Improved Crop Acreage Estimation Utilizing LANDSAT

Galen F. Hart, Richard S. Sigman, and George A. Hanuschak
U. S. Department of Agriculture
Economics, Statistics, and Cooperatives Service
Statistical Research Division
Washington, D. C. 20250

JUNE 1978

Introduction

Since the launch of LANDSAT I in July, 1972, the Statistical Research Division of Economics, Statistics, and Cooperatives Service has conducted research investigations toward utilizing spectral reflectance data to improve crop acreage estimating ability. The interest in LANDSAT data stemmed from the potential for complete or census like coverage for large areas in a very short time span. The general plan for investigations was to develop a methodology taking advantage of the best features of an existing ground data collection system and LANDSAT digital data.

"Baseline" Methodology

To discuss what is meant by improved acreage estimation it must first be established what existing methodology provides. The baseline methodology referred to by this paper is developed around the land area sampling frame of ESCS. The concept is straightforward. The entire area of the United States is partitioned or stratified by agricultural land use. For a particular state the number of partitions or strata may vary but for a typical state about ten uniquely separable categories of land use are delineated. The task of dividing the land area of a state into strata is accomplished by interpreting photography obtained primarily from the USDA Agricultural Stabilization and Conservation Service.

Urban or non-agricultural land must be separated. By using a percent cultivation criteria degrees of cropping intensity can be interpreted and delineated. Woodland and grazing land are separated. When the task of stratification is completed all land in a state has been uniquely assigned to a particular strata. Within each stratum the total land area is sub-divided into sampling units. This collection of all sampling units for all strata is called the area sampling frame. A probability sample of units is selected from each stratum and the selected unit is delineated on a small scale aerial photography (8 inches to the mile).

A major national survey of about 16,000 sampling units is conducted in late May of each year. This survey is known as the June Enumerative Survey (JES). About (sixteen hundred) part-time interviewers employed by the Service obtain a complete set of agricultural information for each of the selected sampling units during a two week interview period. Intense training of field supervisors and interviewers is conducted providing rigid controls to minimize potential error. Each parcel of different land is delineated on the "8 inch to the mile" photograph and the land use and acreage is recorded on a questionnaire. The interviewer also obtains and records on the questionnaire information on crop utilization, grain storage, livestock inventory by various weight classes and agricultural labor and economic items. The data collected during this Survey serves a wide range of estimating programs of the Agency. This same set of sampling units or sub-samples thereof are visited several other times during the year to obtain such information as yield data, production practice data and to update information obtained at the time of the JES. For major crops at the state level this survey provides estimates with relative sampling errors on the order of 3 to 8 percent.

At the national level the sampling errors for major items are on the order of 1.5 to 3.5 percent relative sampling error. Sample units drawn from the area sampling frame provide estimates of items relevant to current agricultural needs at state and national levels that are timely (publication within three to four weeks after data collection), accurate and are acquired at a reasonable cost. The cost of annual surveys, including maintenance of the sampling frame, is about 3.5 million dollars.

Potential for Improvement Offered by LANDSAT

Sampling methodology offers an apparent cost-effective means of collecting data at the state and national levels. But it is an attribute of sampling methodology that sample size is almost independent of population size. For a typical state with about 400 sampling units, estimates for major crops have relative sampling errors of about 3 to 5 percent. The sample size required to achieve the same precision at the national level would be less than 1000 sampling units. In order to provide a target 3 to 5 percent sampling error at the county level, several hundred sampling units would be required and therefore, sampling is not considered to be a cost-effective method to provide small area statistics. LANDSAT, however, being a complete or census coverage method for collected data offers potential for small area land use statistics since four bands of spectral data are acquired for each acre of land.

Remotely sensed data from LANDSAT does have a limited scope in that the sensors provide data related to general land use and perhaps plant vigor. Also, there is the task of converting spectral data to crop acreage land use or yields information. Past research has shown that spectral data is of widely ranging quality. Information content can be relatively high for cloud free imagery with minimal atmospheric disturbance at optimal times during the growing season. However, cloud cover and atmospheric disturbance can result in either very low or no information value in LANDSAT data.

It has been demonstrated through many experiments that information extraction from LANDSAT digital data is very directly associated with the amount of ground data available to convert spectral data to acreage or productivity information. Ground data are needed to obtain the "signatures" of spectral data. Selected sampling units from the land area sampling frame provides a substantial amount of ground data to "train" a computer to classify land use from spectral data. Also since the area sample is a probability sample the data collected stands on its own plus transferring inference power to the process of combining this data with spectral data. The statistical procedure of combining these two data sources is known as "double sampling" and estimation is performed by the "regression estimator." LANDSAT data are appropriate as an auxiliary data set to apply the theory. If the correlation between ground and spectral data for a particular land use is sufficiently high then, since LANDSAT data is without sampling error, estimates resulting from the combination would have a lower overall net sampling error. In fact, if there was a perfect one-to-one relationship between ground and spectral data, estimates resulting from the combination would be without sampling error.

The task of forming methodology for combining spectral and ground data has been completed and is in a computer network (ARPA) environment with software developed cooperatively between ESCS and the Center for Advanced Computation, University of Illinois. The network permits communication between researchers in Washington, D. C. and the Center for Advanced Computation and further permits the large scale computer classification task required on an efficient processor known as ILLIAC IV at the NASA Ames Laboratory, Moffett Field, California. Some of the processes involved in this software, known as EDITOR, include interactive digitizing, storage, and retrieval of the ground data, extracting reflectance data for the sampling units where the ground data is collected, determining the statistics necessary for establishing a relationship between spectral and land use data and generating the combined estimates and sampling variances. Methods have also been developed to handle the situation when cloud cover or lack of LANDSAT coverage allows only for the use of ground data.

Presently, the most time consuming task in utilizing LANDSAT data is the one of registration. Registration is the process of accurately aligning LANDSAT data to a map coordinate base so that individual LANDSAT data points in the sample units can be tagged with their known crop identity. This task presently takes about 100 person hours per LANDSAT scene (a scene being a 100 nautical mile sqaure). This time requirement assumes approximately 40 sample units and 50 common control points with a trained person performing the task. In the near future a portion of this task and perhaps a significant portion will be eliminated by the delivery of "registered" LANDSAT. The digitizing process deletes field boundaries by assigning a high density of coordinates. This process requires about one person hour per sample unit or about 40 person hours per LANDSAT scene. Computer classification of an entire LANDSAT scene presently requires about 300 dollars for computer time. This cost will likely be reduced significantly in the future and is not a significant cost in the total operation.

An Example
1975 Illinois Acreage Estimation Project

The entire state of Illinois was the test area for this research activity. The objective was to estimate the acreage of major spring planted crops at the county level. There were 300 selected sample units. Sample unit size varies by stratum and boundary delineation constraints but a typical sample unit is about one square mile. There were two counties in the center of the state that were not covered by a single cloud fee LANDSAT scene. It was decided not to utilize the necessary special techniques required for analysis of these two counties--all other counties in the state were analyzed.

Due to the different LANDSAT scenes and passes, the state was divided into analysis areas. Seven such areas were defined for the study (Fig. 1).

Our evaluation criteria for success was reduction of the relative sampling error (r.s.e.). Both estimation procedures, direct expansion of data from sample units and regression using both sample unit and LANDSAT data provide estimates of r.s.e. that can be compared directly. A third data set was also used for comparison--the Illinois State Farm Census. This is a post-growing-season accounting of specified crop and livestock items obtained as an adjunct to a state tax accounting. The census is not a controlled accounting, and adjustments are made for consistency. But these data provide independent comparisons.

The use of LANDSAT data did result in significant reductions in the relative sampling errors of corn and soybeans from the use of sample unit data alone for the analysis areas but the reductions are not overwhelming (Table 1). The estimates themselves are within sampling error. The relative sampling errors of minor crops for analysis areas and county estimates are unacceptably high by the standards ESCS normally places on estimates (Table 2). We are reasonably certain however that with refined analytical techniques and improved sensors it will be possible to produce estimates for major land uses at the county level with the same precision that we now have at the state level with sample unit expansions.

Cost-Effective Application

Since the technical feasibility for using LANDSAT digital data to generate improved estimates of major land uses at the county or multi-county levels seems assured the question then becomes "Can the incorporation of this methodology in operating programs be cost justified based on improved information value?" The question is relevant since no USDA operating program utilizes LANDSAT digital data as an integral part of operations. For ESCS the question becomes, "Is the user value of quality small area crop acreage estimates sufficient to justify the cost of implementing the methodology?" Neither question has been answered to date but it seems likely that the answer to the second question would be no--at least in the near future timeframe.

At present USDA is exploring to determine if it might be possible to generate a series of outputs, including statistics, image products, and special overlays, that would benefit a number of different federal and state program responsibilities.

Within USDA the Forest Service, Soil Conservation Service, and Agricultural

Stabilization and Conservation Service, and Economics, Statistics, and Cooperatives Service all have program responsibilities that require land use inputs. Many state governments are also involved in land use planning and are seeing a greater need to monitor changes in land use. If a "core" processing system similar to that discussed in this paper could be modified so that many users could obtain their products as marginal outputs then the cost of core processing could be distributed over a number of "benefited" programs. This approach assumes the "core" processing is the major cost of a particular product output. In other words the "marginal" cost of generating the user specified product is small as compared to the cost of the user independently generating the product. Also the distributed core cost plus the product marginal cost must be favorable as compared to the value of program improvement as a result of including the product. Some possibilities are: (1) Forest and range inventoring and monitoring, (2) inventoring and monitoring of irrigated croplands for planning agricultural water supply demands, (3) integrating land cover and topographic data for erosion potential and water quality management, and (4) monitoring of urban development patterns as they relate to important or prime agricultural lands.

If it is possible to satisfy some of these needs then the next likely occurrence would be to create a land use data base or information system in a geographic format. This would include point data or summary data for configuration of geographic area that could be digitized in a common map base. The creation and utilization of such a base would expand land use analysis capability almost beyond imagination. The next two to five years should be an exciting period in the utilization of space technology for current very down to the earth problems relating to land use and its changes.

Table 1. Estimated Acres of Corn and Soybeans for Wholly Contained Counties in Each Analysis Area.

Analysis Area	No. of Counties Wholly Contained In the Analysis Area	Estimator	Corn		Soybeans	
			Acres	r.s.e.	Acres	r.s.e.
W123[a]	29	Direct Expansion	4,110,150	3.6%	1,539,200	7.7%
		Regression	4,125,400	2.5	1,681,800	5.2
		SSO	3,682,300		1,657,800	
C1A	7	Direct Expansion	1,191,400	7.1	532,700	13.9
		Regression	1,180,500	2.9	523,200	8.2
		SSO	1,196,900		502,900	
C12	20	Direct Expansion	2,907,700	4.5	2,217,200	5.5
		Regression	2,945,100	4.3	2,127,200	5.1
		SSO	2,939,700		1,990,400	
C33+	16	Direct Expansion	1,158,000	9.5	1,675,100	8.6
		Regression	1,077,000	8.6	1,540,000	6.8
		SSO	1,233,000		1,246,000	
E12	12	Direct Expansion	1,781,300	5.6	1,439,500	6.3
		Regression	1,577,300	4.1	1,290,700	6.5
		SSO	1,792,000		1,383,000	
E23+	32	Direct Expansion	1,669,500	7.5	2,431,950	5.2
		Regression	1,615,000	6.9	2,357,850	3.8
		SSO	1,767,000		2,045,000	
West[b] CRD	9	Direct Expansion	1,316,000	8.5	562,000	13.1
		Regression	1,269,000	4.6	574,100	10.6
		SSO	1,125,000		680,000	

[a]W1 and W2 (Fig. 1) were analyzed individually and joined with W3 (not shown on Fig. 1 but follows W2) to form W123

[b]Wholly contained within W2

Table 2. Regression Estimates for Corn and Soybeans in Individual Counties in Western Pass W 123 (Fig. 1)

County	Corn		Soybeans	
	Acres	r.s.e.	Acres	r.s.e.
Adams	166,600	24.0%	83,600	35.3%
Brown	53,700	33.4	24,300	50.7
Bureau	254,000	18.7	110,600	33.4
Calhoun	56,700	25.1	23,300	39.9
Carroll	126,500	17.5	57,200	29.6
Cass	91,700	20.3	54,100	25.5
Fulton	172,100	29.0	91,400	37.8
Greene	136,800	19.2	76,000	24.8
Hancock	190,500	19.3	74,800	36.2
Henderson	104,000	17.3	37,100	36.4
Henry	276,800	17.2	79,400	46.6
Jersey	85,700	21.6	48,900	27.0
Jodaviess	108,300	34.1	27,100	94.2
Knox	174,100	19.5	79,600	31.6
Mason	129,100	21.3	76,100	27.9
McDonough	162,500	17.4	82,500	26.3
Mercer	139,800	18.7	43,900	43.4
Morgan	147,200	17.6	93,700	20.9
Ogle	223,000	19.0	51,500	64.2
Peoria	124,000	24.0	65,300	32.6
Pike	160,100	25.7	78,300	37.3
Rock Island	107,000	18.7	27,500	52.7
Schuyler	84,000	29.0	36,650	46.2
Scott	61,100	19.9	31,500	28.6
Stark	92,000	18.2	40,600	32.1
Stephenson	172,100	18.6	30,600	81.8
Warren	161,800	16.5	64,100	32.2
Whiteside	242,800	16.2	62,400	49.0
Winnebago	121,500	21.5	29,600	68.0

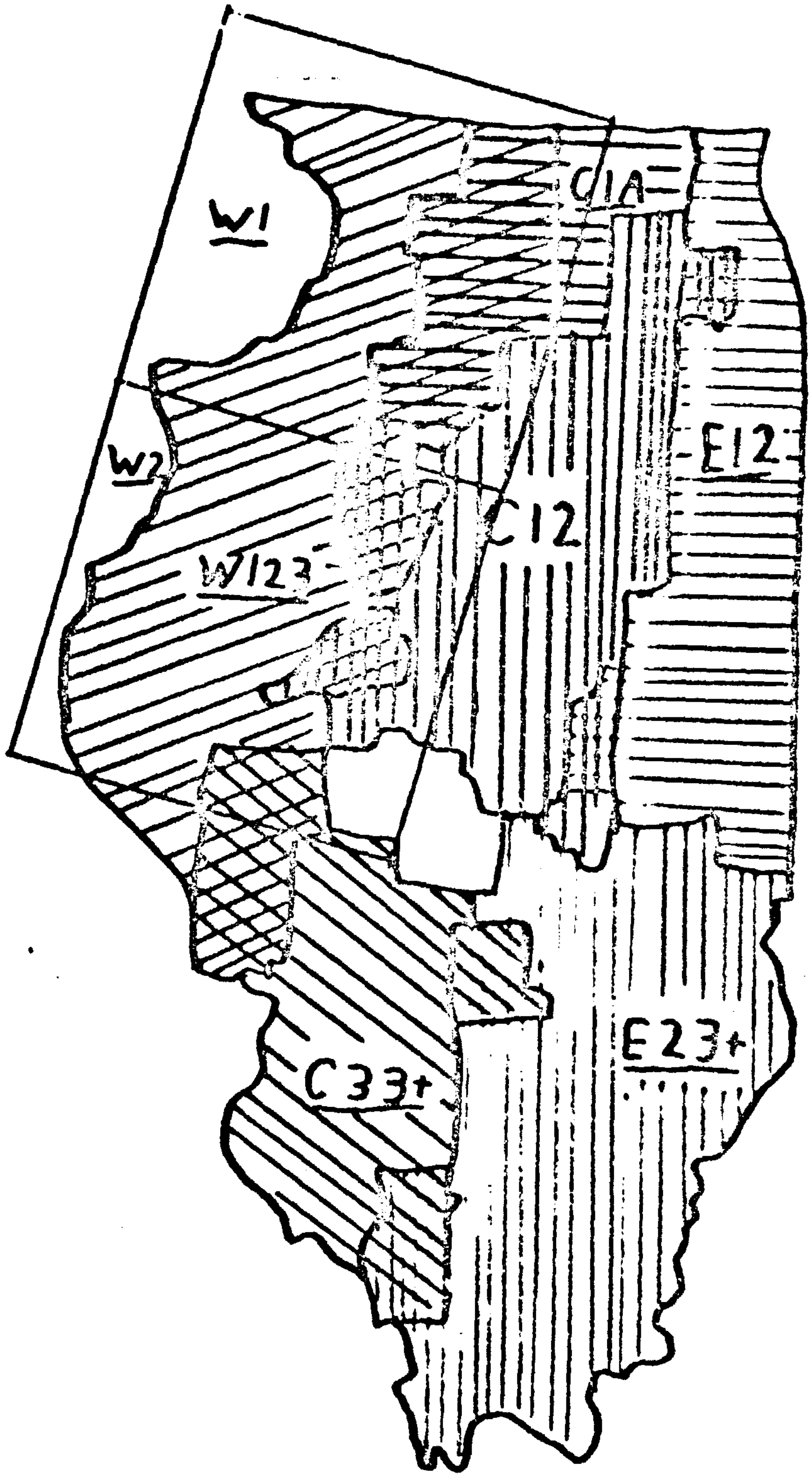

Figure 1. Analysis Areas for 1975 Illinois Acreage Estimation Project.