a misclassified pixel. For the purpose of correcting the boundary, it is assumed that misclassification occurs mainly along boundaries. To locate a boundary, a simple derivative-type edge-detector is used. The algorithm of such a detector is presented at the end of this discussion.

One-dimensional Cleaning Process. The following discussion considers a situation in which a boundary is north-south or vertical, and the direction of cleaning is west-east or horizontal. If no misclassification has occurred at the boundary, there should be a clear-cut stepwise transition of pixel class, thus the output of the edge-detector should show a jump of one pixel width at the boundary. However, if some pixels are misclassified to some other classes, other than those on both sides of the boundary, the output of the detector should show consecutive jumps of total width longer than one pixel. The objective of cleaning is to make such a "thickened" boundary "thin". To do this, we estimate the position of the boundary by re-assigning

the pixels on the "thickened boundary" to the classes of the pixels before and after the jumps according to some criterion. For example, a simple half-and-half re-assignment may serve this purpose. Figure 2.2b-8 is an illustration of the cleaning process and the algorithm is listed at the end of this section.

```
                              1    2
                              ↓    ↓
            scan line:        AAAAAAABECCCCDDDDD
edge-detector output:         100000011100010000
      cleaned line:           AAAAAAAACCCCCDDDDD
```

Illustration of the 1-dimensional cleaning process. A "thickened" boundary occurs at position 1, pixel B is re-assigned as A, pixel E is re-assigned as C. A "thin" boundary occurs at position 2, no re-assignment takes place.

Figure 2.2b-8 Cleaning Algorithm

Cleaning process for the classification map. It is obvious that the 1-dimensional cleaning process is effective only when the boundary is not parallel to the direction of cleaning. But the classification map requires a 2-dimensional cleaning process. However, a direct extension of the 1-dimensional method to a 2-dimensional case is handicapped by some problems, such as detection of parallel-running edges; thus a more complex decision scheme is required. A simple but effective method to avoid these difficulties is to apply the 1-dimensional cleaning processing horizontally and then vertically. A major disadvantage of this method is that it requires the interchange of line and column and may be difficult to accomplish if the size of the map is too large.

Accuracy and Resolution. It should be noted that the cleaning process is nothing more than a re-assignment of misclassified pixels based on an estimate of the positions of boundaries. Because the color-coded features are well separated by distinct boundaries that can still adequately reappear after maximum likelihood classification, the estimate of boundary position by the edge-detector is very likely to be sufficiently accurate. It is subsequently expected that pixel re-assignment of boundaries is reasonably accurate. Even with this, the uncertainty of the position of the cleaned boundaries is not greatly affected, hence resolution remains unimproved. As a matter of fact, resolution is expected to be lowered because the re-assignment may move the boundaries slightly. How much a boundary can possibly be moved depends on a parameter L in the 1-dimensional cleaning process. This parameter is the minimum length of consecutive pixels of the same class that are regarded as correctly classified. For example, if L = 2, then in the direction of cleaning, a new class of more than one pixel in length is considered as correctly classified and will not be re-assigned; but a new class of length only one pixel is regarded as misclassified and will be re-assigned. In this case, maximum shift of boundaries is one pixel. This uncertainty of boundary positions has the same effect of doubling the sampling interval. Since both sampling interval and aperture size can affect spatial resolution, which factor will actually be more predominant depends on which is larger. If both the sampling interval and aperture size are the same, then doubling the sampling rate will lower spatial resolution by half. However, the land use map of Kansas was scanned with sampling interval 50 μm and aperture size 100 μm. Doubling the sampling interval will only make the factors the same, so does not lower spatial resolution.

Results of Cleaning. Figure 2.2b-9 shows the classification map pro-

Figure 2.2b-9 Land use map classification processed by cleaning algorithm with parameter L = 2.

Figure 2.2b-10 Map classification processed by cleaning algorithm with parameter L = 3.

Figure 2.2b-11 Land use map classification processed by cleaning algorithm with parameter L = 4.

cessed by the 1-dimensional cleaning process horizontally and then vertically, with the parameter L set to 2. Results show that boundaries are more distinct and many of the misclassifications at the boundaries are removed. Furthermore, it is found that some other misclassified pixels in field centers, especially those corresponding to marks and lettering on the map, are removed. The general classification quality is significantly improved. Better cleaning can be achieved by setting L = 3 (Figure 2.2b-10) and L = 4 (Figure 2.2b-11). The misclassifications that still exist in Figure 2.2b-7 gradually disappear. The price of this is loss of resolution and details. Loss of resolution is observed among some boundaries which appears to be "stair-like". Loss of details is observed among some small features which disappear or are broken into several pieces. This implies that the parameter L must be set to some value which compromises between the degree of cleaning, loss of resolution and detail.

Task 7 - Evaluation of Scanner Digitization Method. Scanner digitization method was proposed as an alternate method to table digitization method because it can be highly automatic, reduces production cost in certain cases and improves the quality of digitization. The cost evaluation is delayed to the Task 9 discussion where it is more meaningful to talk about cost if it is compared with table digitization method. The evaluation here will be mainly the quality of digitization.

Figure 2.2b-12 shows a portion of the land use map classification of Kansas. The overlay transparency is a re-plot of table-digitized data. The gray scale image is the cleaned classification map of the scanner-digitized data.* Their scales are re-adjusted so that they can be overlaid

*Only 8 different gray levels can be displayed, so the same gray level has been assigned to green and orange, causing an illusion of losing patches of colors.

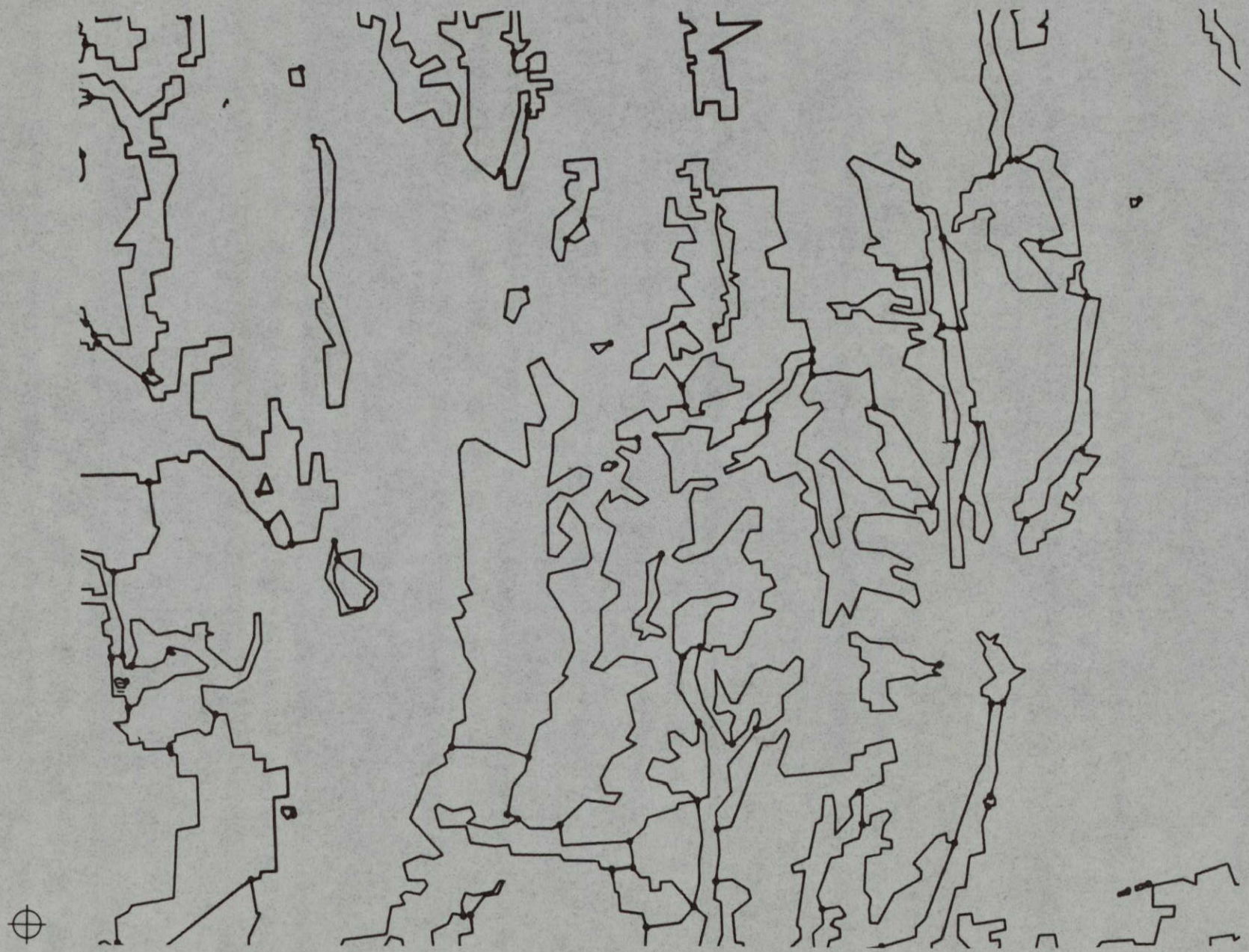


Figure 2.2b-12 Comparison of land use map classification and table digitized land use map boundaries.

for comparison. Regarding the table-digitized data as "ground truth", the cleaned classification matches extremely well with it, especially the boundaries of color patches in the central portion of the figure. The colors are orange, brown, green, yellow and light yellow. There are some misclassifications, most of them belonging to small patches of colors, partially contaminated with black marks and/or writings. On the whole, the scanner-digitized maps do have comparable quality with the table-digitized maps.

It is noticed with interest that scanner digitization has a problem of losing or gaining patches of colors. These are not due to misclassifications occurring along field boundaries and the cleaning process cannot remove them. Actually the cleaning process may enhance their existence by regarding them as correctly classified. This problem significantly hurts feature fidelity and this problem does not exist in table digitization. Tracing back to the origin of these misclassifications, we find that the writing and marks on the map are responsible. However, most maps are available with marks and writing, so direct application of scanner map digitization will eventually encounter this problem. Another trouble source is spatial resolution. As indicated before, effective spatial resolution cannot be increased without limit because maps are usually printed with dots of colors, rather than patches of color. Most maps are printed with 100 dots per inch, allowing three dots for color averaging. The best resolution is only 33 points per inch. This is slightly lower than the accuracy of maps, which is about 50 points per inch. If it happens that the full 50 points per inch resolution of the original map is needed, direct use of scanner method will not be able to attain that high resolution. Furthermore, the cleaning process inherently sacrifices a certain amount of resolution.

We strongly feel that there should be some modifications before scan-

ner digitization can actually be applied to map digitization. As a suggestion, a possible modification may be to touch up or color-code the entire map before it is photo-reduced. Thus, all writings and marks are eliminated. The boundaries are clear-cut without black lines. This will give a better transition of colors across boundaries. The dot density problem is eliminated because dots do not exist anymore. Thus, photo-reduction ratio may be adjusted simply to whatever the resolution is desired to be. Furthermore, the original map need not be a color-coded map.

Certainly there are other modifications that can improve scanner digitization method. For example, one may use a better classifier to handle the intermediate classes so that misclassifications may be reduced. One may improve cleaning algorithm by using more complex criterion of cleaning, or one may even abandon the use of color. The use of black edges against white background may be an alternative.

However, even with some misclassifications and a resolution of 27 points per inch, scanner digitization method, as experimented, shows comparable or even better quality as table digitization method whose curve resolution is only 10 points per inch because spacings between points can only be as close as 0.1 inch. With the advantage of being highly automatic cost can be sharply cut and job time much shorter. Scanner digitization would be an attractive alternative.

Task 8 - Coordinate Systems for Data Base. Previous work on digitization and registration of physical maps with satellite data requires a data conversion process from polygon storage to grid storage mode. This process consists of a series of steps of extracting data information from polygons, converting them to some intermediate form, and finally producing data in grid structure. Depending on the complexity and size of the polygons, such a polygon-to-grid conversion is generally time-consuming, dis-

crediting the use of polygons as a storage method if the data are needed for frequent analysis where fast economical access is highly demanded. Another storage method is considered below. Its space requirement is usually more than that of polygons but still substantially less than that of grid storage method, in fact, for the ancillary data of the four variables registered for this project, each set of data can comfortably reside in the core memory, providing quick random accessibility. We call this method a compressed-line storage method or simply, line compression. Table 2.2b-5 is a direct comparison of polygon, line compression and grid storage methods in a general sense.

As an illustration of these methods, a simple hypothetic map of three polygons is considered as in Figure 2.2b-13. This map has five lines by five columns, thus grid storage requires only 25 bytes. Polygon storage method requires a list of all line segments which compose the boundaries of all the polygons, as shown in Table 2.2b-7. Depending on the actual storage assignment, the space needed may possibly be more than 25 bytes because the polygons are too small. Line compression storage method requires a list of column-content pairs for each scan line, as shown in Table 2.2b-8. This is how the line is compressed: scanning from the first column of each, the data or content is stored and the column number before the content changes. Depending on how frequent the data changes, each line may have many and/or different number of these pairs. Saving of space is noticeable if data does not change much along each scan line. A storage assignment of line compression of the hypothetic map is shown in Figure 2.2b-14, where three arrays are used. The array NP is a book-keeper of all addresses to the column-content pairs which are stored in array NC and ND. Store requirement for this map is $4 \times 6 + 2 \times 8 + 1 \times 8 = 48$ bytes, again more than the 25 bytes required by the grid storage method. However, this situation is quickly reversed when size of map increases. Typical polygon storage

Table 2.2b-5 Comparison of Storage Methods

	Polygon	Line-Compression	Grid
1) Space efficiency			
a) general dependence	on complexity and size of polygons	on complexity, size, and arrangement of polygons	independent
b) efficiency for large area	extremely high typical less than 1%	very high typical 1-2%	100%
2) Possibility of residing in core memory	excellent	good	unlikely
3) Conversion to grid storage			
a) algorithm	complex	simple	-----
b) CPU time	lengthy	reasonably short	-----
4) Conversion from grid			
a) algorithm	(don't know)	simple	----
b) CPU time	(don't know)	reasonably short	-----
5) Random accessibility if residing in core			(restricted access to small sectior
a) algorithm	complex	simple	simple
b) CPU time	lengthy	reasonably fast	fast

Table 2.2b-6 Comparison of various storage methods when applied to county, soil and land use map of Kansas (one LANDSAT frame, 2340 lines by 3226 columns).

	County Map	Soil map	Land use map
1) Space requirements (bytes)			
a) Polygon	5558	19040	89936
b) Compressed-line	48838	48478	136240
c) Grid	7548840	7544880	7544880
2) Conversion CPU time (hour)			
a) Polygon to grid	0.375	0.512	0.762
b) Compressed-line to grid	0.058	0.059	0.062
c) Grid to compressed-line	0.081	0.083	0.086
3) Random-accessibility (sec/32260 pixels)			
a) For compressed-line, by point	0.621	0.664	1.04
b) For compressed-line, by line	0.116	0.117	0.118
c) For grid, by point	0.818	0.818	0.818
d) For grid, by line	0.068	0.068	0.068
f) Sequential accessibility for grid (sec/32260 pixels)	0.050	0.050	0.050

Note: All figures are estimates only; actual time depends on computer and system load.

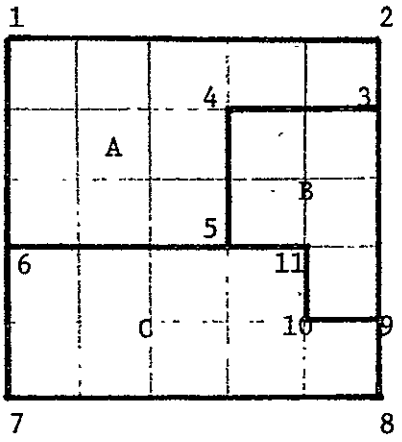


Fig. 2.2b-13: A simple hypothetic map of three polygons of total size = 5x5=25 pixels.

Table 2.2b-7: Polygon data structure of the map in Figure 2.2b-13.

line segment	first line	first col	last line	last col	CL/CR
1-2	0	0	5	0	0 A
2-3	5	0	5	1	0 A
3-4	5	1	3	1	B A
4-5	3	1	3	3	B A
5-6	3	3	0	3	C A
1-6	0	0	0	3	A 0
6-7	0	3	0	5	C 0
7-8	0	5	5	5	C 0
8-9	5	5	5	4	C 0
9-10	5	4	4	4	C B
10-11	4	4	4	3	C B
11-5	4	3	3	3	C B
9-3	5	4	5	1	B 0

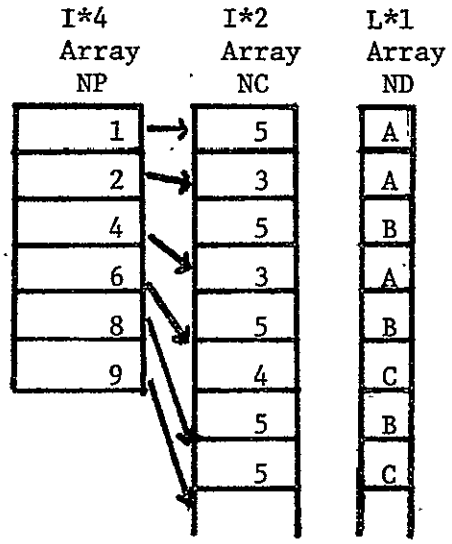


Fig. 2.2b-14: Contents of arrays to store compressed-line data of Fig. 8.1. Array NP contains pointer to array NC & ND. NC contains the column number; ND contains the content of data.

Table 2.2b-8: Compressed-line data structure of the map in Figure 2.2b-13.

line	col/C	col/C
1	5 A	- -
2	3 A	5 B
3	3 A	5 B
4	4 C	5 B
5	5 C	- -

requires only 0.5 to 1% of space of grid storage; line compression is about 1 to 2%.

Unlike polygons, line compression may be handicapped by certain arrangements of polygons. An example of this is shown in Figure 2.2b-15(a), which has five rectangles with the long side parallel to the columns. A lengthy compressed-line list is required as shown in Table 2.2b-9(a). But an extremely simple list is obtained when the long side is parallel to the scan lines, as shown in Figure 2.2b-15(b) and Table 2.2b-9(b).

In addition to very minimal space requirements, line compression has two significant advantages.

1. Easy conversion to and from grid-storage: Following is an algorithm to perform polygon-to-grid conversion (USGS-LARS boundary algorithm). It requires lengthy sorting of all line segments and heavy file manipulations. Grid-to-polygon conversion may not require sorting but some form of multiple searching which may again be complex. Compressed-line-to-grid conversion and reverse, as outlined by Alg. 2.2b-4 and Alg. 2.2b-5, requires comparisons only which are certainly simple and economical to perform.

2. High random accessibility: both polygon and compressed-line files are generally small enough to be loaded into core memory, thus providing quick random accessibility. However, algorithm to retrieve data of a particular pixel is excessively complex and the CPU time requirement may be excessively long. This is not time for line compression because algorithms to retrieve data

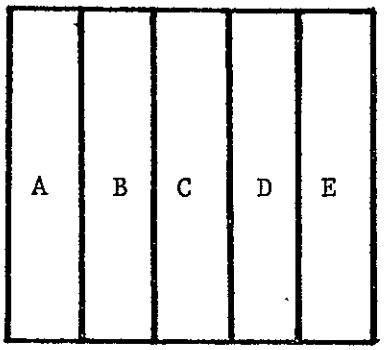


Fig. 2.2b-15(b): A hypothetic map of polygons which handicaps efficient compressed-column storage.

Table 2.2b-9(b): Compressed-column data structure of the map in Fig. 2.2b-15.

line	col/C	col/C	col/C	col/C	col/C
1	1 A	2 B	3 C	4 D	5 E
2	1 A	2 B	3 C	4 D	5 E
3	1 A	2 B	3 C	4 D	5 E
4	1 A	2 B	3 C	4 D	5 E
5	1 A	2 B	3 C	4 D	5 E

Table 2.2b-9(a): A line column data structure of the map in Fig. 2.2b-15.

line	col/C
1	5 A
2	5 B
3	5 C
4	5 D
5	5 E

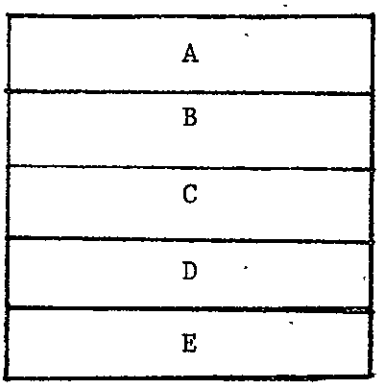


Fig. 2.2b-15(a): A hypothetic map of polygons which promotes efficient compressed-column storage.

require comparisons only (see Alg. 2.2b-6 and 2.2b-7). Grid files are usually too large to be loaded into core memory; however, it can reside in core only a small section at a time, but buffer management is complex and lengthy to execute.

Table 2.2b-6 provides comparisons obtained by some experiments of storage efficiency, conversibility of storage modes, and random accessibility of these methods for the county, soil and land use map of Kansas. It seems that line compression is an excellent alternative to polygon or grid storage method.

Task 9 - Evaluation of Resources for Digitization and Conclusion. Digitization of physical feature maps requres close human attention, often laborious manual operations assisted slightly by machines. Such heavy dependence on human interaction affects not only the cost, but the quality as well as the time required to make data available to potential users. On the other hand, registration is a highly automatic process whose quality and job time are controllable and its cost does not depend on map complexity which varies a lot from map to map.

Our comparison here mainly concentrates on cost of digitization of the maps digitized for this project, plus a map that was digitized for another LARS project. The reason for including the map of Bolivia is because it is extremely simple and gives more insights about how complexity of maps affects cost. Table 2.2b-10 is a list of the digitization cost on an item-to-item basis. It can be seen that cost increases with map complexity for the table digitization method. The major expenditure is editing/plotting (item (7)) and processing (item (9)). These two items account for over 80% of total cost and it is evident that it depends on number of line seg-

Table 2.2b-10 Cost of digitization of various maps.

	(Table Digitization Method)				(Scanner Method)
	Bolivia Map +	County Map	Soil Map	Land Use Map	Land Use Map
1) Number of line segments	200	490	2250	11428	11428
2) Map area (in.^{2})	28x26	14x25	12x23	13x25	13x25
3) Size of intermediate file	1824	43092	57114	79800	-----
4) Map preparation *^{1}	-----	2/$6	4/$12	10/$30	10/$30
5) Photography	-----	-----	-----	-----	/$60
6) Digitization *^{2}	3/$27	5/$45	9/$81	16/$144	2/$140
7) Editing/plotting	/$100	/$150	/$300	/$1100	-----
8) Reformatting	-----	-----	-----	-----	/$100
9) Processing					
a) CPU	0.7/$185	3/$795	4.1/$1086	6.1/$1616	2.5/$662
b) labor *^{3}	5/$50	18/$180	26/$260	35/$350	10/$90
10) Post-processing	-----	-----	-----	-----	-----
11) Estimated cost	$362	$1176	$1739	$3240	$1744

+: Digitization of Bolivia map does not belong to this project but is listed here for cost comparison.

*^{1}: Non-technical labor rate of $3/hr is used.

*^{2}: For table digitization, machine rate is $6/hr; labor is $3/hr.
For scanner digitization, machine rate is $60/hr; labor is $10/hr.

*^{3}: Technical labor rate of $10/hr is used.

NOTE: All figures are estimates only; actual costs may vary.

ments and size of maps. On the other hand, the major cost of scanner digitization is processing and post-processing (items (9) and (10)) which accounts for over 75% of total cost. It is expected that map complexity will only affect the total cost slightly. Figure 2.2b-16 graphically displays the cost of each digitization method versus map complexity. It shows an interesting "break even" point of the two methods. This suggests that one method is advantageous over another for some types of maps and reversed for another type of map.

When it comes to the choosing of digitization methods, several factors should be considered; among these are: quality, efficiency of mass production, and job time. Quality here is meant by spatial resolution and accuracy of map information. Because of limitations of human eyes, spatial resolution may not be very high, but map information can be made to be extremely accurate. Because of very small aperture, scanner method may provide better resolution, but due to classification error and post-cleaning, map information may be distorted.

Table digitization does not gain much in efficiency when several maps are digitized simultaneously because time depends on map complexity. However, scanner digitization does get more efficient because actual scanning time is relatively small, and most of the time cost is in setting up the machines. Finally, job time to table-digitize a map depends on map complexity; but scanner digitization has a relatively fixed time cost. Table 2.2b-11 contains recommendations for future map digitization. As a general rule, use of table digitization is recommended if the map is simple, otherwise for more complex polygon maps, scanner digitization is more attractive.

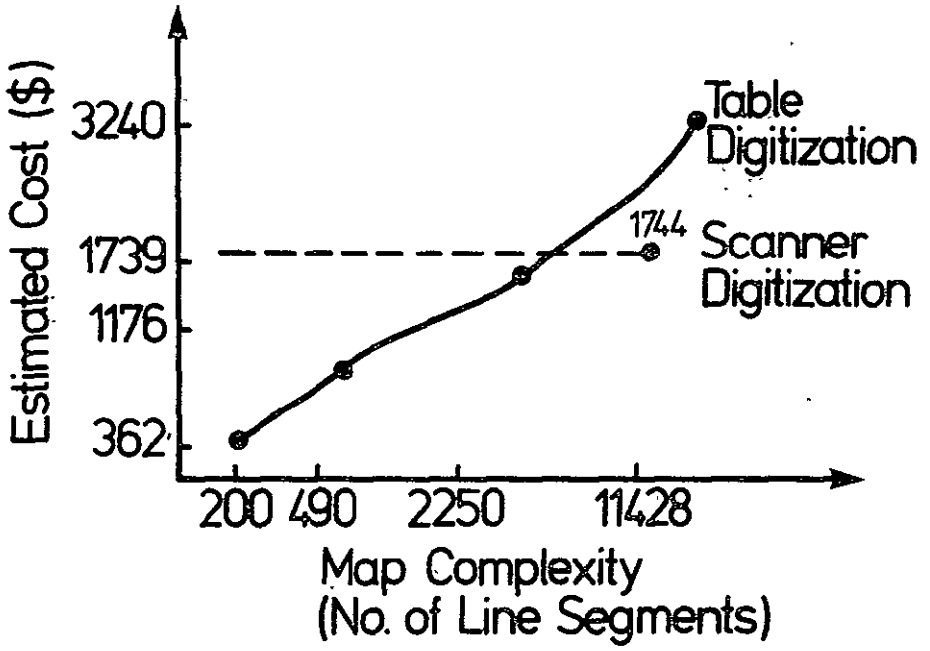


Figure 2.2b-16 Estimated cost versus map complexity for table and scanner digitization methods.

Table 2.2b-11 Advantages of the two digitization methods for maps of various complexity.

	simple map	average map	complicated map
1) Cost	TD is lower	about the same	SD is much lower
2) Quality			
a) feature	TD is better	TD is better	TD is better
b) resolution	SD is higher	SD is higher	SD is higher
3) Job time	TD is shorter	SD is shorter	SD is much shorter
4) Efficiency gain by mass production	SD slightly advantageous	SD is advantageous	SD is advantageous
5) Recommended method	TD	TD or SD	SD

TD -- table digitization
SD -- scanner digitization

Algorithm for a Simple Edge-Detector (2.2b-1)

Input: a line of pixels of length NN in array PIC

Output: the derivative of above in array IG of dim NN

Step 1: Set IG(1) = 1 (a convention)

Step 2: Do step 3 from K = 2 until NN

Step 3: If PIC(K) differs from PIC(K-1), then IG(K) = 1, otherwise IG(K) = 0

The 1-Dimensional Cleaning Algorithm (2.2b-2)

Step 1: Define L = length of consecutive pixels that are regarded as correctly classified

Step 2: Do step 3 through step 10 from first line to last line

Step 3: Get output from the edge-detector in IG

Step 4: From IG, search a consecutive block of zeros of length longer than NJ = L-1 pixels

Step 5: If search is unsuccessful, go to step 6, otherwise go to step 7

Step 6: (only left side of a boundary found) re-assign pixels according to left side of the boundary. Go to step 10.

Step 7: Test width of boundary. If equals 1, go to step 9 (i.e., no re-assignment)

Step 8: (a "thickened" boundary found) If left side of boundary lies outside the line, re-assign pixels according to right side of the boundary. Otherwise re-assign according to half-left-half-right criterion.

Step 9: Test end of line. If yes, go to step 10. Otherwise, go to step 3.

Step 10: Continue.

Alg. 2.2b-3 Polygon-to-grid conversion (USGS-LARS boundary algorithm)

Input: arcs or boundaries describing the polygons

Output: a grid file

Step 1: breakdown of arcs into line segments

Step 2: generation of an array of endpoints only in the form of

LINE COL SLOPE CL CR

where CL and CR are the data to the left and to the right of the segments

Step 3: From this array, for each line segment, the extend (i.e., the smallest rectangle in which the line segment lies diagonally in) is computed. For each line within the extend, the first and last column of the line segment is computed. A file of the form

LINE START-COL END-COL CL CR

is generated.

Step 4: This file is sorted by line and then by starting column.

Step 5: For each line (now in ascending order), fill in data according to the CL and CR of the boundaries (also in ascending order).

Alg. 2.2b-4 Compressed-line-to-grid conversion

Input: data in compressed-column storage

Output: a grid file

Step 1: For each line, execute Alg. 8.5 to expand the compressed data.

Step 2: output the expanded data

Alg. 2.2b-5 Grid-to-compressed line conversion

Input: a grid file

Output: a compressed-line file

Step 1: For each line, do step 2.

Step 2: Starting from first column, compare data of adjacent columns; if the same, repeat comparison for next column; otherwise store column number and data content.

Step 3: Store address to the first 'col-content' pair of each line.

Alg. 2.2b-6 Random-access by point to compressed-line storage

Input: line and column number, NLINE, NCOL

Output: return data in IL.

Step 1: Set K1=NP(LINE), K2=NP(LINE+1)-1 (NP is the array containing address to array NC and ND).

Step 2: If NCOL is greater than NC(K1), go to step ; otherwise IL=ND(K1) and exit.

Step 3: For K from K1 to K2, do step 4.

Step 4: If NCOL is greater or equal NC(K), repeat this step with K incremented; otherwise IL=ND(K) and exit.

Step 5: IL=ND(K2) and exit.

Alg. 2.2b-7: Random access by line to compressed-line storage

Input: line number

Output: data of one line in LL

Step 1: Set K=1, KX=NP(LINE), and K2=NC(NP(LINE+1)-1).

Step 2: If K is greater than K2, exit. Otherwise, do step 3.

Step 3: If K is greater than or equal to NC(KX), go to step 4. Otherwise increment KX.

Step 4: Set LL(K)=ND(KX) and increment K. Go to step 2.

2.2c Crop Inventory Using Full-Frame Classification

INTRODUCTION

Wheat area estimates have been made by LACIE for the state of Kansas utilizing a sample segment approach. The procedure used has been to select 84 5x6 nm segments in Kansas, allocated to the counties in proportion to their historical area of wheat, and to classify the segments to make county estimates of the area of wheat which are then aggregated to make crop reporting district and state estimates.

Desirable attributes of any estimate are accuracy and precision. Accuracy refers to the size of deviations from the true parameter (i.e. the bias of the estimate), while precision refers to the size of deviations from the mean of all the estimates of the parameter obtained through repeated applications of the sampling procedure (i.e. the scatter of the estimates). Studies need to be made to determine the accuracy and precision of the LACIE method of estimation and compare it to alternatives.

Studies of classification error and of estimation accuracy are difficult to carry out since absolute "truth", either in terms of reference data or in terms of the area of wheat in Kansas, is not available. The accuracy of sampling schemes is related to training and classification methods and thus can only be evaluated relative to the specific training and classification methods used. Due to the above factors, this study has been limited to the sampling error or precision of estimates derived from several different sampling schemes.

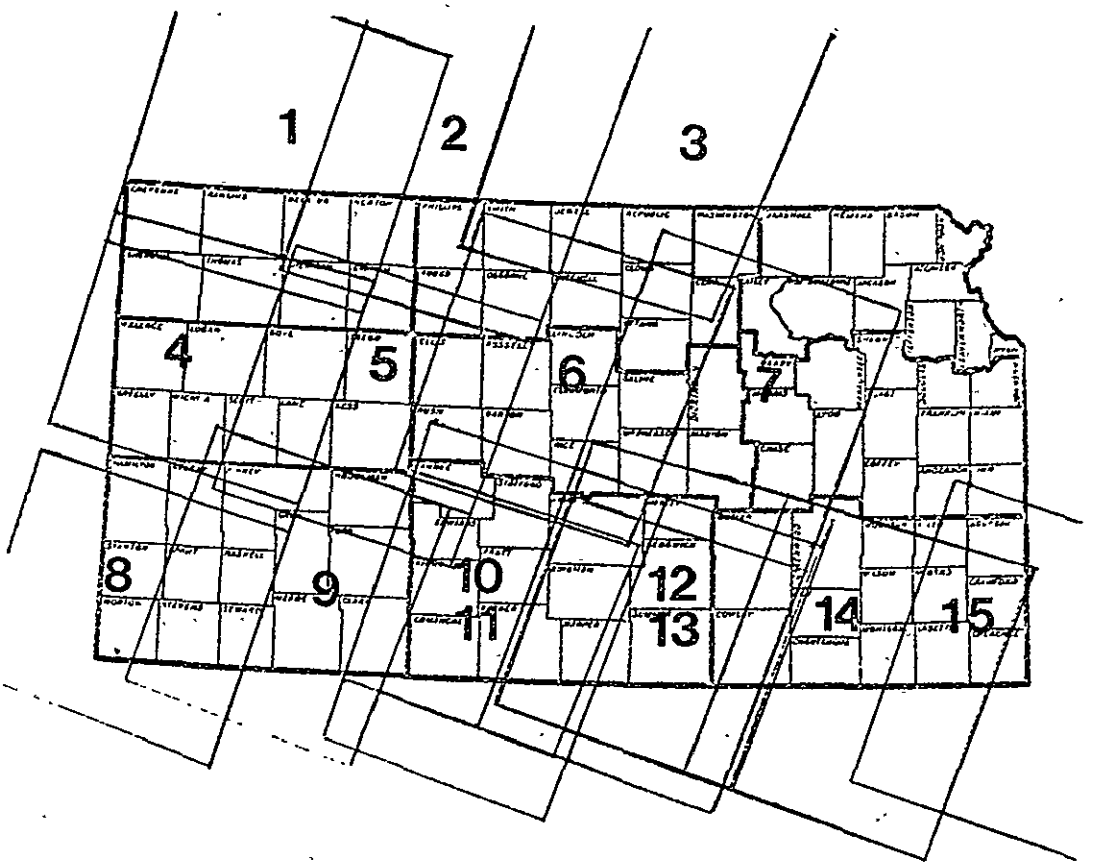
OBJECTIVE

The objective of this task was to compare the sampling errors associated with crop inventory methods using full-frame classification and the LACIE approach of sample segment classification.

GENERAL APPROACH

Full-frame classifications of Kansas into wheat and non-wheat made in another investigation [1] were used in this study. The classifications were done on a county basis using every other line and every other column of Landsat data. Training statistics were developed using aerial infrared photography taken along several flightlines dispersed throughout the state and were extended to counties lacking reference data but known to have similar land use, crops, and soils. Eighty counties comprising seven crop reporting districts (referred to as the pseudo-state) were included. The Landsat scene ID, LARS run number, and date of acquisition for the Landsat frames used in these classifications are given in Figure 1. The full-frame classifications were considered to have negligible sampling error and were repetitively sampled to simulate alternative sampling plans.

The experimental design consisted of three parts: (1) comparison of sampling error achieved by using a fixed total sample size (number of pixels), but varying the segment size and number of segments, (2) calculation of the number of segments of a given size required to achieve the sampling error of the LACIE-like allocation of 84 5x6 nm segments, and (3) comparison of the area estimates achieved by all sampling methods.



Key

	Landsat Scene ID	LARS Run Number	Date
1	2165-16450	75013800	July 6
2	2146-16392	75005800	June 17
3	2163-16334	75006500	July 4
4	2165-16453	75004600	July 6
5	2146-16395	75005900	June 17
6	2163-16340	75006600	July 4
7	2144-16282	75005600	June 15
8	2147-16460	75006200	June 18
9	5032-16310	75007200	May 21
10	2073-16342	75001500	April 15
11	2109-16341	75005000	May 11
12	2072-16284	75000900	April 9
13	2144-16284	75005700	June 15
14	2107-16225	75004900	May 9
15	2142-16171	75005400	June 13

Figure 1. Landsat Coverage for Kansas.

In the first part, the total number of pixels in the sample was held constant and the segment size and number of segments were varied. Four segment sizes were used:

Segment Size	No. of Segments in State	No. of Segments in Pseudo-State
5x6 nm	84	75
4x4 nm	157	137
2x2 nm	630	560
Pixel	481,572	427,587

The sampling error for each of these plans was then calculated and an analysis of variance performed to determine if a significant difference existed among the precision of estimates derived under these plans. Although the relative precision of the estimates associated with the segment sizes tested can be compared, the optimal sampling scheme for a given application depends on additional information such as the cost of acquiring the segments.

In the second part, the sampling error was held constant at the level achieved by the 5x6 nm segments and the number of segments of each size needed to achieve this error was calculated. The theoretical number of segments of size 5x6 required was calculated to test the validity of the assumptions made.

Thirdly, the area estimates obtained by each of these plans for the pseudo-state were compared to determine if a bias was introduced through any of the sampling schemes.

EXPERIMENTAL PROCEDURES

Part One. LACIE procedures for determining the allocation (number) and location (geographic placement) of segments were followed whenever

possible and applicable. Some changes had to be made in the procedure outlined in [3] to: (1) correct inconsistencies in the allocation method and (2) permit extension of the general method to other segment sizes.

Allocations strictly according to the LACIE procedure produced county allocations which did not add to the total number allocated for the crop reporting district (CRD). It was found that LACIE had also encountered this problem and had adjusted their allocations to achieve consistency. Determination of the number of segments per county generally followed the scheme given below because more consistent results were obtained than with the method given in the LACIE documentation:

Value of n_k [1]	Segments Allocated
0 - .3	0
.3 - .6	PPS [2]
.6 - 1.6	1
1.6 - 2.6	2

etc.

The county allocations for each sampling scheme are given in Table 1.

The location of sample segments differed in two respects from the location of the LACIE segments: first, in the definition of nonagricultural areas and, second, in the number of segments permitted in a window or extended rectangle about a given segment.

Nonagricultural areas of at least 2x2 miles in size were excluded from consideration as sample segments. The boundaries of urban areas, federal lands, reservoirs, etc., appearing on county maps prepared by

---

(1) For each county, n_k is determined as a function of the county land area and the historical proportion of wheat. More details are available in [3].

(2) Probability proportional to size (0 or 1 segment).

TABLE 1. NUMBER OF SAMPLE SEGMENTS ALLOCATED PER COUNTY FOR EACH SAMPLING SCHEME.

COUNTY	5X6	4X4	2X2	PIXEL
CHEYENNE	1.	2.	9.	6677.
DECATUR	1.	2.	7.	5615.
GRAHAM	1.	2.	7.	5578.
NORTON	1.	2.	7.	5362.
RAWLINS	1.	2.	9.	7111.
SHERIDAN	1.	2.	8.	5818.
SHERMAN	1.	2.	9.	7047.
THOMAS	1.	3.	11.	8231.
CLAY	1.	1.	6.	4283.
CLOUD	1.	2.	7.	5094.
JEWELL	1.	2.	7.	5567.
MITCHELL	1.	2.	7.	5537.
OSBORNE	1.	2.	8.	5863.
OTTAWA	1.	2.	7.	5335.
PHILLIPS	1.	2.	7.	5174.
REPUBLIC	1.	1.	6.	4547.
ROOKS	1.	2.	8.	5810.
SMITH	1.	2.	7.	5462.
WASHINGTON	1.	2.	7.	4958.
GOVE	1.	2.	9.	6622.
GREELEY	1.	2.	7.	5562.
LANE	1.	2.	7.	5121.
LOGAN	1.	2.	9.	6719.
NESS	1.	2.	10.	7632.
SCOTT	1.	2.	7.	5255.
TREGO	1.	2.	8.	5897.
WALLACE	1.	1.	6.	4497.
WICHITA	1.	1.	6.	4819.
BARTON	1.	2.	9.	6917.
DICKINSON	1.	2.	8.	6124.
ELLIS	1.	2.	8.	5932.
ELLSWORTH	1.	2.	6.	4887.
LINCOLN	1.	2.	7.	5038.
MCPHERSON	1.	2.	9.	6994.
MARION	1.	2.	8.	6144.
RICE	1.	2.	7.	5669.
RUSH	1.	2.	7.	5528.
RUSSELL	1.	2.	8.	5863.
SALINE	1.	2.	7.	5255.
CLARK	1.	2.	7.	5366.
FINNEY	1.	3.	11.	8634.
FORD	1.	3.	11.	8287.
GRANT	1.	1.	5.	3773.
GRAY	1.	2.	8.	6383.
HAMILTON	1.	2.	8.	6135.
HASKELL	1.	1.	6.	4360.
HODGEMAN	1.	2.	8.	5879.
KEARNEY	1.	1.	6.	4871.
MEADE	1.	2.	9.	6570.
MORTON	1.	1.	5.	3635.
SEWARD	1.	1.	5.	4128.
STANTON	1.	1.	6.	4593.
STEVENS	1.	1.	6.	4461.
BARBER	1.	2.	9.	7074.
COMANCHE	1.	1.	6.	4722.
EDWARDS	1.	1.	6.	4630.
HARPER	1.	2.	9.	6477.
HARVEY	1.	1.	5.	4044.
KINGMAN	1.	2.	9.	6719.
KIOWA	1.	1.	6.	4811.
PAWNEE	1.	2.	8.	5943.
PRATT	1.	2.	7.	5675.
RENO	2.	3.	13.	9889.
SEDGWICK	1.	2.	10.	7609.
STAFFORD	1.	2.	8.	5900.
SUMNER	2.	3.	13.	9656.
ALLEN	0.	1.	2.	1850.
BOURBON	1.	1.	2.	1852.
BUTLER	1.	2.	8.	5804.
CHAUTAUQUA	0.	1.	3.	2109.
CHEROKEE	1.	1.	4.	3303.
COWLEY	1.	2.	9.	6653.
CRAWFORD	0.	1.	3.	2512.
ELK	0.	1.	2.	1846.
GREENWOOD	0.	1.	3.	2571.
LABETTE	1.	1.	4.	3378.
MONTGOMERY	1.	1.	4.	3167.
NEOSHO	0.	1.	3.	2598.
WILSON	1.	1.	4.	2761.
WOODSON	0.	0.	2.	1416.

the State Highway Commission of Kansas, Department of Planning and Development were found using a coordinate digitizer at LARS. The boundary definitions of nonagricultural areas were more crude than those defined by LACIE. The reasons for this include: (1) constraints of time (including computer time) and resources (including maps which were not as detailed as might have been possible) and (2) the belief that only major nonagricultural areas needed to be excluded because of experience in another investigation [1] showing that even when very few nonagricultural areas are excluded, estimates of high accuracy can be obtained. The constraint that a sample segment not fall within a nonagricultural area was ignored with the pixel sampling method due to excessively high costs of computer checking for each of the nearly four million samples.

The constraints concerning the number of segments permitted in a given size rectangle centered about the sample segment and its east-west and north-south extensions to 80 nm and 100 nm, respectively, were adjusted by number and size of the rectangle to be relatively consistent with the constraints for the LACIE 5x6 nm segments and are presented in Table 2. This type of constraint was not feasible and, thus, not included in the pixel selection procedure.

The selection of sample segments was computer-implemented. This allowed a large number of segments to be chosen with little personnel time and also facilitated choice of any segment size or number of segments. The Landsat coordinates of each of the sample segments chosen for this investigation are available, but are not given in this report due to the large volume of segments selected. The greater number of samples which could be taken through automated selection permitted statistical tests of precision.

Eight replications were drawn for each of the four sampling plans.

Table 2. LACIE - type constraints for the different segment sizes.

Segment Size	Rectangle Considered	Segments Allowed in Extended Rectangle	
		E-W	N-S
5x6 nm	10.5 x 12	4	8
4x4 nm	8.4 x 8	6	10
2x2 nm	4.2 x 4	12	20

Wheat areal estimates were calculated for each replication for the counties, crop reporting districts, and pseudo-state according to [4]. For each sampling plan, a sampling error was computed for the estimate using four replications. Since at least two errors were necessary for each plan to permit a good statistical analysis, eight replications were drawn and two sampling errors, based on two subsets of four replications, were computed for each plan. Analysis of variance was performed on these errors to assess differences in precision of the sampling plans.

Part Two. The objective of this part of the analysis was to calculate the number of segments of each segment size required to achieve the sampling error found from the 5x6 nm segments. Assuming stratified sampling with each county as a stratum the number of samples required to achieve a fixed variance V (the variance of the estimates obtained from 5x6 nm segments) is given by:

$$n = \frac{(\Sigma\, W_h S_h \quad \sqrt{c_h} \quad) \; \Sigma \, \frac{W_h S_h}{\sqrt{c_h}}}{V + \frac{1}{N} \; \Sigma \; W_h S_h^{\,2}}$$

where c_h is the cost per unit in stratum h, $S_h^{\,2}$ is the variance within stratum h, W_h is the weight of stratum h, and N is the total number of units in the population [2]. If the cost per unit in each stratum is assumed to be the same,

$$n = \frac{(\Sigma\, W_h S_h)^2}{V + \frac{1}{N} \; (\Sigma\, W_h S_h^{\,2})}$$

The assumptions necessary to carry out these calculations will be discussed in the experimental results section.

Part Three. Analysis of variance was performed to investigate differences among the area estimates found using the various sampling plans. Significance of the test statistic would indicate the introduction of a bias through one or more of the estimation plans but, since the true area is unknown, no conclusions can be drawn as to the most accurate estimate.

EXPERIMENTAL RESULTS

Part One. The sampling errors of the wheat area estimates of the pseudo-state for each sampling plan are given in Table 3. A plot of the results (Figure 2) illustrates the variability of each plan. It clearly shows that estimates obtained using smaller segment sizes are more precise.

Since the variances were homogeneous, an analysis of variance was run to examine the differences. The results showed that the variances differed at the 8% level of significance. A multiple range test (the least significant difference) showed that the 5x6 nm segments had a significantly higher variance than any of the other schemes.

Part Two. Using stratified random sampling, the number of segments of each segment size required to achieve the sampling error of the 5x6 nm scheme was to be calculated. The number of 5x6 nm segments which should have been required to achieve the fixed error was calculated but the unreasonable result showed that the assumptions necessary for this calculation apparently were not met. We have concluded that this approach to determining the required number of segments was not possible because the within stratum (county) variability must be calculated in terms of wheat area, and insufficient 5x6 nm segments were taken in any given replication to obtain a variance estimate. In

Table 3. Sampling errors (standard deviations) achieved by different sampling plans with a fixed total sample size.

Plan	Standard Deviation	Coefficient of Variation
	(Ha)	(%)
5x6 nm	134820 *	2.4
	266143	4.5
4x4 nm	99657	1.8
	39641	0.7
2x2 nm	82429	1.5
	46273	0.8
pixel	11801	0.2
	12603	0.2

* The first value is found from the first four replications of a sampling scheme and the second value is from the last four replications.

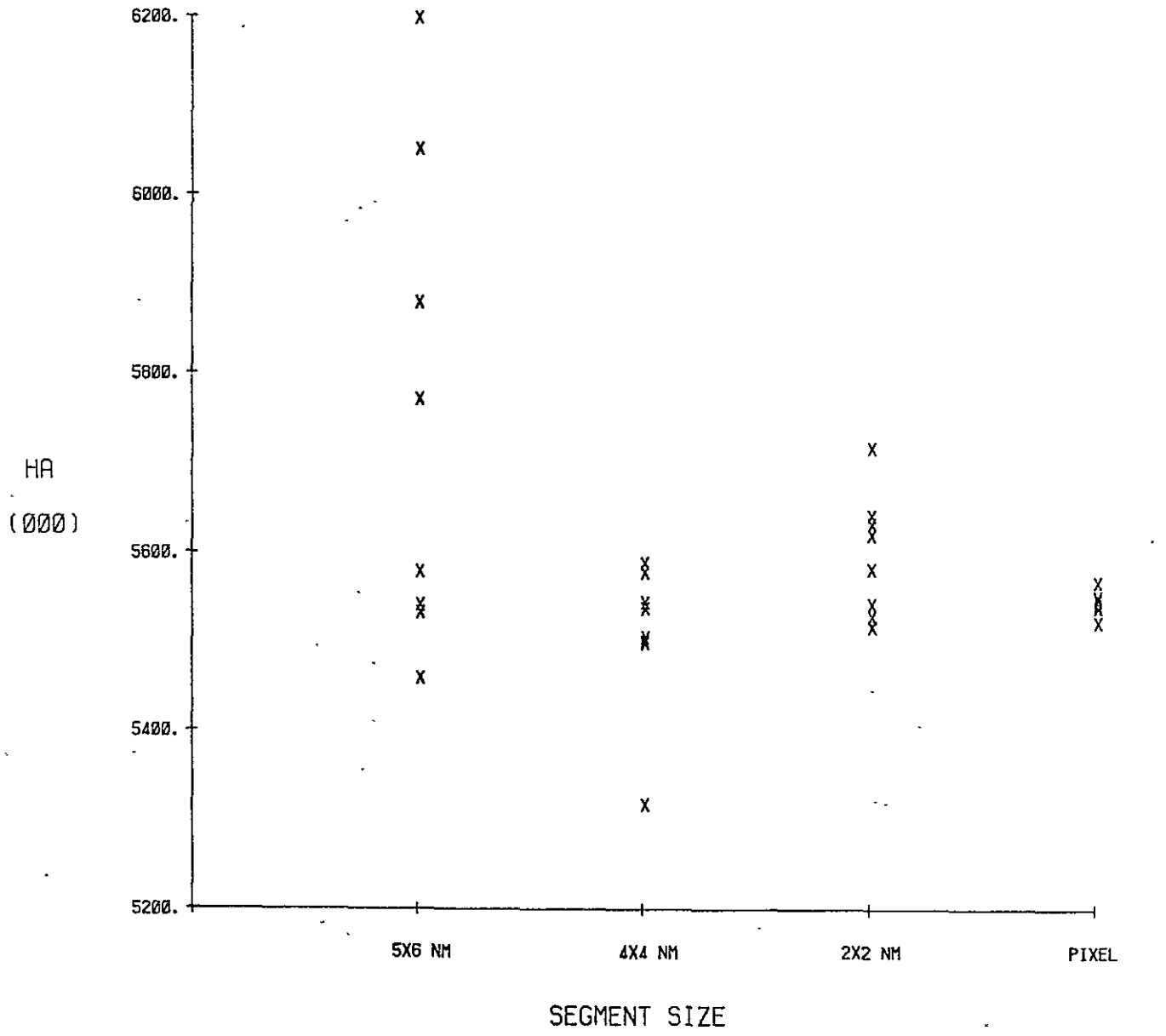


FIGURE 2. COMPARISON OF WHEAT AREA ESTIMATES MADE WITH DIFFERENT SEGMENT SIZES.

addition, the restrictions on placement of sample segments prohibit the use of replications to estimate the variance and, furthermore, it is not clear that these restrictions could be fully taken into account in any theoretical calculation.

Part Three. Area estimates achieved for each crop reporting district for each of the eight replications of the four sampling schemes were calculated and compared to determine if any bias had been introduced in the estimates by the sampling scheme employed. An analysis of variance showed a difference at the 5% level of significance. Multiple range tests showed the estimates obtained by sampling 5x6 nm segments were different from the other estimates, possibly because they are more likely to include some nonagricultural areas; the other area estimates did not differ from one another. The pseudo-state area estimates for each replication and the average estimate achieved by each scheme are presented in Table 4.

SUMMARY AND CONCLUSIONS

The results of this investigation are well illustrated in Figure 2. The area estimates found by the use of 5x6 nm segments cover a much larger range of values and thus have a larger variability than any of the other segment sizes. The estimates become more and more precise as the segment size decreases and more segments are taken. The estimates obtained from the 5x6 nm segments were different from the estimates obtained by the other segment sizes, indicating a bias in the 5x6 nm sampling scheme.

The estimates achieved using the 5x6 nm segments have the least precision of any sampling scheme tested. However, to assess the implications of this result for the LACIE sampling scheme, additional

Table 4. Area estimates achieved by different sampling plans with a fixed total sample size.

Plan	Area of Wheat	
	Each Replication (Ha)	Average (Ha)
5x6 nm	5,526,460	5,744,360
	5,534,130	
	5,452,630	
	5,763,840	
	5,872,010	
	6,043,350	
	6,191,050	
	5,571,410	
4x4 nm	5,310,490	5,501,740
	5,530,430	
	5,498,490	
	5,491,940	
	5,581,020	
	5,537,270	
	5,571,060	
	5,493,230	
2x2 nm	5,510,240	5,590,640
	5,613,220	
	5,634,280	
	5,710,130	
	5,625,250	
	5,521,900	
	5,574,300	
	5,535,820	
pixel	5,540,820	5,538,320
	5,541,670	
	5,559,670	
	5,531,420	
	5,515,120	
	5,538,990	
	5,543,820	
	5,535,070	

factors must be considered. In order to fully evaluate the scheme, the method of training and classification which would be used in conjunction with a sampling plan must be considered. Also, even though the precision of choosing more but smaller segments may be higher, this gain in precision must be weighed against the costs of sample selection and classification.

REFERENCES

1. Bauer, M.E. and Staff. Crop Identification and Area Estimation over Large Geographic Areas Using Landsat MSS Data. LARS Technical Report 012477. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

2. Cochran, W.G. Sampling Techniques. Second Edition. John Wiley and Sons, Inc., New York, 1963, p. 96.

3. LACIE-00200, Volume IV: Crop Assessment Subsystem (CAS) Requirements, Level III Baseline (Reference CCBD #III-0001, Dec. 16, 1974), pp. B1-B6.

4. MacDonald, R.B., F.G. Hall, and R.B. Erb. "The Use of Landsat Data in a Large Area Crop Inventory Experiment (LACIE)." Proceedings, Symposium on Machine Processing of Remotely Sensed Data, June 3-5, 1975, Purdue University, West Lafayette, Indiana, pp. 1B-10 through 1B-11. (IEEE Catalog No. 75 CH 1009-9-C).