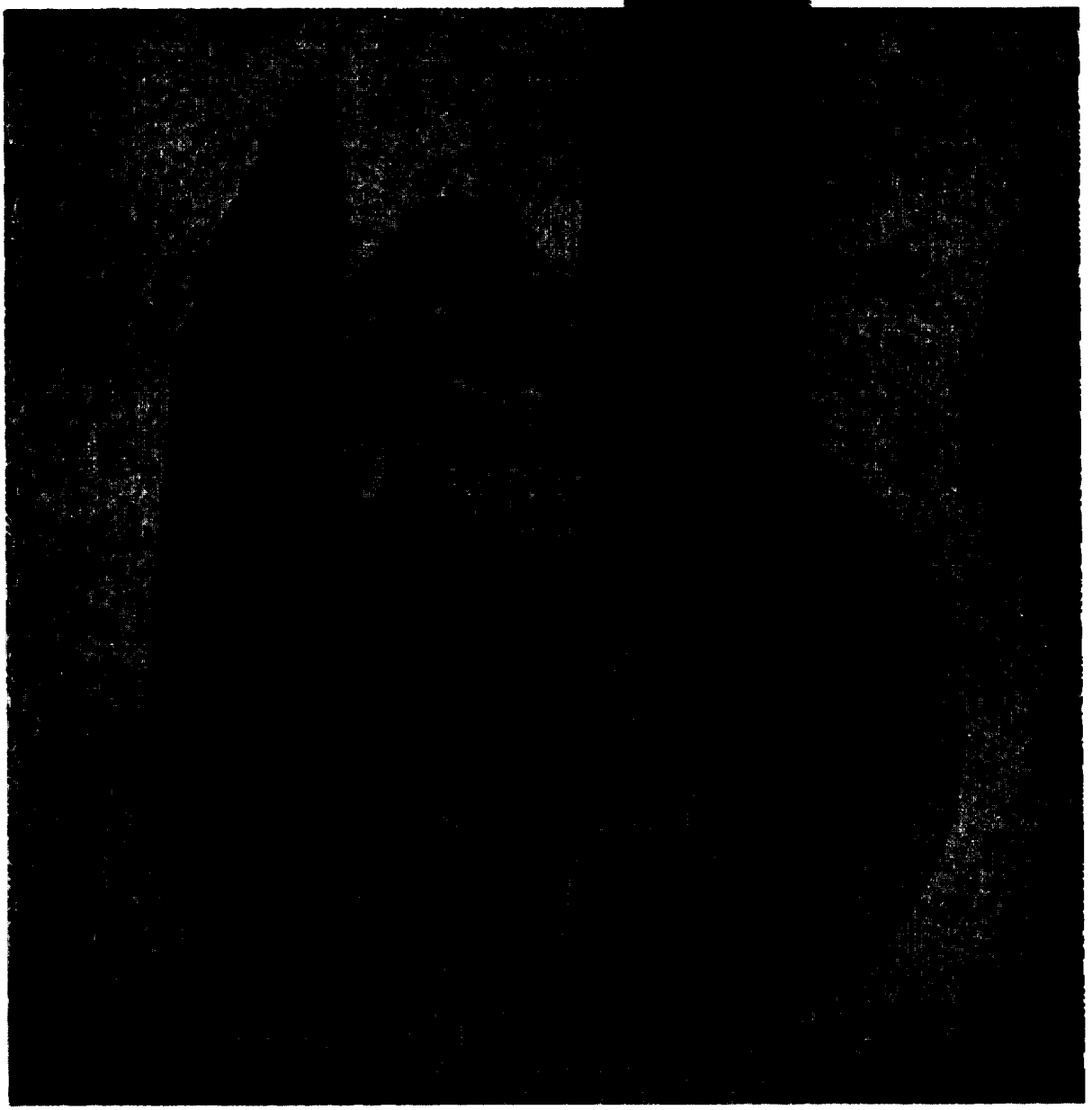
Obtaining Timely Crop Area Estimates Using Ground-Gathered and LANDSAT Data

George Hanuschak, Richard Sigman,
Michael Craig, Martin Ozga,
Raymond Luebbe, Paul Cook,
David Kleweno, and Charles Miller



United States
Department of
Agriculture

Economics,
Statistics, and
Cooperatives Service

Technical
Bulletin
No. 1609

OBTAINING TIMELY CROP AREA ESTIMATES USING GROUND-GATHERED AND LANDSAT DATA. By George Hanuschak, Richard Sigman, Michael Craig, Martin Ozga, Raymond Luebbe, Paul Cook, David Kleweno, and Charles Miller. Statistical Research Division; Economics Statistics, and Cooperatives Service; U.S. Department of Agriculture. Technical Bulletin No. 1609.

ABSTRACT

The report describes how NASA earth resources monitoring satellites, LANDSAT II and III, were used with conventional ground-gathered data to estimate planted crop areas for the 1978 Iowa corn and soybean crops. Estimates that used LANDSAT data and ground data jointly were substantially more precise than those made from ground data alone. These estimates were one of several data sources used in determining the official year-end Annual Crop Summary for Iowa issued January 16, 1979, by USDA's Crop Reporting Board. Problems associated with total project cost, timely delivery of LANDSAT data to the USDA, and cloud cover must be solved prior to any planning for an operational program.

Key words: Satellite, LANDSAT, crop areas, Crop Reporting Board, NASA, cloud cover, regression estimate, Iowa

ACKNOWLEDGMENTS

The strong support of several members from the following groups made this project a reality:

1. New Techniques Section, Statistical Research Division, Economics, Statistics and Cooperatives Service (ESCS), U.S. Department of Agriculture (USDA).

2. Iowa State Statistical Office and Enumerative Staff, ESCS, USDA.

3. Data Collection Branch, Survey Division, ESCS, USDA.

4. Systems Branch, Survey Division, ESCS, USDA.

5. Methods Staff, Estimates Division, ESCS, USDA.

6. National Aeronautics and Space Administration (NASA) - Goddard Space Flight Center.

7. Institute for Advanced Computation, Sunnyvale, Calif.

8. Bolt, Beranek, and Newman, Data Processing Facility in Cambridge, Mass.

The authors wish to extend a special thanks to the participants for their hard work and contributions. Thanks also goes to Charles Caudill, Galen Hart, Harold Huddleston, and William Wigton, ESCS, for their overall management and technical support; Kathy Barr, ESCS, for her ground data management; and Tricia Brookman and Kathy Whyte for their fine typing efforts.

Washington, D.C. 20250 August 1979

SUMMARY

Crop area estimates using NASA's LANDSAT satellites and USDA ground-gathered data were developed for the USDA's 1978 Annual Crop Summary. These estimates of Iowa's 1978 planted crop areas for corn and soybeans had smaller sampling errors than conventional estimates.

The statistical methodology used was a regression estimator. Estimates were developed at the State, multicounty (analysis district), and individual county levels. At the State and multicounty level, the estimates for the regression estimate--using LANDSAT and ground-gathered data--were substantially more precise than the direct expansion estimate, which used ground data only.

Significant gains in time and cost efficiency were realized in the Iowa project. Improvements were made in all phases of the LANDSAT data processing. Problems with total project cost, delivery of LANDSAT data to ESCS in time for analysis, and with cloud cover, however, remain.

CONTENTS

Obtaining Timely Crop Area Estimates Using Ground-Gathered and LANDSAT Data

George Hanuschak, Richard Sigman, Michael Craig, Martin Ozga,

Raymond Luebbe, Paul Cook, David Kleweno, and Charles Miller

INTRODUCTION

One function of the Economics, Statistics, and Cooperatives Service (ESCS), USDA, is to estimate the size of crop areas planted at national and State levels. These estimates are published by ESCS's Crop Reporting Board starting on June 30 of the crop year. Estimates are updated monthly through mid-January when final national and State estimates are made for the crop year. Estimates for individual counties (or in some States for multicounty areas, called Crop Reporting Districts) are made by ESCS's State Statistical Offices (SSO) in cooperation with State government agricultural agencies. Small area estimates, however, are often not published until April of the year following the crop year.

This paper describes efforts by ESCS to develop timely crop area estimates for the 1978 Iowa corn and soybean crops, using LANDSAT and ground-gathered survey data. The LANDSAT II and III used are satellites with earth resource monitoring instruments that measure energy reflected and emitted from the earth's surface. The estimates obtained by this method proved considerably more precise than those that used only ground data.

This was not the first study for which the LANDSAT satellites were used. During 1972-77, ESCS investigated the ability of LANDSATS I, II, and III to improve crop area estimates at State, multicounty, and individual county levels. The results from these studies were mixed. While the precision of winter wheat crop area estimates improved substantially, results for corn and soybeans improved only in a subset of LANDSAT investigation areas. These previous research studies took over a year, on the average, to complete.

For the 1978 study, however, ESCS strove to develop timely LANDSAT-based crop area estimates to supplement current area survey estimates, used in the 1978 Annual Crop Summary for Iowa released by the Crop Reporting Board on January 16, 1979. These estimates were also used by the Iowa SSO in making multicounty estimates.

This report, intended for those with some knowledge of remote sensing applications, will be useful to researchers considering the use of LANDSAT data in estimating crop areas.

DATA SOURCES

The two major sources of data necessary for the Iowa Project were LANDSAT II and III data and ground-gathered survey data.

Ground Data Acquisition and Processing

In late May and early June each year, ESCS conducts a nationwide agricultural survey known as the June Enumerative Survey (JES). The JES uses area frame sampling. Areas of land called segments are selected through stratified random sampling (see appendix A). The land use strata are primarily based on the percentage of land area cultivated.

Nationally, the JES consists of approximately 16,000 sample segments which make up about 0.5 percent of the total U.S. land area. These segments are typically 2.59 square kilometers in size; there were 298 such segments in the Iowa sample in 1978. The crop or land use is recorded for all land area within each segment. Interviewers identify farm operators in these segments using an aerial photograph at a scale of 8" = 1 mile and delineate each farmer's fields. The size of a field as well as a crop or land use is recorded on a questionnaire (fig. 1) for each field inside the segment. The questionnaire data in Iowa is recorded, keypunched, and edited at the individual farm level by Iowa Crop and Livestock Reporting Service personnel. Data are then transmitted to the Washington Computer Center (WCC) via Control Data Corporations INFONET Systems.

In the 1978 LANDSAT studies, each field was specially edited to ensure accurate field boundary locations, using both the photo and questionnaire data. After the JES data were edited by ESCS's Statistical Research Division personnel, a computer tape with all ground data information was sent to the Bolt, Beranek, and Newman (BBN) data processing facility in Cambridge, Mass. Not all fields had been planted at the time of the survey, however, thus, a followup survey was conducted from July 21 to August 1. The followup survey questionnaire (fig. 2) and aerial photography were used to determine the land cover for any fields not planted at the time of the JES. The followup survey was then used to update the ground data computer files at BBN.

For the LANDSAT projects, the fields in the JES segments had to be located on the LANDSAT imagery. To facilitate this process, the field boundaries were recorded on computer data files in latitude-longitude coordinates by a process called digitization. This process began in mid-July and ended in mid-September for the Iowa project. Digitization involves several procedures. First, the aerial photograph and 7-1/2-minute or 15-minute U.S. Geological Survey (USGS) maps are mounted on a 1.27-meter by 1.51-meter digitizing data tablet. Common points such as road intersections are then found on both the photo and the map. These points are used to establish a relationship between the photo and the map. Next, all the field boundaries are then transformed to the latitude-longitude coordinates of the map. Average time to calibrate and digitize a segment was 1 hour. A digitized segment is displayed in figure 3.

Since crop area estimation is usually done within a land use stratum, the strata boundaries also had to be located on the LANDSAT imagery. The land use strata boundaries were digitized directly from county highway maps. This process began in mid-January and was completed in May for the 99 counties in Iowa. Average time to digitize an individual county was 1 day.

Figure 1—1978 JES Questionnaire Crop Section A

SECTION A – ACREAGES OF FIELDS AND CROPS INSIDE BLUE TRACT BOUNDARY

How many acres are **inside** this blue tract boundary drawn on the photo *(or map)*? . . . . . **Acres** [ ]

Now I would like to ask about each field inside this blue tract boundary and its use in 1978.

FIELD NUMBER	827 1	827 2	827 3	827 4
1. TOTAL ACRES IN FIELD	828 .	828 .	828 .	828 .
2. CROP OR LAND USE *(Specify)*				
3a. WOODS, WASTE, IDLE LAND, ROADS, DITCHES, ETC. *(Less than 5.0 acres)*	829 .	829 .	829 .	829 .
3b. WASTE, IDLE LAND ROADS, DITCHES, ETC *(5.0 acres or more)*	830 .	830 .	830 .	830 .
3c. WOODS, *(Including grazed wood land)* *(5.0 acres or more)*	831 .	831 .	831 .	831 .
4. OCCUPIED FARMSTEAD OR DWELLING	843 .	843 .	843 .	843 .
5. TWO CROPS PLANTED IN THIS FIELD for harvest this year or two uses of the same crop?	NO ☐ YES _____ 844 .	NO ☐ YES _____ 844 .	NO ☐ YES _____ 844 .	NO ☐ YES _____ 844 .
6. ACRES LEFT TO BE PLANTED?	61_ .	61_ .	61_ .	61_ .
8. PASTURE	842 .	842 .	842 .	842 .
11. WINTER WHEAT — Planted	540 .	540 .	540 .	540 .
12. WINTER WHEAT — For Grain	541 .	541 .	541 .	541 .
13. RYE — Planted and to be planted	547 .	547 .	547 .	547 .
14. RYE — For Grain	548 .	548 .	548 .	548 .
15. OATS — Planted and to be planted	533 .	533 .	533 .	533 .
16. OATS — For Grain	534 .	534 .	534 .	534 .
19. CORN — Planted and to be planted	530 .	530 .	530 .	530 .
20. CORN — For Grain	531 .	531 .	531 .	531 .
21. SORGHUM *(Excl. crosses)* — Planted and to be planted	570 .	570 .	570 .	570 .
22. SORGHUM *(Excl. crosses)* — For Grain	571 .	571 .	571 .	571 .
23. OTHER USES OF GRAINS PLANTED . Use				
Acres abandoned, cut for hay, silage, etc. Acres	.	.	.	.
24. HAY Cut and to be cut — ALFALFA and ALFALFA MIXTURES	653 .	653 .	653 .	653 .
25. HAY Cut and to be cut — OTHER HAY Kind				
Acres	65_ .	65_ .	65_ .	65_ .
26. SOYBEANS — Planted and to be planted	600 .	600 .	600 .	600 .
36. POTATOES — Planted and to be planted	552 .	552 .	552 .	552 .
38. OTHER CROPS — Acres planted or in use	--- .	--- .	--- .	--- .

Figure 2—Follow-up Survey Questionnaire

1978 Iowa (July) Follow-up Survey of

June Enumerative Intention Fields

At the time of the June Enumerative Survey a few months ago, a few fields were identified as not yet having been planted. For various reasons, the farmer's early intentions may not have been realized. For this reason it is necessary that there be a follow-up of the June Enumerative intention fields. Below is a listing of all of the fields in segment _____ which were not actually planted at the time of the JES interview. For each field review the ASCS segment photo and indicate whether the field boundaries are accurate as drawn. Make any corrections according to the enumerator instructions. Also, verify the JES field acreages and crop types as recorded in columns 3 and 4 for each field.

NOTE: Field information should be verified primarily from field observations. Personal interviews should only be conducted when absolutely necessary.

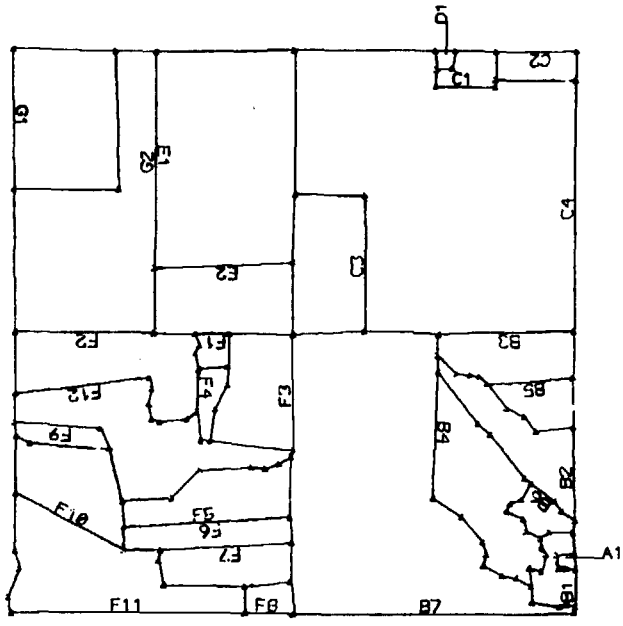
1	2	3	4	5	6	7	8	9
JES TRACT	JES Field Number	JES Field Acreage	JES Indicated Crop Type	Are Photo Boundaries Correct? [1]	JES Field Acreage & Crop Type Correct? [2]	JES Acreage (ACRES)	Land-Use or Crop Type (SPECIFY)	NOTES
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			
				__YES --→ __NO	__YES __NO --→			

[1] If not, indicate new boundaries with green dashed lines. Do NOT erase JES boundaries.

[2] If yes, proceed to next field. If no, record correct data for acreage & crop type.

Figure 3

Plot of Digitized Segment



LANDSAT Data Acquisition and Processing

LANDSAT is an earth resources monitoring satellite in a sun-synchronous polar orbit at an altitude of approximately 570 miles. A multispectral scanner onboard LANDSAT measures energy reflected and emitted in four bands of the electromagnetic spectrum for each .45 hectare (pixel) on the earth's surface. A further description of LANDSAT data can be found in appendix B.

Twelve LANDSAT scenes were required to cover most of Iowa. The LANDSAT scene covering the northwest corner of Iowa was not analyzed because it showed only 200 kilometers not covered by LANDSAT scenes further to the east--an amount less than 0.2 percent of the total area of the State. The location of the 12 LANDSAT scenes can be seen in figure 4.

Based on ESCS's previous LANDSAT analysis experience in Illinois (6) [1/] and on the 1978 planting times, researchers sought LANDSAT imagery especially for early to mid-August. However, due to problems with cloud cover, image dates ranged from August 6 to September 4, 1978 (table 1).

Attempts to obtain cloud-free imagery were not successful. For path 29, row 31, both August 18 and September 5, were cloud free. However, the August 18 image was of poor quality, while the September 5 image was not delivered to ESCS by December 15 in

1/ Underscored numbers in parentheses refer to references listed at the end of this report.

Figure 4

LANDSAT Scene Locations

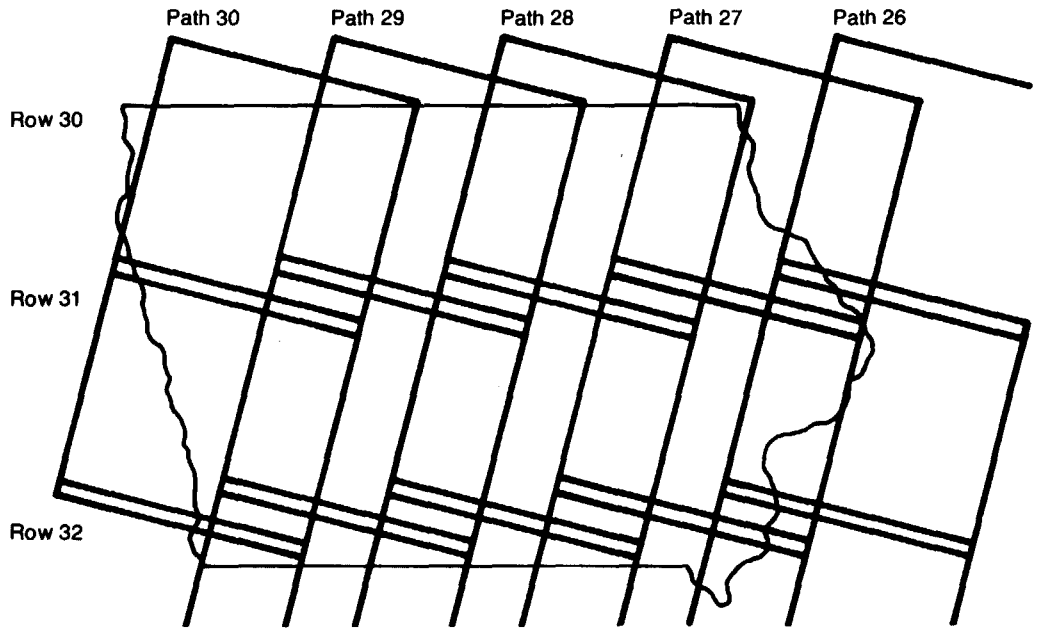


Table 1--Dates of LANDSAT imagery, Iowa project, 1978

Path	Row	Date	Percentage Iowa cloud cover	Scene ID
30	30	August 19	0	30167-16274
	31	August 19	0	30167-16280
29	30	August 9	0	21295-16013
	31	August 9	40	21295-16020
	32	August 18	0	30166-16224
28	30	September 4	60	30183-16162
	31	September 4	0	30183-16164
	32	September 4	0	30183-16171
27	30	August 7	10	21293-15500
	31	August 7	15	21293-15502
	32	August 7	10	21293-15505
26	31	August 6	0	21292-15444

time for it to be registered and analyzed by December 31. Consequently, the partially cloud-covered August 9 scene was registered for path 29, row 31. Path 27 on August 16 was cloud free, but this imagery was never received by ESCS. Thus, partially cloud-covered imagery for August 7 was used for path 27.

Due to various dates of the Iowa LANDSAT imagery, associated cloud-cover problems, and the different times at which ESCS received LANDSAT data, Iowa was partitioned into 10 separate areas, called analysis districts (fig. 5). The smallest analysis district, number 2C, contained three counties; the largest district, number 1, had 20. Analysis district 3A consisted of the 13 cloud-covered counties.

A number of analysis districts, such as 3B, 3C, and 3D, have the same image date. Separate analysis districts were formed in such cases instead of a single large one because the LANDSAT data were received by ESCS for the separate areas at different times. To save time, analysis districts were formed when data were received, instead of waiting until all data for a given image date were on hand.

For each LANDSAT scene used in crop area estimation, three major processing activities transpired from time of satellite overpass to completion of crop area estimates These were:

1. NASA delivery of LANDSAT data products to ESCS,

2. LANDSAT tape reformatting and scene registration, and

3. LANDSAT data analysis and calculation of crop area estimates.

Figure 6 displays the beginning and ending dates for the LANDSAT processing activities by analysis district. The first analysis district completed was 2A on October 26; the last 2B, was completed December 21. Data delivery averaged the longest and was the most variable in duration of the three processing activities (table 2).

Figure 5

Iowa Analysis Districts

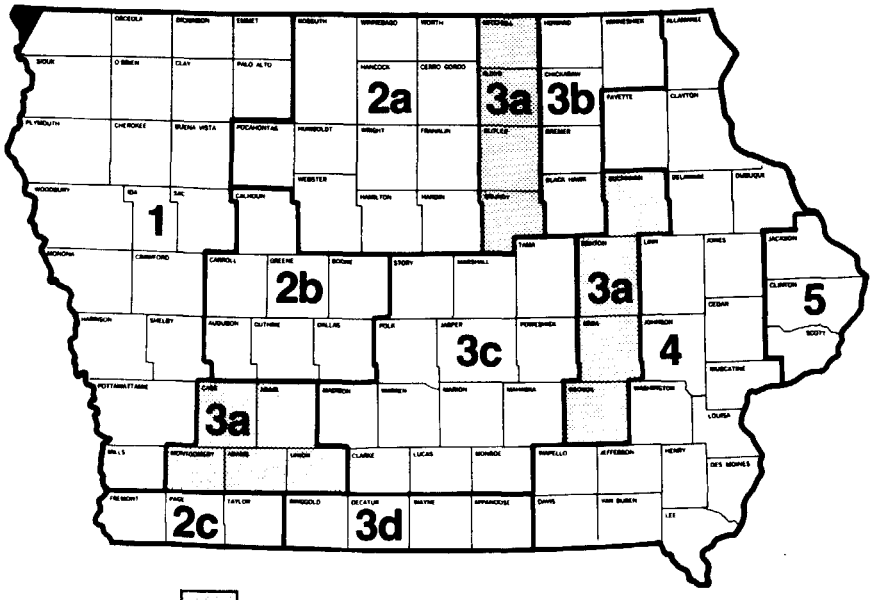
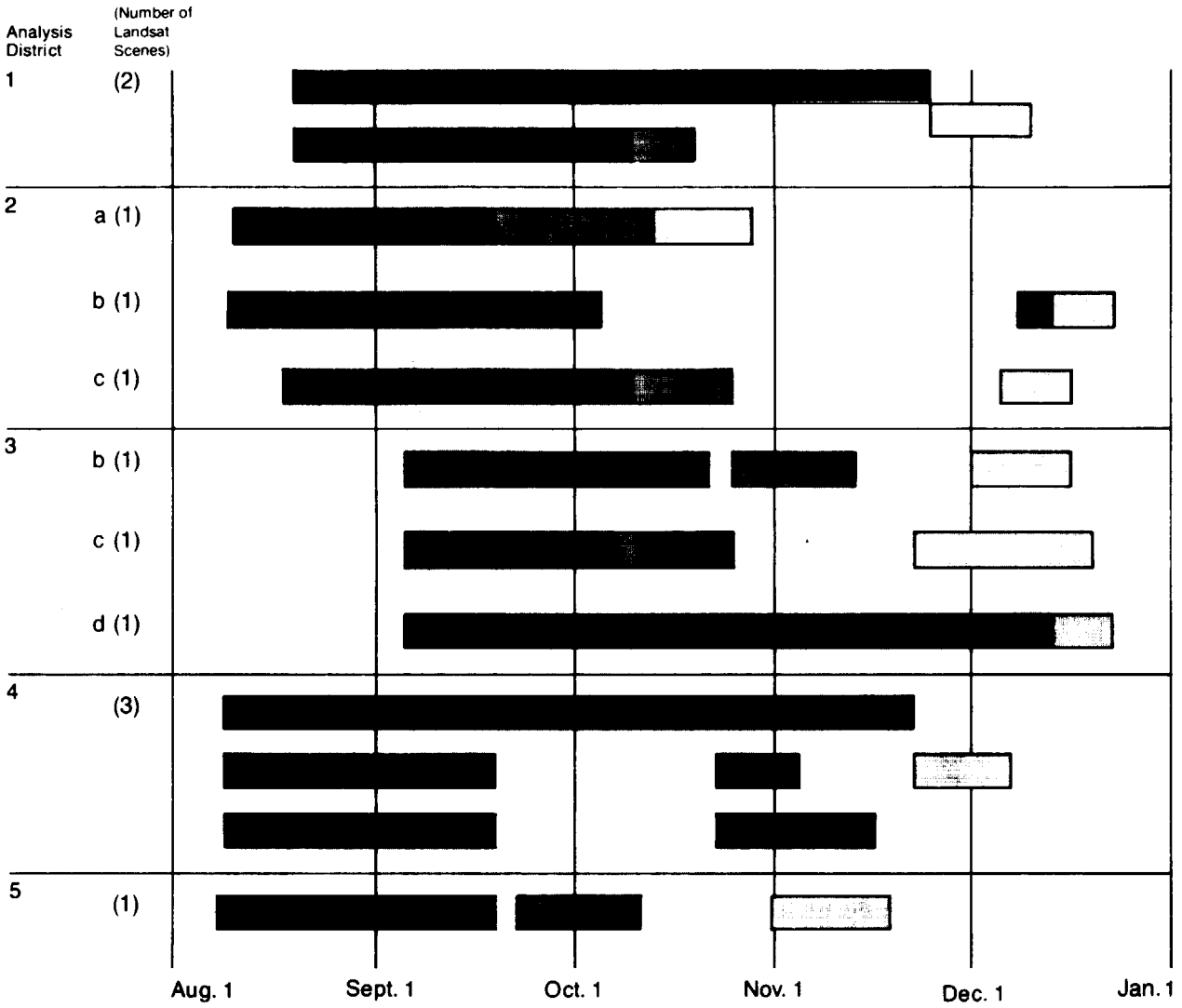
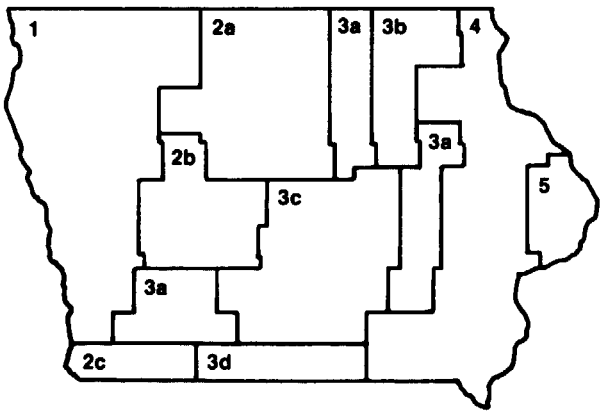


Figure 6

Time Required for Major Project Activities





Data delivery

Bad initial tape
Date tape reordered

Reformatting and registration

Analysis and estimation

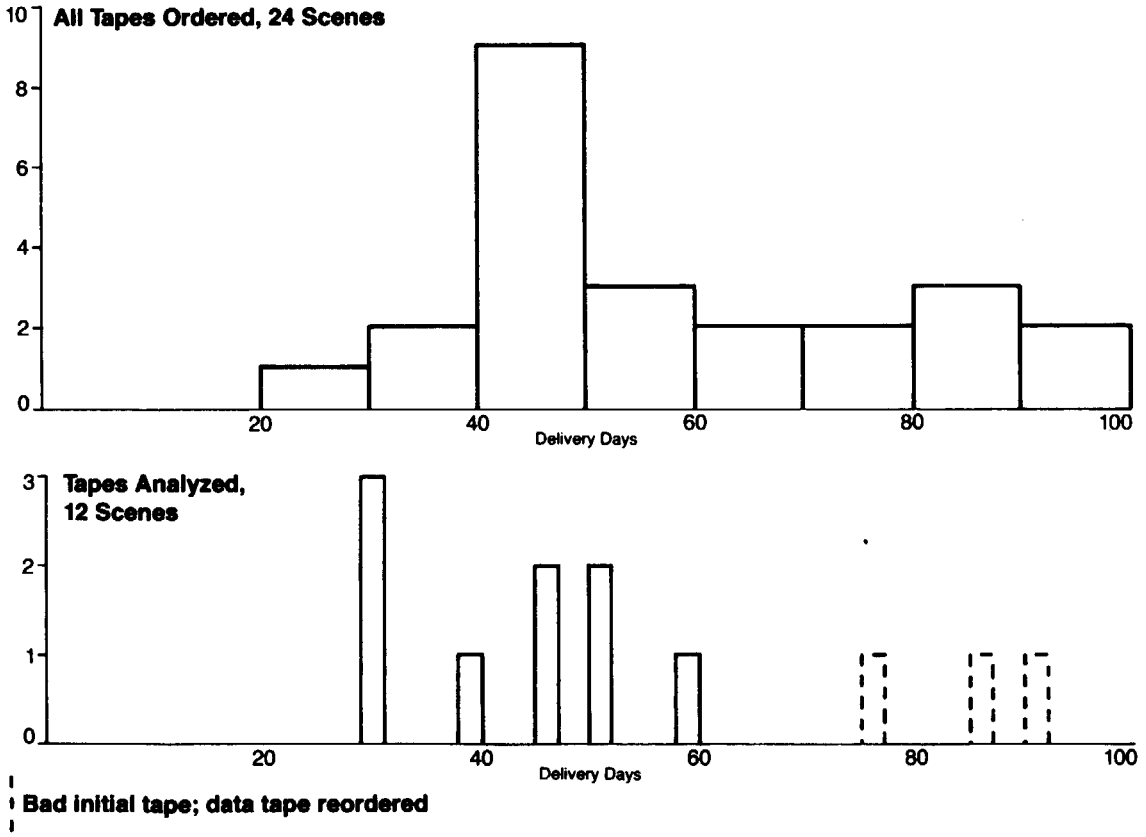
Table 2--Time required for major project activities

Activities	Duration			
	Median	Minimum	Maximum	Quartiles
	Days			
Data Delivery	49	32	93	37, 66
Reformatting, registration	16	4	25	8, 20
Analysis, estimation	13.5	7	26	10, 18

By examining Geo-Synchronous Orbiting Earth Satellite (GOES) satellite weather photos daily, ESCS was able to select candidate cloud-free LANDSAT scenes 2 days after a LANDSAT overpass. LANDSAT computer compatible data tapes and 1:1,000,000 black and white transparencies were supplied to ESCS by NASA's Goddard Space Flight Center (GSFC). Twenty-four tapes were ordered from GSFC, 12 of which were registered for the calculation of crop area estimates. A histogram of delivery times (that is, time from date of satellite overpass to receipt by ESCS) for the 24 tapes is shown in figure 7. Figure 7 also displays the tape delivery times for the 12 scenes which were registered.

Figure 7

LANDSAT Tape Delivery Times, 24 and 12 Scenes



DATA PROCESSING SYSTEMS HARDWARE

ESCS purchases computer time on the following types of computers:

1. A PDP10 in Cambridge, Mass.(the BBN facility) used for interactive processing such as photo and map digitization, LANDSAT analysis for sample segments, and calculation of crop area estimates.

2. An IBM 370-168 at the USDA's Washington Computer Center (WCC) used for computer editing of ground truth data, reformatting LANDSAT tapes, and batch printing of grayscales (for costs see fig. 8).

3. The Illiac IV computer in Sunnyvale, Calif., used by ESCS for clustering and wall-to-wall classification of LANDSAT scenes.

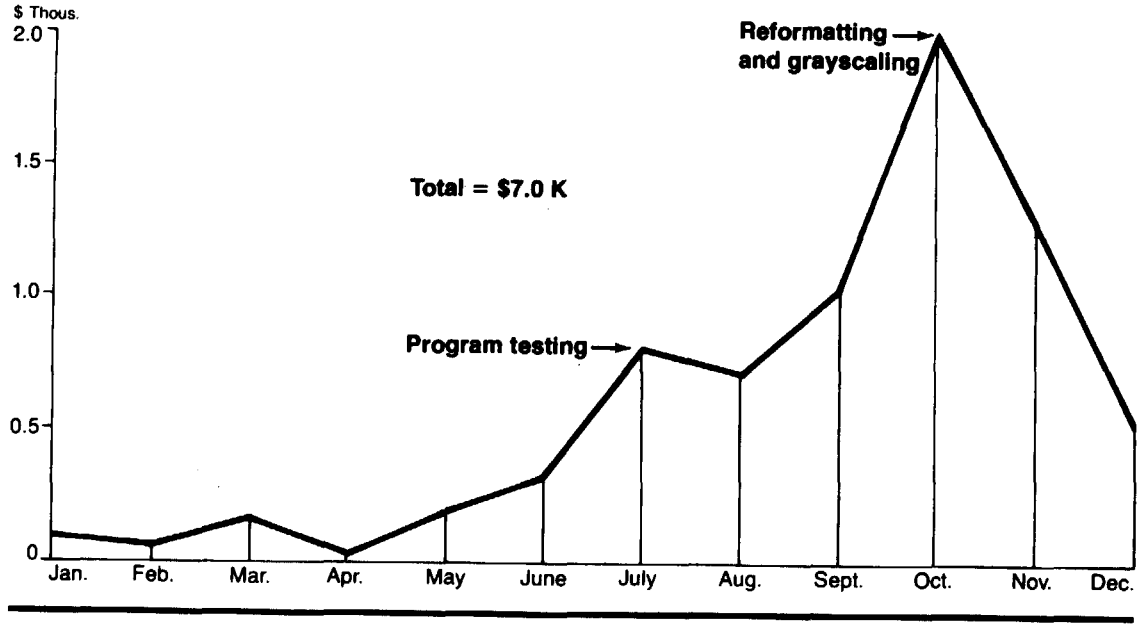
For electronic data transmission, ESCS uses Computer Science Corporation's INFONET data network and the Department of Defense's ARPANET computer network. Additional pieces of hardware used by ESCS for LANDSAT data analysis include:

- two digitizer tablets,
- zoom transferscope,
- terminal plotter with controller,
- leased phone line with multiplexor, and
- 15 KSR (keyboard send-receive) terminals of various types

The total purchase price of this equipment is approximately $90,000.

Figure 8

Washington Computer Center



Total IBM 370-168 computer charges for the Iowa project were $7,000, including usage for computer program testing. PDP10 computer usage for the Iowa project (including usage for development and testing of associated computer programs) was approximately $69,000. Central Processing Unit hours are shown in figure 9. Total ILLIAC IV computer charges for the Iowa project were $25,000.

SOFTWARE AND DATA MANAGEMENT

LANDSAT data analysis for Iowa was done using the EDITOR software system (9), with the exception of reformatting tapes and some of the grayscale printing for registration (see appendix C). For the Iowa project, EDITOR was not changed in any substantial or basic manner, but a number of improvements were made to facilitate its use.

The overall flow of data for the Iowa project was as follows:

1. Ground-truth data were keypunched in Des Moines, Iowa, and transmitted via INFONET to WCC in Washington, D.C.

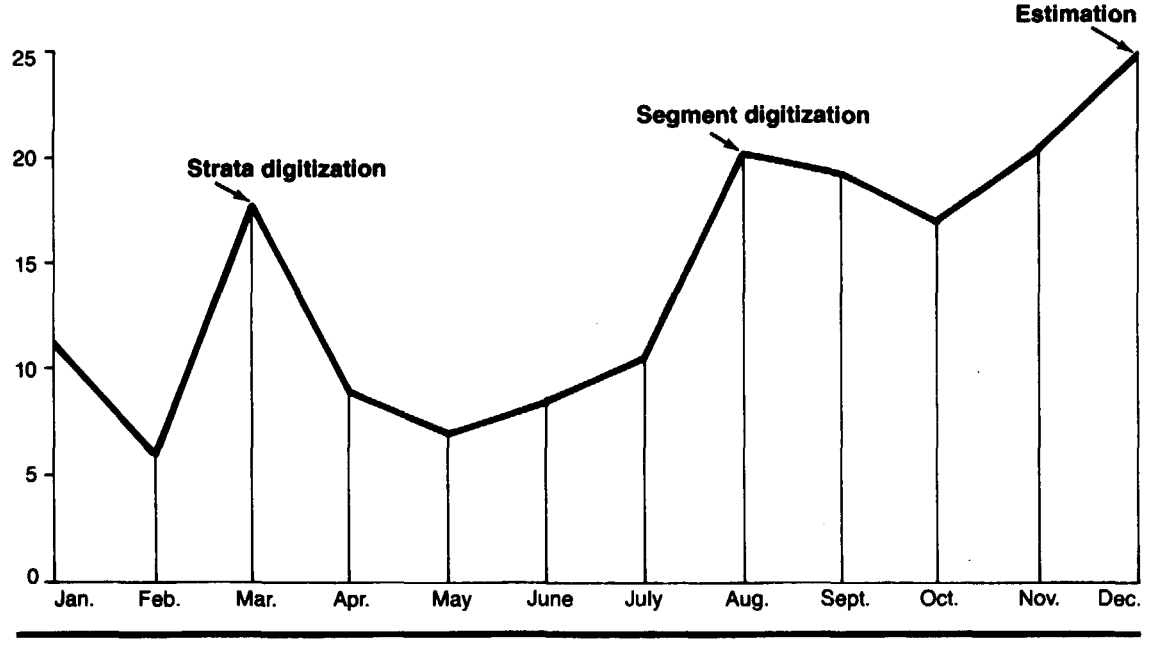
2. Ground-truth data were edited in Washington, D.C., and a ground-truth tape mailed to BBN in Cambridge, Mass.

3. LANDSAT tapes from NASA's Goddard Space Flight Center in Greenbelt, Md., were reformatted and tapes mailed to Cambridge, Mass., and Sunnyvale, Calif.

4. The PDP10 in Cambridge, Mass., was accessed via ARPANET or leased line for interactive processing of LANDSAT data of sample segments for developing crop/land use classification parameters.

Figure 9

Bolt, Beranek, and Newman CPU Hours



5. Classification parameters were transmitted to Sunnyvale, Calif., via ARPANET for wall-to-wall LANDSAT scene classification.

6. Aggregated ILLIAC IV classification results were transmitted back to Cambridge over ARPANET for interactive crop-estimate calculations.

The total project cost was estimated at $300,000.

LANDSAT DATA REGISTRATION

To make effective use of the LANDSAT data, one must know quite accurately the geographic location of each resolution element (pixel) in a given scene. This process of relating the LANDSAT row-column coordinates with map latitude-longitude coordinates by means of appropriate mathematical equations is known as registration. In the course of working with LANDSAT data, the method of registration has been refined. Its major components are presented in the following outline.

Certain materials and equipment are presumed available. These include: 1) black and white positive transparencies at 1:1,000,000 scale for bands 5 and 7 of the LANDSAT scene considered, 2) the USGS index maps of the State(s) covered as well as all 7-1/2-minute, 15-minute, and 2-degree USGS maps listed for the area, 3) a coordinate digitizer with 0.001-inch resolution, 4) a teletype (TTY) compatible terminal for connection to BBN and, 5) a digital LANDSAT tape prepared for use at BBN.

One would then proceed in the following manner:

1. Select control points from the 1:1,000,000 LANDSAT transparency. The locations of points are chosen near the intersections of 5 x 5, 6 x 6 or 7 x 7 grid. The selected features are road and/or rail intersections, small lakes, and other time invariant topographic features.

2. Overlay the transparency on a mosaic of USGS index maps and digitize the control points. This produces an output file containing the row-column and latitude-longitude coordinates for each point as well as a map index for storage and retrieval of USGS maps.

3. Print grayscales for all control points. Grayscales are replicates of portions of the LANDSAT scene, obtained from a printer. Each pixel is represented by one print character and either special characters or overprinting is used. The printing was done at 10 characters per inch on a given line and 8 lines per inch to produce an image with 1:24,000 scale approximately.

4. Determine corresponding points on grayscales and USGS topographic maps. The corresponding points were selected either by overlaying grayscales and maps or by visual identification of features. The grayscales could be overlayed directly on 7-1/2-minute maps. Overlaying was done for 15-minute maps and 2-degree maps by reduction xeroxing or with a zoom transfer scope. The points determined by this step are known as the precision control points.

5. Enter the selected points from both the maps and grayscales into a file. This is done by digitizing the map points while entering the row-column values from the grayscales.

6. Predict row and column coordinates from a third order polynomial of latitude and longitude by means of regression. The third order polynomial is used to determine the accuracy of the chosen points and to make corrections as needed. The polynomial coefficients are output as the final precision calibration file to be used in JES sample segment and land use strata location.

This procedure has been refined to the point that at a complete State level, we can expect the tape reformatting and the registration accuracy to be within one pixel. Examples for the Iowa project are included in table 3. For the 12 Iowa scenes, approximately one week per scene was required to complete the procedure.

To determine labeled pixels for classifier training, each segment must be accurate to one-half pixel or better. This procedure follows:

1. At the scale of LANDSAT grayscales (approximately 1/24,000), plots showing field boundaries were obtained for each segment.

2. The segment plots were then overlayed on the segment grayscales at the locations predicted by the precision registration polynomial.

3. By examining the grayscale's lightness and darkness patterns corresponding to segment fields in conjunction with the segment photo and USGS map, it was determined whether the segment was correctly located. If not, row and column shifts needed to move the segment to its correct location were determined and used to generate local calibration files.

LANDSAT DATA CLASSIFICATION

The estimated average field size in the Iowa study was 12 hectares for corn and 13 hectares for soybeans, based on the stratified random sample of ground data segments. The number of pure field interior pixels was thus approximately 59 percent of the total pixels for these two major crops.

For each crop or land cover type, "training signatures"[2/] were developed using several methods (see appendix B for a discussion of classifying LANDSAT data into crop types). Methods used were (1) resubstitution, in which all the field interior pixels for the cover type are used; (2) the 1/2 sample partition method, in which the data for 50 percent of the sample segments are used; and (3) a method where small fields (less than 5 hectares) were excluded from the training data. Once the training data for a cover type were established, the use of prior probabilities for a cover type and clustering within a cover type's training data had to be considered. Types of prior probabilities used were those proportional to the reported hectares in the sample segments or equal prior probabilities.

As seen in the variance formula for the regression estimate (see appendix A for the statistical methodology), the variance is at a minimum when the r_h^2 (sample correlation coefficient squared for h = 1, ..., L) are at maximum values. This is the main criterion used to evaluate the precision of the regression estimates. A traditional criterion for evaluating LANDSAT crop/land cover classification is percentage correct measures (table 4). Using corn, for example, percentage correct is the number of pixels known to be corn that are computer labeled as corn divided by the total known number of pixels of corn and converted to a percentage.

The r_h^2 values are also plotted against crop maturity stages in figure 10 for corn and soybeans. As seen from this figure, the r_h^2 for corn using the September 4 imagery were drastically reduced from results of the August imagery.

2/ Training signatures are the 4-band mean vectors and covariance matrices for the various land cover categories (5).

Table 3--LANDSAT data registration accuracy

Path	Row	Scene ID	Control points	Mean Square Errors			Max. residuals		
				Line	Column	Line & column	Line	Column	Line & column
			Number	--Pixels--		Meters	--Pixels--		Meters
30	30	30167-16274	36	.535	.724	59.6	1.33	2.06	129
30	31	30167-16280	33	.442	.616	49.9	0.98	1.70	109
29	30	21295-16013	42	.417	.548	45.3	1.25	1.66	111
29	31	21295-16020	20	.296	.713	46.7	.80	1.42	90
29	32	30166-16224	18	.587	.900	69.8	1.13	2.21	142
28	30	30183-16162	26	.654	1.325	92.6	1.70	3.33	213
28	31	30183-16164	33	1.234	1.010	113.8	3.56	2.87	314
28	32	30183-16171	16	.290	.775	50.4	0.61	1.44	84
27	30	21293-15500	24	.837	1.011	87.5	2.24	2.55	181
27	31	21293-15502	36	.633	.819	68.2	1.51	1.86	142
27	32	21293-15505	23	.492	.655	53.8	.88	1.39	97
26	31	21292-15444	21	.755	.955	80.6	1.47	2.68	181

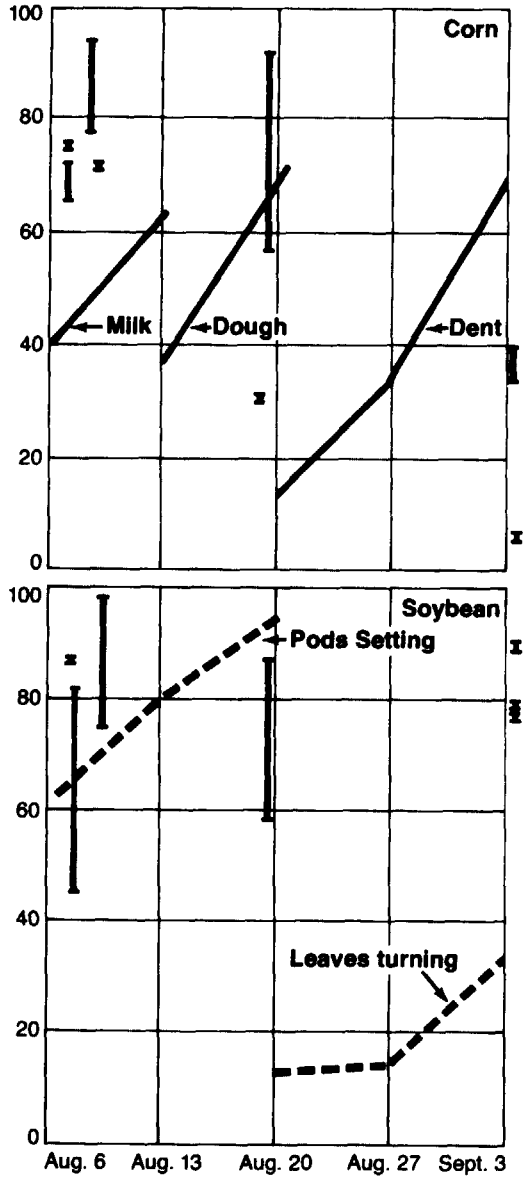
Table 4--Percentage correct classification - corn and soybeans

Analysis district	Corn			Soybeans		
	Using all pixels	Using interior pixels	Range of r^2 1/	Using all pixels	Using interior pixels	Range of r^2 1/
	--Percent--			--Percent--		
1	72.13	79.98	.57-.92	67.34	76.37	.58-.88
2A	81.46	87.03	.71	71.21	79.43	.74
2B	79.59	90.39	.78-.94	71.36	85.14	.74-.98
2C	50.55	63.17	.30	59.63	74.31	.80
3B	77.58	77.41	.38	37.49	44.15	.79
3C	56.57	65.24	.34-.40	59.37	68.97	.77
3D	33.71	51.27	.07	54.93	70.47	.89
4	56.68	60.94	.65-.71	26.52	29.36	.45-.83
5	50.00	54.35	.75	45.23	75.51	.86

1/ Range by land use strata.

Figure 10

Corn and Soybean Stages of Development Versus Sample Coefficients of Determination



I = r^2 x 100 or range of r^2 x 100 based on results by land use strata

——— = Percentage of crop at stage of maturity

▬ ▬ ▪ = Percentage of acreage with 75 percent of plants in fields showing pod setting or leaves turning

CROP AREA ESTIMATES

Crop area estimates for corn and soybeans were developed at the State, multicounty (analysis district), and individual county levels. At the State and multicounty level, improvements in precision for the regression estimate (LANDSAT and ground data) versus the direct expansion estimate (ground data only) were substantial. At the analysis district level, the range of relative efficiencies for corn was 0.93 to 5.98 and soybeans ranged from 2.73 to 7.59. Specific values for all analysis district estimates and their corresponding relative efficiencies are listed in tables 5 and 6. Clouds covered 13 of the 99 counties in Iowa for the available LANDSAT data. Loss of LANDSAT data for portions of a State during the optimum period for crop discrimination due to cloud cover is not unusual. The conventional direct expansion estimate of ground data had to be used for the 13-county area in Iowa (7). Individual county estimates had coefficients of variation (CVs) ranging from 7.1 to 59.9 percent for corn and 9.0 to 100 percent for soybeans. CVs above 20 percent are not suitable for operational data use by ESCS.

The State level estimates were input to USDA's Crop Reporting Board's 1978 Annual Crop Summary for Iowa; the analysis district estimates were input to the Iowa Crop and Livestock Reporting Service's multicounty level estimates. These LANDSAT-based regression estimates, however, were not the sole source of data in determining the State and multicounty estimates. Conventional data sources such as ESCS's June Enumerative Survey, June Acreage Survey, Fall Acreage and Production Survey, as well as USDA export data and other check data were considered.

CONCLUSIONS

The major benefit of LANDSAT regression estimates to ESCS is substantial improvement in precision with no increase in respondent burden associated with ground surveys. The repeatability of such an effort, however, depends crucially on timely delivery of LANDSAT data to ESCS. It is important to note that these estimates must be considerably more precise than those provided by ESCS's efficient June Enumerative Survey to be useful to USDA's

Table 5--Corn area estimates and relative efficiencies

Analysis district	Classified pixels 1/	$\hat{Y}_{DE}$ direct expansion estimate	CV $\hat{Y}_{DE}$	$\hat{Y}_{R}$ LANDSAT regression estimate	CV $\hat{Y}_{R}$	Range of r^2 for h=1,..,L	Relative efficiency
	--Hectares--			Hectares			
1	1,306,217	1,462,074	3.48	1,460,234	2.20	.57-.92	2.51
2A	923,626	828,772	4.47	818,892	2.50	.71	3.28
2B	463,957	332,050	11.50	454,252	3.40	.78-.94	5.98
2C	124,767	106,036	10.98	109,959	9.50	.30	1.24
3A	2/	657,462	4.36	2/	2/	2/	2/
3B	345,293	276,112	10.05	268,022	8.47	.38	1.49
3C	589,898	550,581	7.46	542,081	6.02	.34-.40	1.58
3D	58,843	83,658	17.76	82,798	18.65	.07	0.93
4	1,058,692	1,029,688	6.72	896,084	4.47	.65-.71	2.99
5	132,166	148,148	11.10	149,820	6.03	.75	3.32
State total	5,660,921	3/ 5,525,807	2.3	5,439,604	1.5	.07-.94	2.43

1/ Converted to hectares.

2/ LANDSAT data not available.

3/ This is the JES direct expansion estimate.

Table 6--Soybean area estimates and relative efficiencies

Analysis district	Classified pixels 1/	$\hat{Y}_{DE}$ direct expansion estimate	CV $\hat{Y}_{DE}$	$\hat{Y}_{R}$ LANDSAT regression estimate	CV $\hat{Y}_{R}$	r^2 Range of for h=1 . . , L	Relative efficiency
	--Hectares--			Hectares			
1	760,215	747,759	8.11	781,566	4.04	.58-.88	3.70
2A	650,382	655,049	6.75	675,293	3.42	.74	3.68
2B	244,275	256,944	12.91	255,540	6.11	.74-.98	4.55
2C	93,828	95,196	24.97	97,497	11.67	.80	4.37
3A	2/	401,671	9.20	2/	2/	2/	2/
3B	84,102	86,550	28.00	125,300	9.37	.79	4.26
3C	369,662	328,662	14.51	338,363	7.06	.77	3.98
3D	78,841	82,633	32.55	95,933	10.20	.89	7.59
4	343,162	441,032	12.68	424,782	7.97	.45-.83	2.73
5	34,575	47,060	29.20	48,580	12.53	.86	5.10
State total	3,060,122	3/ 3,205,320	3.91	3,244,525	2.50	.45-.98	2.38

1/ Converted to hectares.

2/ LANDSAT data not available.

3/ This is the JES direct expansion estimate.

Crop Reporting Board. Cloud cover is a serious problem in estimating crop areas at the sub-State level. At the individual county level, the sampling errors associated with the crop area estimates are generally too large to warrant use of the data. Problems with overall project costs and timely acquisition of LANDSAT data remain. Their solutions must precede any official use of LANDSAT data for monthly crop area reports or year-end reports covering large regional areas.

REFERENCES

1. Baker, J. R. and E. M. Mikhail, Geometric Analysis and Restitution of Digital Multispectral Scanner Data Arrays. Laboratory for Application of Remote Sensing, Purdue University, Information Note 052875.

2. Cardenas, M., M. Blanchard and M. Craig, On the Development of Small Area Estimators Using LANDSAT Data as Auxiliary Information. Economics, Statistics, and Cooperatives Service, U.S. Dept. of Agriculture, August 1978.

3. Cochran, William G. Sampling Techniques. Third edition, John Wiley and Sons, 1977.

4. Craig, M., R. Sigman and M. Cardenas, Area Estimates by LANDSAT: Kansas 1976 Winter Wheat. ESCS-Economics, Statistics, and Cooperatives Service, U.S. Dept. of Agriculture, August 1978.

5. Fleming, M. D., J. S. Berkebile, R. Hoffer, Computer Aided Analysis of LANDSAT-1 MSS Data: A Comparison of Three Approaches Including a Modified Clustering Approach. Proceedings of the 1975 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Ind.

6. Gleason, C., R. Starbuck, R. Sigman, G. Hanuschak, M. Craig, P. Cook, and R. Allen, The Auxiliary Use of LANDSAT Data in Estimating Crop Acreages: Results of the 1975 Illinois Crop Acreage Experiment. Statistical Reporting Service, SRS-21, U.S. Dept. of Agriculture, October 1977.

7. Hanuschak, G. LANDSAT Estimation with Cloud Cover. Proceedings of 1976 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Ind.

8. LANDSAT Data Users Handbook. National Aeronautics and Space Administration Document No. 76SDS4258, Goddard Space Flight Center, September 1976.

9. Ozga, M., W. Donovan and C. Gleason, An Interactive System for Agricultural Acreage Estimates Using LANDSAT. Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Ind.

10. Swain, P. H., Pattern Recognition: A Basic for Remote Sensing Data Analysis. LARS information note 111572.

11. Wigton, W., The Technology of LANDSAT Imagery and Its Value in Crop Estimation for the U.S. Dept. of Agriculture, Statistical Reporting Service, March 1976.

Appendix A--Statistical Methodology

This appendix describes the direct expansion estimator (ground data only) and the regression estimator (LANDSAT and ground data jointly) used in this study.

Direct Expansion Estimation (Ground Data Only)

Aerial photography obtained from the Agricultural Stabilization and Conservation Service, USDA is visually interpreted using the percentage of cultivated land to define broad land-use strata. Within each stratum, the total area is divided into N_h elementary area frame units. This collection of area frame units for all strata is called an area sampling frame. A simple random sample of n_h units is drawn within each stratum.

In the general purpose JES survey, area devoted to each crop or land use is recorded for each field in the sampled area frame units or segments. The scope of information collected on this survey is much broader than crop area alone. Items estimated include crop area by intended utilization, grain storage on farms, livestock inventory by various weight categories, agricultural labor and farm economic data. Intensive training of field statisticians and interviewers helps minimize nonsampling errors. The notation used for the stratified random sample in the survey is:

Let h = 1,2,...and L be the land use strata. For a specific crop (corn, for example), total crop area for all purposes and the variance of the total area is estimated as follows:

Let y = total corn area for a state (Iowa, for example)

$\hat{Y}_{DE}$ = estimated total of corn area for a state.

y_{hj} = total area in the j^{th} sample unit in the h^{th} stratum.

Then,

$$\hat{Y}_{DE} = \sum_{h=1}^{L} N_h \left(\sum_{j=1}^{n_h} y_{hj} \right) / n_h$$

The estimated variance of the total is:

$$v(\hat{Y}_{DE}) = \sum_{h=1}^{L} \frac{N_h^2}{n_h (n_h - 1)} \frac{N_h - n_h}{N_h} \cdot \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$$

Note that we have not yet made use of an auxiliary variable such as computer classified LANDSAT pixels. The estimator is commonly called a direct expansion estimate,(3) and we will denote this by $\hat{Y}_{DE}$.

Regression Estimation (Ground Data and Computer Classified LANDSAT Data)

The regression estimator utilizes both ground data and classified LANDSAT pixels. The estimate of the total Y using this estimator (3)is:

$$\hat{Y}_R = \sum_{h=1}^{L} N_h \cdot \bar{y}_h \text{ (reg)}$$

where

$$\bar{y}_h(\text{reg}) = \bar{y}_h + \hat{b}_h(\bar{X}_h - \bar{x}_h)$$

and $\bar{y}_h$ = the average corn area per sample unit from the ground survey for h^{th} land use stratum.

$\hat{b}_h$ = the estimated regression coefficient for the h^{th} land-use stratum when regressing ground-reported corn area on classified pixels for the n_h sample units.

$\bar{X}_h$ = the average number of pixels of corn per frame unit for all frame units in the h^{th} land-use stratum. Thus, entire LANDSAT scenes must be classified to calculate $\bar{X}_h$. Note that this is the mean for the population and not the sample.

X_{hi} = number of pixels classified as corn in the i^{th} area frame unit of the h^{th} stratum.

$\bar{x}_h$ = the average number of pixels of corn per sample unit in the h^{th} land-use stratum.

x_{hj} = number of pixels classified as corn in the j^{th} sample unit in the h^{th} stratum.

The estimated (large sample) variance for the regression estimator is:

$$v(Y_R) = \sum_{h=1}^{L} \frac{N2_h}{n_h} \frac{N_h - n_h}{N_h} \frac{1 - r_h^2}{n_h - 2} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$$

where

r_h^2 = sample coefficient of determination between reported corn area and classified corn pixels in the h^{th} land-use stratum.

$$r_h^2 = \frac{\left[\sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)(x_{hj} - \bar{x}_h)\right]^2}{\left[\sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2\right]\left[\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)^2\right]}$$

Note that

$$v(\hat{Y}_R) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h - 2} (1 - r_h^2)\, v(\hat{Y}_h)$$

and so $\lim v(\hat{Y}_R) = 0$ as $r_h^2 \to 1$ for fixed n_h. Thus a gain in lower variance properties is substantial if the coefficient of determination is large for most strata.

The relative efficiency of the regression estimator compared to the direct expansion estimator will be defined as the ratio of the respective variances:

$$R.E. = v(\hat{Y}_{DE}) \,/\, v(\hat{Y}_R)$$

Since the entire State of Iowa cannot be covered by LANDSAT imagery of the same date, it was necessary to define post-strata (analysis districts) which were wholly contained within a LANDSAT pass or scene. The formulas for the direct expansion estimate and regression estimate hold for post-strata as presented by Gleason (6). The regression estimator is called the separate form of the regression estimator. An alternate form for the regression estimator, called the combined form, is described by Craig (4). Conditions under which use of the combined form are appropriate are discussed by Cochran (3). Several types of estimates have also been developed for individual counties. (2, 6).

Appendix B--Categorization or Classification Procedures for LANDSAT Data

This appendix gives a brief description of LANDSAT data. The statistical procedures of discriminant analysis and clustering also included are as applied to LANDSAT data for the purposes of crop area estimation.

Description of LANDSAT Data (11)

The satellite data used in this report is LANDSAT Multispectral Scanner (MSS) data described in section 3 of the Data User's Handbook (8).

The MSS is a passive electro-optical system that can record radiant energy from the scene being sensed. All energy coming to earth from the sun is either reflected, scattered, or absorbed, and subsequently emitted by objects on earth (1). The total radiance from an object is made up of two components, reflected radiance and emitted radiance. In general, the reflected radiance forms a dominant portion of the total radiance from an object at shorter wavelengths of the electromagnetic spectrum, while the emitted radiance becomes greater at the longer wavelengths. The combination of these two sources of energy would represent the total spectral response of the object. This, then, is the "spectral signature" of an object; it is the difference between such signatures which allows the classification of objects using multivariate statistical techniques. This particular product in system-corrected images refers to products that contain the radiometric and initial spatial corrections introduced during the film conversion. Every picture element (pixel) is recorded with four variables corresponding to one of the four MSS bands.

Appendix table 1--Multi-spectral scanner band relationships

Spectral band number	Wavelengths (micrometers)	Color	Band code
1	.5 - .6	Green	4
2	.6 - .7	Red	5
3	.7 - .8	Near Infrared	6
4	.8 - 1.1	Infrared	7

Discriminant Analysis

This background (11) is intended to enable the reader to understand the detailed computations and results in this report. Discriminant analysis is the process used in attempting to differentiate between two or more populations of interest based on multivariate measurements.

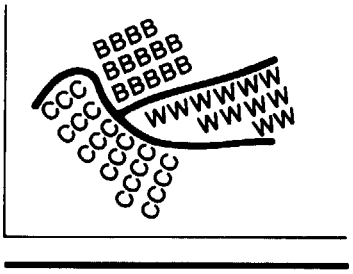
Suppose the land population of interest is a portion of the San Joaquin Valley in California and that cotton, wheat, and barley are the major crops. From every acre in the San Joaquin Valley, we have light intensity readings for green light, red light, and two infrared wavelengths. These light intensities are multivariate measurements that will be used to allot or classify each data point into a crop type such as cotton, wheat, or barley.

A sample of fields from each crop type is selected and their respective light intensities obtained. These sample points are plotted on a two-dimensional graph showing relative positions of each crop in the measurement space (MS). The problem is to partition the MS in some optimal fashion so that points are alloted as nearly correct as possible.

There are many ways to partition a MS. We have done a simple nonstatistical partition above, merely by drawing lines (appendix fig. 1). Visually partitioning the MS may work when it is one or two dimensional but for a more than two-dimensional MS, a visual partition is not possible. For most LANDSAT and aerial photography classification studies, a four-dimensional MS has been used.

Appendix figure 1

Two-Dimensional Measurement Space



The method used in this report was that of constructing contour "surfaces" in the MS. These dividing surfaces were constructed so that points falling on the dividing surface have equal probabilities of being in either group on each side. Those points not on the dividing surface always have a greater probability of being classified into the crop for which the point is interior to the contour surface. If prior knowledge of the population density function indicates that the density is multivariate normal, then a multivariate normal density distribution will be estimated for each crop. It is hoped that the data is approximately multivariate normal, since only the mean vector and covariance matrix is required to estimate a discriminant function. Usually small departures from normality will not invalidate the procedure, but certain types of departures (for example, bimodal data) may be very detrimental to the statistical technique. However, the error rate and estimator properties depend on the assumptions of the distributions and prior information.

A multivariate normal density was assumed in this study so it becomes quite simple to estimate the density functions and the discriminant scores which in turn determine boundaries.

The discriminant score for i^{th} population is:

$$P_i \, (2\pi)^{-\frac{q}{2}} \left| \Sigma_i \right|^{-\frac{1}{2}} \, e^{-\frac{1}{2} (\chi-\mu_i)' \Sigma_i^{-1} (\chi-\mu_i)}$$

where P_i is the prior probability for the i^{th} crop

Σ_i is the covariance matrix (qxq) for the i^{th} crop

μ_i is the mean vector (q length) for the i^{th} crop

χ is the set of measurements of an individual from the i^{th} population or its equivalent discriminant score the $\log_{(e)}$ of

$$D_i = \log_e (P_i) - 1/2 \log_e \left| \Sigma_i \right| - 1/2 \, (\chi-\mu_i)' \, \Sigma_i^{-1} \, (\chi-\mu_i) \, .$$

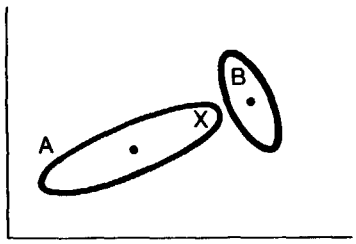
The boundary between two populations is quadratic (curved) and the point χ that falls in the boundary has an equal probability of being in either population.

When an unknown land point is classified, its measurement vector is compared to the mean vector for each crop represented. The point is assigned to the crop whose mean is "nearest" from a statistical point.

The procedure used for finding the nearest mean uses the Mahalanobis measure of distance not the Euclidean (see appendix fig. 2).

Appendix figure 2

Measurement Space Showing Two Crop Density Functions and an Unknown Point (x)



The point is actually closest (Euclidean distance) to the mean vector (center point) of B. However, when one takes into account the variance and covariances, χ is found to be closest to Group A based on a probability concept and an outlier of Group B. Therefore, the point would be classified into Group A, because the probability that the point (χ) is a member of Group A is much greater than for Group B.

Thus, the MS is partioned by computing the means for each crop type and using the Mahalanobis distances from this mean. This distance depends on the covariance matrix and is a measure of probability. The discriminant functions without prior probabilities are:

(1) $(X - \bar{X}_i)' S_i^{-1} (X - \bar{X}_i)$ which is a sample estimate of $(X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)$ if linear discriminant functions are used, and

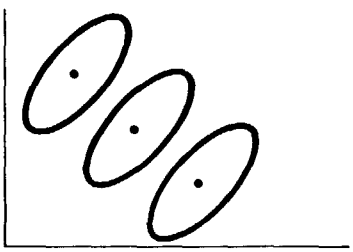
(2) $-1/2 \log_e S_i - 1/2 (X - \bar{X}_i)' S_i^{-1} (X - \bar{X}_i)$ if quadratic discriminant functions are used. These functions involve the exponent of the density formula of the multivariate normal distribution

$$C_{exp} = -1/2 (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)$$

of the i'th crop. If $\Sigma_i = \Sigma_j$ for all $i \neq j$, linear discriminant functions are used.

Appendix figure 3

Measurement Space Where Crop Types Have Same Covariance Matrix and Slope



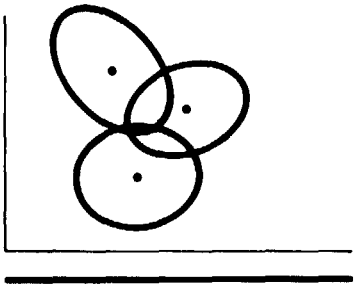
It is worth pointing out that if linear discriminant functions are used one assumes (1) that $\Sigma_i = \Sigma_j$, (2) that for all crops in the MS the major and minor axes are equal, and (3) the sample data for each crop has the same slope. Such an event in two-space is shown in appendix fig. 3.

This space can be partitioned effectively with straight lines. Thus, we can use linear discriminant functions.

Appendix fig. 4 shows a MS where covariance matrices are not equal, and therefore, linear discriminant functions are not appropriate. In either case, the Mahalanobis distance is used.

Appendix figure 4

Measurement Space When Crops Have Different Covariance Matrices



In appendix fig. 3, even though a common center point is not present, a common covariance (ellipse) matrix would be computed. In appendix fig. 4, a different covariance matrix will be needed for each crop type. When the off-diagonal elements in the covariance matrix are unequal, the slopes of the data are different and linear discriminant functions are not appropriate.

The above techniques follow from our first assumption that the data is normally distributed in the MS. In practice, however, one does not decide what the distribution of the population density is in the MS and program the correct procedure. One uses the available procedures for analyzing data instead. Most available programs assume multivariate normal data because the program and the calculations are greatly simplified.

In order to better explain how a parametric procedure can reduce the workload, consider that the first step in the discriminant analysis (DA) is to estimate the population density function in the MS, with a sample of points from each crop. Once these population density functions have been estimated, partitioning the space is extremely simple.

To estimate a multivariate population density in MS for cotton where we have no prior information except sample data on cotton is extremely difficult. If a sample of 1,000 points were available, each of these 1,000 data points would need to be stored in the computer. On the other hand, if we are working a multidimensional normal distribution, theory tells us that the sufficient statistics are computed (mean vector and covariance matrix) and stored in the computer.

The individual data points could be discarded because no additional information about the population distribution in the MS is available in these points. (There would be information about how well the data fits the normal distribution in these 1,000 data points.)

Another consideration is that all the techniques we have described require independent random samples from each crop in order to estimate the population density in the MS (training data). This point is mentioned because most remote sensing analysts do not work with randomly selected points. In this study, we have tried to work with randomly selected fields. The points within these fields are not a random sample of all possible points in a given crop. Therefore, the data are nested within fields. Consequently, the random selection is restricted to the selection of fields within the randomly selected segments.

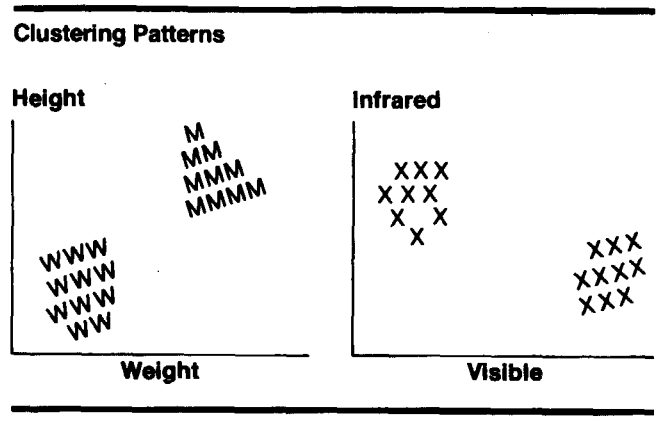
One type of prior information that can be used in the classification procedure is the relative frequency or occurrence (prior probabilities) for each of the K crop or land use populations in the total land population. For example, one-third of all land is cotton, and one-fourth is barley, this information would be used and would affect the partitioning of the measurement space accordingly. If a crop has a high chance of selection, then the area in the MS would be increased. Conversely, if a certain crop has a very low chance of occurrence, then the area in MS would be adjusted downwards.

Clustering

Clustering is a data analysis technique by which one attempts to determine the natural or inherent relationships in a set of observations or data points (10). To get an intuitive idea of what is meant by natural or inherent relationships in a set

of data, consider the examples in appendix fig. 5. If one were to plot height versus weight for a random sample of students without regard to sex on a college campus, it is likely that two relatively distinct clusters of observations would result, one corresponding to the men in the sample (heavier and taller) and another corresponding to the women (lighter and shorter). Similarly, if the spectral reflectance of vegetation in a visible wave band were plotted against reflectance in an infrared wave band, dry vegetation and green vegetation could be expected to form discernible clusters.

Appendix figure 5



If the data of interest never involved more than two attributes (measurements or dimensions), cluster analysis might always be performed by visual evaluation of two-dimensional plots such as those in figure 5. But beyond two or possibly three dimensions, visual analysis is impossible. For such cases it is desirable to have a computer perform the cluster analysis and report the results in a useful fashion.

In regard to the application of clustering to remote sensing research, the greatest use of cluster analysis has been for ensuring that the data used to characterize the crop or land use classes do not seriously violate the assumption of Gaussian statistics. In general, it may be expected that each distinct clustering center will correspond to a mode in the distribution of the data. Therefore, with the objective of defining a crop or land use subclass for each cluster center, the possibility of multimodal (and hence, definitely non-Gaussian) crop or land use distributions is essentially eliminated.

A more detailed report on the technical development of several clustering algorithms, is provided by Swain (10).

Appendix C--Software Improvements

All Iowa analysis was done using the EDITOR processing system with the exception of reformatting tapes and some of the grayscale printing for registration. The latter functions were performed using the IBM 370 at WCC (the Washington Computer Center, USDA).

EDITOR is an interactive image processing system which runs under the TENEX operating system on the DEC SYSTEM-10 and provides a link via the ARPA network to the ILLIAC IV for large-scale batch processing. EDITOR is a large collection of programs all called from a single main program using simple commands describing the program functions. The programs communicate with each other through various files. The TENEX system specifically used for the analysis is at BBN; the ILLIAC IV complex is located at Moffett Field near Sunnyvale, Calif.

EDITOR is used to digitize segments, register scenes, and locally register segments to scenes. Once the segment has been located on the scene, a mask file is created so that pixels may be associated with the fields of the segment. A program in EDITOR is also used to transform the ground truth information obtained from the State offices to a form more readily used by other programs. This ground truth information consists of various attributes such as size and ground cover of the fields in the segments.

When the masks and ground truth information are available, EDITOR is used to create packed files of pixels corresponding to ground covers of interest. These packed files constitute the training data and are cluster analyzed, either at BBN or at the ILLIAC IV for larger data sets to generate statistics files (means, variances, and covariances) representing the ground covers. The categorized packed files are used to create a tabulation file by ground cover and category which is one of the principal inputs needed for sample estimation along with the ground truth information. The sample estimation process creates an estimator parameter file to be used for large-scale estimation and also some values to indicate the expected quality of the estimate.

Large-scale estimation is performed in EDITOR by county and land-use strata. Thus, it is necessary to digitize strata boundaries within the counties, register these to the LANDSAT scene, and create county masks. Each scene is then classified and aggregated on the ILLIAC IV using the masks for counties contained in the scene. The result of the aggregation, taking into account the categories corresponding to the various covers as determined in training, are used with the estimator parameter file to generate the final large-scale estimates. This entire procedure of using EDITOR to obtain estimates is described in more detail in (9).

For the Iowa project, EDITOR was not changed in any substantial or basic manner. However, a number of improvements were made to facilitate its use. An important improvement made to several programs was the addition of a DIRECTORIES command, a facility to allow access of standard named files in various directories. This command, previously available in certain EDITOR programs, was also included in the mask generation, segment plotting, and local calibration (local registration of segments) programs. Use of different directories allows improved organization of data and helps to circumvent certain limitations of TENEX.

Another improvement was to make standard repetitive use of certain programs easier. Changes were made allowing the user to enter a number of parameters and file names upon starting a program and then let it process the data for a possibly lengthy run without further intervention. Among the programs improved in this manner were those for creating packed files, printing scattergrams (of packed files), and performing large-scale estimation.

*U.S. GOVERNMENT PRINTING OFFICE : 1980 O-310-945/ESCS-189

Other improvements included use of a batch method to transfer files to and from the ILLIAC IV site, the display of estimation results in both hectares and acres, creation of a new program to prorate estimates by county and strata based on the number of frame units, and allowing direct input of a file of segment shift distances (from the estimated positions based on the scene registration) to eliminate an unnecessary step in segment registration.

UNITED STATES DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C. 20250

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF
AGRICULTURE
AGR 101
THIRD CLASS