CROP-AREA ESTIMATES FROM LANDSAT: TRANSITION FROM RESEARCH AND DEVELOPMENT TO TIMELY RESULTS

by

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

Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture
Washington, D.C.

JUNE 1979

I. INTRODUCTION

The utility of LANDSAT data in developing data in crop-area statistics has been demonstrated by a number of investigators. Generally such studies have had "proof of concept" in a research and development (R&D) mode as their primary objective. Obtaining timely results for consumption by agricultural data users has been of secondary importance in most R&D studies. In contrast, this paper describes recent efforts by the Economics, Statistics, and Cooperatives Service (ESCS) of USDA in developing timely crop-area estimates from LANDSAT with measurable and improved precision. The 1978 Iowa corn and soybean crops were the key items of interest for this project.

One of the functions of ESCS is to estimate 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 up until mid-January at which time final national and state estimates for the crop year are made. Estimates for individual counties and in some states for multi-county areas, called Crop Reporting Districts, are made by ESCS's State Statistical Offices (SSO's) in cooperation with state government agricultural agencies. Small area estimates, however, are often not published until April of the year following the crop year.

From 1972 to 1977, ESCS has investigated the ability of LANDSAT data to improve crop-area estimates at state, multi-county and individual county levels. The results from these studies have been mixed. For winter wheat, substantial improvements in the precision of crop-area estimates were obtained in Kansas. For corn and soybeans, however, good results were obtained only for a subset of investigation areas. These previous R&D efforts took over a year on the average to complete and thus the results were not useable for setting final area estimates for the current crop year. In 1978, on the other hand, ESCS strove to develop timely LANDSAT-based crop-area estimates to supplement current area survey estimates. These estimates were then input to the 1978 Annual Crop Summary released by the USDA's Crop Reporting Board on January 16, 1979. The estimates were also used by the Iowa SSO in making multi-county estimates.

II. STATISTICAL METHODOLOGY

A. DIRECT EXPANSION ESTIMATION (GROUND DATA ONLY)

Aerial photography obtained from the Agricultural Stabilization and Conservation Service is visually interpreted using the percent 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. ESCS conducts a survey in late May, known as the June Enumerative Survey (JES). In this general purpose survey, area devoted to each crop or land use is recorded for each field in the sampled area frame units (segments). The scope of information collected on this survey is much broader than crop area alone. Items estimated from this survey 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 is conducted providing rigid controls to minimize nonsampling errors. The notation used for the stratified random sample is as follows.

Let $h=1,2,\ldots,L$ be the land use strata. For a specific crop (corn, for example) the estimate of total crop area for all purposes and the estimated variance of the total area is 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,[1] and we will denote this by $\hat{Y}_{DE}$.

B. REGRESSION ESTIMATION (GROUND DATA AND COMPUTER CLASSIFIED LANDSAT DATA)

The regression estimator utilizes both ground data and classified LANDSAT pixels.[1] The estimate of the total Y using this estimator 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:

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

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 $r_h^2 \rightarrow 1$ 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:

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

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 where 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, et. al.[2] The regression estimator described above is called the separate form of the regression estimator. An alternate form for the regression estimator, called the combined form, is described by Craig, et. al.[3] Conditions under which use of the combined form are appropriate are discussed by Cochran.[1] Several types of estimates have also been developed for individual counties.[2,4]

III. GROUND DATA

The ground data required were: land use strata boundaries digitized from a latitude-longitude map base, individual field data for all fields in the JES sample segments, aerial photographs with accurate field boundary locations and a follow-up survey of JES fields that were not yet planted at the time of the JES.

Digitization of the land use strata boundaries began in mid-January. The land use boundaries were located and digitized from Iowa County Highway Maps. The vertices were digitized from a latitude-longitude coordinate system and then transformed to the row-column system of LANDSAT data using each individual scene's registration transformation. The digitization of the land use strata boundaries was completed in May. Average time to digitize an individual county was one day.

The sample segments had land use or crop type and area recorded for each field during late May and early June during ESCS's June Enumerative Survey. The field boundaries were drawn onto ASCS aerial photographic prints with a scale of approximately 8" = 1 mile. The field level information was then recorded on a questionnaire and drawn onto an aerial photograph. In the LANDSAT study, a special edit of each field using photo and questionnaire data was done to insure accurate field boundary locations. The questionnaire data was collected, edited at the individual farm level, and keypunched by Iowa Crop and Livestock Reporting Service personnel. The data was then transmitted to Washington, D.C. for individual field level editing. After editing the JES data, a computer tape with all ground data information was sent to Bolt, Beranek, and Newman (BBN) data processing facility in Cambridge, Massachusetts. However, at the time of the JES, all fields had not yet been planted. Thus, a follow-up survey was conducted from July 21 to August 1. The follow-up survey questionnaire and aerial photography were used to determine the land cover for any fields that were not planted at the time of the JES. The follow-up survey data was then used to update the ground data computer files at BBN.

Digitization of fields in the JES segments was the next step. All field boundaries were digitized from the aerial photographs using polygonal vertices which were transformed to latitude-longitude. Boundaries of 8600 fields from the 298 JES segments were digitized. The process began in mid-July and ended in mid-September. Average time to calibrate and digitize a segment was one hour.

IV. LANDSAT DATA ACQUISITION

Twelve LANDSAT scenes were required to virtually cover the state of Iowa. The LANDSAT scene covering the northwest corner of Iowa was not analyzed because only 200 square kilometers of Iowa was not contained in LANDSAT scenes further to the east. This unimaged area in northwest Iowa was less than 0.2 percent of the total area of the state. The location of the twelve LANDSAT scenes can be seen in Figure 1.

Based on ESCS's previous LANDSAT analysis experience in Illinois and on the 1978 planting times, LANDSAT imagery was desired during early to mid-August. Table 1 lists images which were registered for the Iowa project. Image dates ranged from August 7 to September 4, 1978. As can be seen from the table, some of the registered images contained clouds.

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 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. However, this imagery was never received by NASA's Goldstone receiving station. Thus, partially cloud covered imagery for August 7 was used for path 27.

Because of the various dates of the Iowa LANDSAT imagery, the associated cloud-cover problems, and the different times at which ESCS received LANDSAT data, Iowa was partitioned into ten separate areas, called analysis districts. (See Figure 2) The smallest analysis district, number 2C, contained three counties; the largest, number 1, contained twenty counties. Analysis district 3A consisted of the thirteen cloud-covered counties.

A number of analysis districts--for example, 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. Because of time pressure, analysis districts were formed when data were received, instead of waiting until all data for a given image data 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 3 displays by analysis district the beginning and ending dates for the LANDSAT processing activities. The first analysis district to be completed was 2A on October 26; the last.

2B, was completed on December 21. Table 2 displays various summary statistics for the time required by each processing activity. As can be seen from the table, on the average, data delivery took the longest and was the most variable in duration of the three processing activities.

By examining daily GOES satellite weather photos, ESCS was able to select candidate cloud-free LANDSAT scenes within one to two 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, twelve of which were registered for the calculation of crop-area estimates. A histogram of delivery times, i.e. time from date of satellite overpass to receipt by ESCS, for the twenty-four ordered tapes is shown in Figure 4(a). Figure 4(b) displays the tape delivery times for the twelve scenes which were registered.

V. DATA PROCESSING SYSTEMS HARDWARE

ESCS purchases computer time on a number of different types of computers. These include:

1. A PDP10 computer in Cambridge, Massachusetts (BBN), used by ESCS 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) in Washington, D.C., used for computer editing of ground truth data, reformatting LANDSAT tapes, and batch printing of greyscales; and

3. The ILLIAC IV computer in Sunnyvale, California, 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 are the following:

- two digitizer tablets,
- Zoom Transferscope,
- Terminal plotter with controller,
- leased phone line with muliplexor, and
- fifteen KRS (keyboard send-receive) terminals of various types.

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

Total IBM 370-168 computer charges for the Iowa project were $7,000 (includes 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.

VI. SOFTWARE AND DATA MANAGEMENT

A. SOFTWARE

All LANDSAT data analysis for Iowa was done using the EDITOR software system with the exception of reformatting tapes and some of the greyscale printing for registration. The latter functions were performed using the IBM 370-168 at WCC.

EDITOR[5] is an interactive image processing system which runs under the TENEX operating system. EDITOR 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 function of the programs. The programs communicate with each other through various files. 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.

B. DATA MANAGEMENT

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

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

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

3. LANDSAT tapes from NASA's Goddard Space Flight Center in Greenbelt, Maryland were reformatted and tapes mailed to Cambridge, Massachusetts and Sunnyvale, California.

4. The PDP10 in Cambridge, Massachusettes was accessed via ARPANET, leased line, or Federal Telephone Service (FTS) dial-up for interactive processing of LANDSAT data for sample segments.

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

6. Aggregated ILLIAC IV classification results were transmitted back to Cambridge, Massachusetts, over ARPANET for interactive calculation of crop-area estimates.

VII LANDSAT SCENE REGISTRATION

LANDSAT data registration procedures used for the twelve scenes were: data reformatting, selection of control points, determination of latitude-longitude from USGS quad maps and row-column from grey-scales, third order polynomial regression analysis, and the matching of predicted segment locations with grey-scales for precise segment location. Root mean square errors for LANDSAT scene registration ranged from 45.3 meters to 91.7 meters. Registration procedures took, on the average, two weeks to complete which was a considerable improvement over previous ESCS LANDSAT projects.

VIII. LANDSAT CLASSIFICATION

Based upon the stratified random sample of ground data segments, the estimated average field size was 12 hectares for corn and 13 hectares for soybeans. Consequently the number of pure field interior pixels was approximately 59 percent of total pixels for these two major crops. The modified supervised approach[6] was used in developing training data for crop signatures.

Within each known cover type, several methods were used to train the classifier:

Resubstitution, in which all the field interior pixels for the cover type are used; the 1/2 sample partition method, in which the data for 50 percent of the sample segments are used; and a method where small fields (< 5 hectares) were excluded from the training data. Once the training data for a cover type was established, there were two additional considerations in developing classification parameters. These were the use of prior probabilities for a cover type and clustering within a cover type's training data. Types of prior probabilties used were those proportional to the reported acres in the sample segments or equal prior probabilities.

The primary objective of classification was to minimize the variance of the resulting regression estimates, thus little attention was given to estimating the traditional percent correct classification measures. As previously shown, the variance of the regression estimate is minimized when the corresponding r^2 is maximized.

IX. CROP-AREA ESTIMATES

Crop-area estimates for corn and soybeans were developed at the state, multi-county (analysis district), and individual county levels. At the state and multi-county 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 3 and 4. 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 isn't an unusual event. The conventional direct expansion estimate of ground data had to be used for the 13 county area in Iowa.[7] Individual county estimates had C.V.'s ranging from 7.1 to 59.9 percent for corn and 9.0 to 100 percent for soybeans. C.V.'s 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 multi-county level estimates. However, these LANDSAT based regression estimates were not the sole source of data in determining the state and multi-county estimates.

X. SUMMARY

The primary project goal of developing timely and precise crop-area estimates at the state and multi-county level utilizing both LANDSAT data and conventional ESCS ground data was accomplished. These estimates were used as input to official USDA crop reports for Iowa. The major benefit of LANDSAT regression estimates to ESCS is substantial improvements in precison with no increase in respondent burden associated with ground surveys. The repeatability of such an effort, however, is crucially dependent upon 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 Crop Reporting Board. Cloud cover is a serious problem is 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.

ACKNOWLEDGMENTS

The authors wish to acknowledge several individuals and organizations for their deeply appreciated contributions to this project: fellow members of ESCS's New Techniques Section; Iowa Crop and Livestock Reporting Service; NASA-Goddard; NASA-Ames; Bolt, Beranek and Newman in Boston; Charles Caudill, Galen Hart, and William Wigton of ESCS for their overall managment support; Kathy Morrissey, ESCS for ground data management, and Tricia Brookman, ESCS, for her fine typing efforts.

REFERENCES

1. Cochran, William G., Sampling Techniques, (Third Edition), John Wiley and Sons, 1977.

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

3. Craig, M., Sigman, R., and Cardenas, M., "Area Estimates by LANDSAT: Kansas 1976 Winter Wheat", Economics, Statistics, and Cooperatives Service, U.S. Department of Agriculture, Washington, D.C., August 1978.

4. Cardenas, M., Blanchard, M., and Craig, M., "On the Development of Small Area Estimators Using LANDSAT Data As Auxiliary Information", Economics Statistics, and Cooperatives Service, U.S. Department of Agriculture, Washington, D.C., August 1978.

5. Ozga, M., Donovan, W., and Gleason, C., "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, Indiana.

6. Fleming, M. D., Berkebile, J. S., Hoffer, R., "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, Indiana.

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

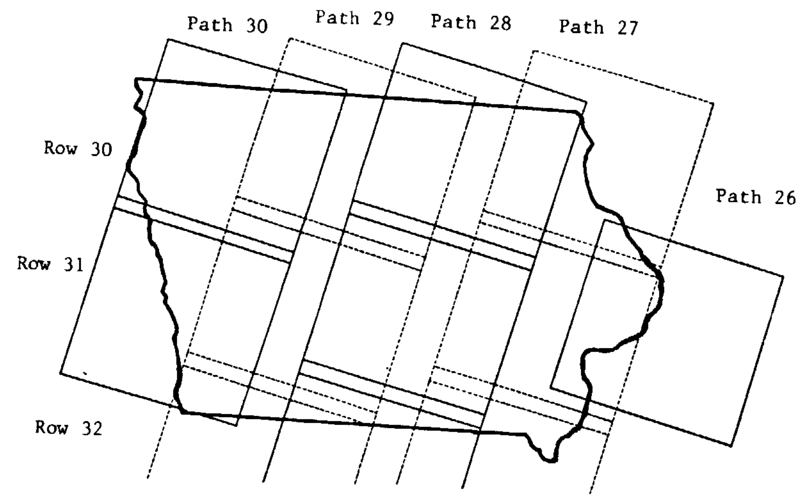

Figure 1. LANDSAT Imagery Paths and Rows

Table 1. Iowa LANDSAT Scenes Used in Crop-Area Estimation

Path	Row	Date	Percent 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 8	10	21293-15500
	31	August 8	15	21293-15502
	32	August 8	10	21293-15505
26	31	August 6	0	21292-15444

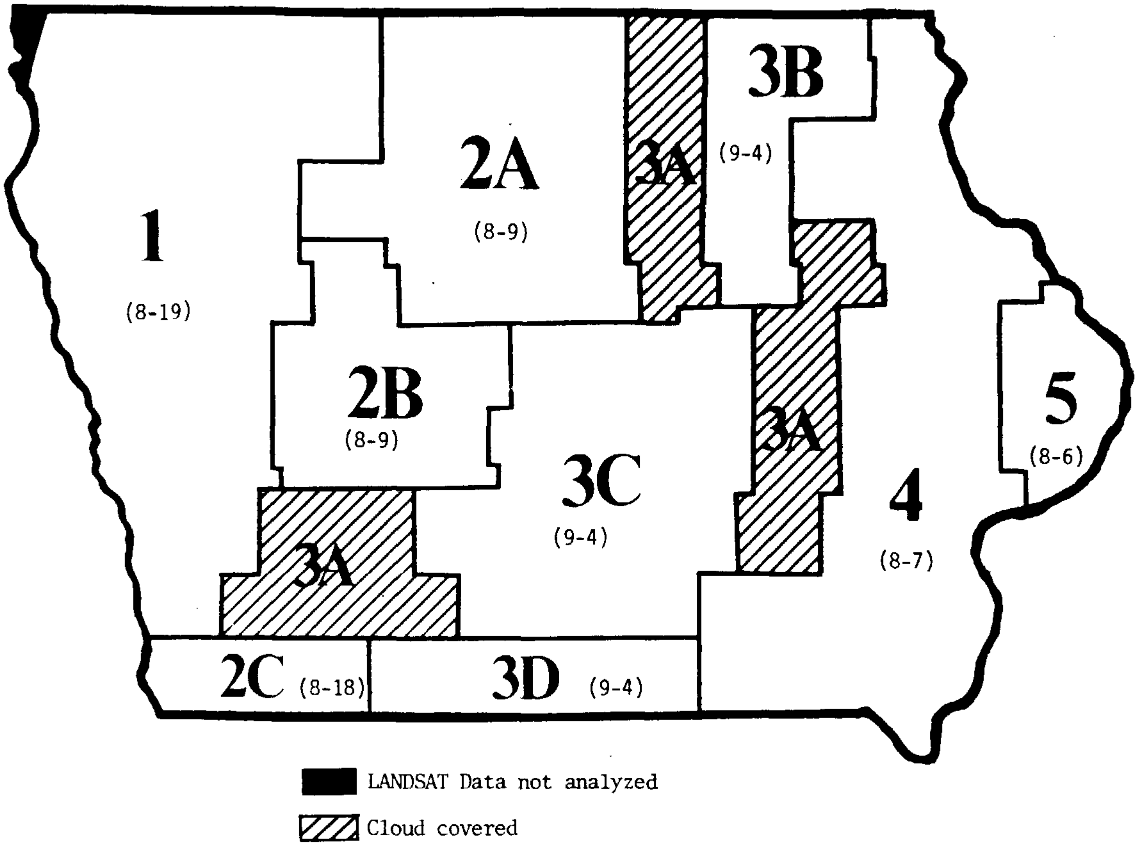

Figure 2. Analysis Districts and Images Dates.

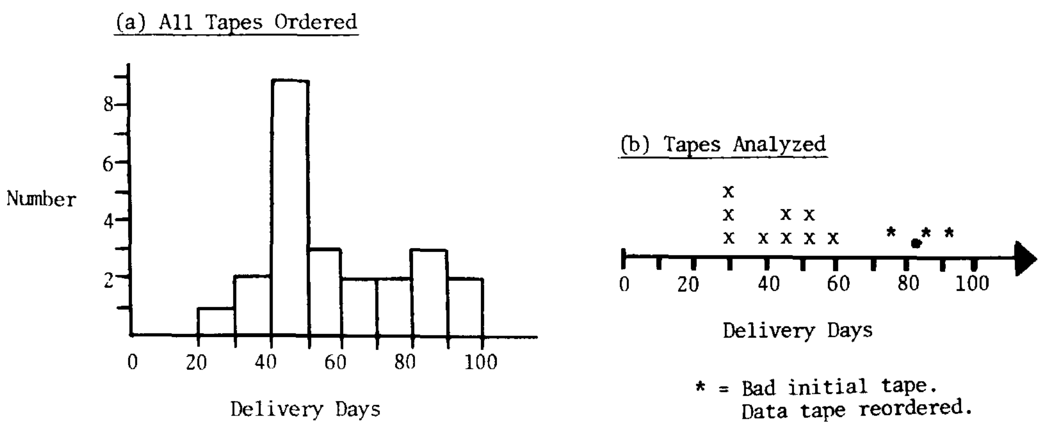

Figure 4. Delivery Times for LANDSAT Tapes (Measured in Calendar Days from Date of Satellite Overpass to Receipt by ESCS):
(a) For all Tapes Ordered, and
(b) For Tapes Analyzed

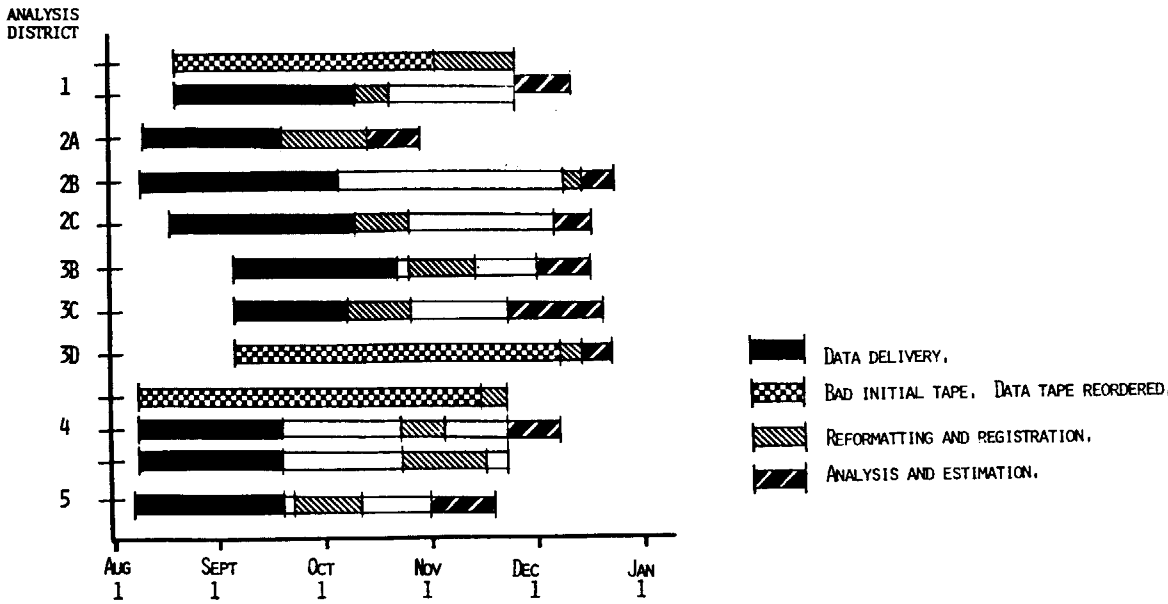

Figure 3. LANDSAT Data Processing Activities Beginning and Ending Dates

Table 2. Durations for LANDSAT data processing activities: Summary Statistics

Activity	Duration in calendar days			
	median	min	max	quartiles
Data delivery	49	32	93	37,66
Reformatting, Registration	16	4	25	8,20
Analysis, Estimation	13.5	7	26	10,18

TABLE 3. 1978 IOWA CORN RESULTS (PLANTED HECTARES)

Analysis District	$\hat{Y}_{DE}$ = 1978 Direct Expansion	Coefficient of Variation for $\hat{Y}_{DE}$	$\hat{Y}_R$ = 1978 LANDSAT Regression	Coefficient of Variation for $\hat{Y}_R$	Range of r^2 for h=1, ..,L	Relative Efficiency
1	1,462,074	3.48	1,460,234	2.20	.57-.92	2.51
2A	828,772	4.47	818,892	2.50	.71	3.28
2B	332,050	11.50	454,252	3.40	.78-.94	5.98
2C	106,036	10.98	109,959	9.50	.30	1.24
*3A	657,462	4.36	-	-	-	-
3B	276,112	10.05	268,022	8.47	.38	1.49
3C	550,581	7.46	542,081	6.02	.34-.40	1.58
3D	83,658	17.76	82,798	18.65	.07	0.93
4	1,029,688	6.72	896,084	4.47	.65-.71	2.99
5	148,148	11.10	149,820	6.03	.75	3.32
State	JES= 5,525,807	2.3	5,439,604	1.5	.07-.94	2.43

*LANDSAT data not available.

TABLE 4. 1978 IOWA SOYBEANS RESULTS (PLANTED HECTARES)

Analysis District	$\hat{Y}_{DE}$ = 1978 Direct Expansion	Coefficient of Variation for $\hat{Y}_{DE}$	$\hat{Y}_R$ = 1978 LANDSAT Regression	Coefficient of Variation for $\hat{Y}_R$	Range of r^2 for h=1, ..,L	Relative Efficiency
1	747,759	8.11	781,566	4.04	.58-.88	3.70
2A	655,049	6.75	675,293	3.42	.74	3.68
2B	256,944	12.91	255,540	6.11	.74-.98	4.55
2C	95,196	24.97	97,497	11.67	.80	4.37
*3A	401,671	9.20	-	-	-	-
3B	86,550	28.00	125,300	9.37	.79	4.26
3C	328,662	14.51	338,363	7.06	.77	3.98
3D	82,633	32.55	95,933	10.20	.89	7.59
4	441,032	12.68	424,782	7.97	.45-.83	2.73
5	47,060	29.20	48,580	12.53	.86	5.10
State	JES=3,205,320	3.91	3,244,525	2.50	.45-.98	2.38

*LANDSAT data not available.

FILE COPY

PRECISION OF CROP-AREA ESTIMATES

TITLE: ↑
AUTHOR: GEORGE A. HANUSCHAK
DATE: APRIL 1979

KEYWORDS
crop area estimates
statistical precision
relative efficiency
controlled accuracy
cloud cover

George A. Hanuschak
Statistical Research Division
Economics, Statistics, & Cooperatives Service
U.S. Department of Agriculture
Washington, D.C. 20250

Invited Paper at the
Thirteenth International
Symposium of Remote Sensing
of Environment
April 1979
Ann Arbor, Michigan

PRECISION OF CROP-AREA ESTIMATES

GEORGE A. HANUSCHAK

Statistical Research Division
Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture
Washington, D.C.

1. INTRODUCTION

The utility of LANDSAT data in developing crop-area estimates has been demonstrated by several investigators. The major issue in evaluating crop-area estimates is how to measure the precision and accuracy of the estimates. This paper describes the methods used by the Economics, Statistics, and Cooperatives Service (ESCS) of the U.S. Department of Agriculture (USDA) for evaluating crop-area estimates.

Annually in late May and early June ESCS conducts a nationwide agricultural survey, referred to as the June Enumerative Survey (JES), consisting of interviews with farm operators in randomly sampled areas of land called segments. Segments enter the JES by selection through stratified random sampling. The strata are land use categories determined by visual interpretation of aerial photography or LANDSAT imagery and delineated on county highway maps. The JES segments are typically 2.59 square kilometers in size. Strict survey quality control methods are used prior to, during, and after the data collection period to minimize nonsampling errors at the elementary sample unit (segment) level. Methods such as training of statisticians and interviewers prior to each JES, use of aerial photographs during the interview with the farm operators, reinterviews by supervisory interviewers, follow-up survey interviews, data editing-manual and machine, current aerial photography for comparison, etc. are used to insure data quality. The relative sampling errors for major crops at the national and regional level are on the order of 1-3 percent. At the state level they are on the order of 2-10 percent.

Any use of LANDSAT data by ESCS must be an improvement over the extremely efficient JES. The Statistical Research Division (SRD) of ESCS has developed techniques using LANDSAT and JES data together that produce lower sampling errors than the JES alone. The basic method used by the SRD is a simple application of regression estimation as described in William Cochran's, Sampling Techniques. This technique has been applied in Illinois (1975), Kansas (1976), and Iowa (1978). The Iowa project was completed during the 1978 crop year and in time for the regression estimates to be input to the official USDA Crop Reporting Board's Annual Crop Summary for Iowa released on January 16, 1979.

2. STATISTICAL METHODOLOGY

2.1 Direct Expansion Estimation (Ground Data Only)

Aerial photography obtained from the Agricultural Stabilization and Conservation Service is visually interpreted using the percent of cultivated land to define broad land-use strata. Within each stratum, the total area is divided into N_h 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. ESCS then conducts a survey in late May, known as the JES. In this general purpose survey area devoted to each crop or land use is recorded for each field in the sampled area frame units. The scope of information collected on this survey is much broader than crop-area alone. Items estimated from this survey include crop-area by intended utilization, grain storage on farms, livestock inventory by various weight categories, and agricultural labor and farm economic data. Intensive training of field statisticians and interviewers is conducted providing rigid controls to minimize nonsampling errors.

The form of an estimated state total for a crop from a stratified random sample is as follows:

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

Let Y = Total corn area for a state (Iowa, for example).

$\hat{Y}$ = Estimated total of corn area for a state.

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

Then,

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

The estimated variance of the total is:

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

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,[1] and we will denote this by $\hat{Y}_{DE}$.

As an example, for the state of Iowa in 1978, the direct expansion estimates were·

Corn $\hat{Y}_{DE}$ = 5,525,807 Hectares

Relative Samping Error = $\sqrt{v(\hat{Y})} / \hat{Y}$ = 2.4%

Soybeans $\hat{Y}_{DE}$ = 3,205,320 Hectares

Relative Sampling Error = $\sqrt{v(\hat{Y})} / \hat{Y}$ = 3.9%

2.2 Regression Estimation (Ground Data and Computer Classified LANDSAT Data)

By means of a regression estimator both ground data and classified LANDSAT data can be utilized to estimate crop hectarage.[2] (Regression estimators are discussed in most sampling texts, e.g. Cochran[1]) The estimate of Y using the separate form of the regression estimator is

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

where

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

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