SRS

84-14 1984

roceedings, SP Meeting, ct. 1984, hatanooga, TN

USE OF GROUND ENUMERATIVE AND LANDSAT DIGITAL DATA TO OBTAIN LAND COVER CLASSIFICATION AND AREA ESTIMATES AT THE STATE LEVEL

Ned Jones, George May, Marty Holko

Research Section
Remote Sensing Branch
Statistical Research Division
Statistical Reporting Service
United States Department of Agriculture
Room 4832 South Building
Washington, D.C. 20250

INTRODUCTION

The USDA's Statistical Reporting Service (SRS) uses digital data from the Landsat satellite to improve crop-area statistics based on ground-gathered survey data. This is accomplished by using Landsat digital data as an auxiliary variable in a regression estimator. Several reports (1, 2, 3) discuss results from this procedure applied to major crops in the midwest. Briefly, the SRS Landsat procedure for major crops in mid-western states consists of the following steps:

- Ground truth, collected during an operational survey, plus corresponding Landsat data are used to develop discriminant functions which in turn are used to classify Landsat pixels as representing specific ground covers,

- Areas sampled by the ground survey are classified and regression relationships developed between classified results and ground truth,

- All of the pixels in the area of interest are classified, and

- Crop-area estimates are calculated by applying the regression relationship to the all-pixel classification results.

In 1981 SRS conducted a research project to determine if it was feasible to extend the Landsat analysis procedure to estimate acreages for desired land-cover categories (4). In this study the Landsat analysis for land-cover was conducted independently of the crop analyses conducted by other Remote Sensing Branch personnel. This research indicated the following: 1) deficiencies in the ground data for certain land covers, 2) a need to conduct a similar study in an area having more diverse land cover types and, 3) a need to use multitemporal Landsat data to improve land-cover classification. In 1982 Missouri was selected for continuing the land-cover research. Ground data were collected but no analyses were conducted due to insufficient Landsat data. Lack of cloud free imagery throughout the growing season resulted in Landsat coverage for only twenty-five percent of the state.

In 1983 the land-cover research was continued in Missouri. The following changes in remote sensing study procedures were made:

1) Areas of land previously defined as non-agricultural land were further categorized into specific land-cover types such as residential, idle, grassland, etc.

2) Additional samples were selected in the non-agricultural strata to improve land cover estimates.

3) Two dates of Landsat imagery were used.

OBJECTIVES

The objectives of the 1983 Missouri crop and land-cover were the following:

1) Provide the Crop Reporting Board (CRB) with estimates of crop area for winter wheat, rice, cotton, corn and soybeans from a combined crop and land-cover Landsat analysis.

2) Provide area estimates for desired land-covers from the combined analysis.

3) Provide a detailed classification of forest covers.

4) Produce classified data tapes of Missouri land covers.

5) Determine the additional cost of land-cover analysis above the cost for crop analysis only.

6) Determine potential users of land-cover analysis and their information needs.

GROUND DATA

During late May and early June each year, SRS conducts a nation-wide survey called the June Enumerative Survey (JES). The JES uses an area-frame sampling technique (5) to sample areas of land called segments. The segments are selected using stratified sampling with the stratification based on percent cultivation. Table 1 lists the stratum definitions for the Missouri area frame and shows the number of segments in the population and the sample size.

During the JES each segment is visited by an enumerator who records all the field boundaries on an aerial photograph. The field acreage and cover type are recorded for each field in the segment.

Because of late planting some fields are recorded as containing crops that the farmer intends to plant. To insure the accuracy of the data for this project, these fields were revisited in August and any discrepancies with the data recorded in June were corrected.

Table 1. Definitions, Population, and Sample Size of Missouri Strata

STRATUM		POPULATION SIZE	SAMPLE SIZE
10	50+cultivated	26,027	100
15	15-50% cultivated	969	4
20	50+% cultivated	13,372	75
25	15-50% cultivated	4,275	17
30	50+% cultivated	23,672	90
35	15-50% cultivated	4,556	17
40	50+% cultivated	14,253	50
45	15-50% cultivated	5,631	21
50	50+% cultivated	7,558	50
55	15-50% cultivated	670	2
91	Agri--urban	7,100	23
92	Agri-urban	4,629	12
1	Woodland	2,959	56
	TOTAL	112,712	517

In order to estimate the acreage of desired land-covers the following modifications to the operational JES were necessary:

Land Cover Definitions

Potential users of SRS generated land-cover data were contacted and asked to determine what land cover types should be included in this study. Land cover terms were defined in a manner that minimized additional training for SRS enumerators and conform with Soil Conservation Service definitions and Anderson level 2 classification. The final list of land covers used, were as follows:

Sown Crops
Row crops
Forest
  Conifer
  Hardwood
  Mixed
Grazed Forest

Residential
Industrial Commercial
  and Services
Transportation,communication
  and utilities
Other urban
Mixed
Water
  3 to 10 acres
  10 to 40 acres
  over 40 acres

Native pasture
Other pasture
Native Hay

Perenarial streams (66 to 660 feet)
Rivers (660)
Dry streambeds

Wetlands

Idle grassland
Brush
Strip mines, quaries, gravel pits
Transitional
Farmsteads

Sample Size

Experience obtained in Kansas indicated that certain land covers were not adequately represented in the operational JES. This resulted in insufficient ground data for classifier training and acreage estimation. Forest is an important and extensive land cover in Missouri. In examining results from previous years, it was apparent that the sample allocation for the operational JES did not adequately sample forest land, especially coniferous forests. To provide additional ground data, 67 additional segments from the non-agriculture strata were selected.

NASA obtained low altitude, infrared aerial photography over the additional segments during early spring of 1981 and 1982. Hard-copy prints at eight inches to the mile were produced for each segment. These segments were photointerpreted for the forest, urban, and water, land cover categories. The reason for photointerpreting instead of adding the segments to the operational JES were 1) adding these segments for ground enumeration would burden the enumerators and increase cost, 2) experience in Kansas indicated that enumerators have a difficult time correctly enumerating large non-agricultural segments due to inaccessibility, and 3) the goal was to improve the estimates for forestland which would easily be identified on infrared photography. Using one or two year old photography did not present a problem since forest, urban and water, land covers change slowly.

Table 2. Allocation of Additional Segments

Stratum	Sample Size
1	49
15	1
25	4
35	3
45	4
91	3
92	3

JES Edit

A detailed edit of the JES data was conducted at the Missouri State Statistical Office (SSO). As an aid for the edit, aerial photography of each JES segment was obtained from the Agricultural Stabilization and Conservation Service (ASCS) in Missouri. These photographs were used to verify field boundaries located by the enumerators on older JES photographs. Since some enumerators had a tendency to miss land cover fields in the three to six acre size range, the ASCS photography allowed the editor to break out these additional fields.

Multiple-Date Ground Data

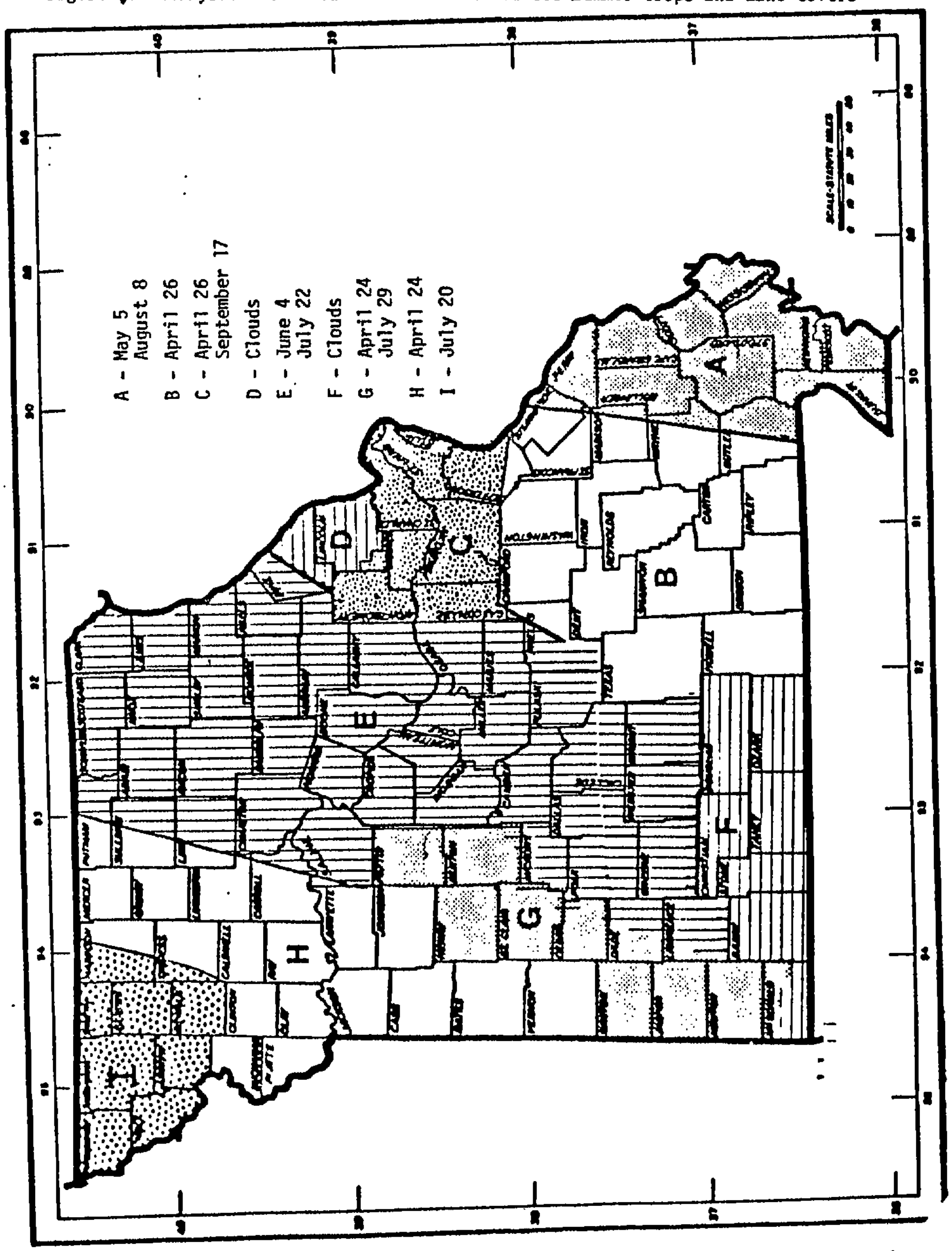
Since two dates of Landsat data were being used for this study it was necessary to maintain a ground data set with two observations (visits) for each field within a segment. Visit 1 corresponded to the ground cover that would appear first during the crop year. Vist 2 corresponded to the cover that would exist second if different from visit one.

LANDSAT DATA

Two dates of Landsat data were used to 1) enable the estimation of crop acreages for a spring crop (winter what) and fall crops (corn, soybeans, rice, cotton), and 2) improve land cover classification results. Only spring imagery was used to produce Landsat regression estimates of winter wheat acreage.

Figure 1. Analysis Districts and Landsat Dates for Summer Crops and Land Covers

Figure 1. Analysis Districts and Landsat Dates for Summer Crops and Land Covers



For the summer-planted crops and the land-cover categories, two dates of imagery were combined to make up the Landsat data set wherever possible. This multitemporal data set was created by overlaying the fall imagery onto the spring imagery (6). The multitemporal data set contained eight channels of Landsat data for each pixel. The first four channels were the reflectance values from the spring date, the second four channels were the values from the fall date. Figure 1 shows the analysis districts used for all covers other than winter wheat. An analysis district is an area of land-covered by Landsat imagery from the same overpass date or combination of dates. Notice that in Figure 1, areas A, C, E and G were covered with multitemporal data. Area I had only fall data, areas B and H had only spring data, and areas D and F had no Landsat coverage.This meant that direct expansion estimates were used for areas B, D, F and H for acreage estimates of corn, soybeans and rice. Direct expansion estimates were used for areas D and F for all land covers estimates.

LANDSAT ANALYSIS

SRS Landsat analysis procedure consists of three primary steps: classifier development, classification-estimation, and accumulation.

Classifier Development

After the Landsat data and the ground data were put in computer-readable form and registered to each other, the segment field boundaries are located in the Landsat digital data. This results in a set of pixels labeled by cover type. When a field is doubled cropped (e.g., winter wheat followed by corn) the double cropping is considered to be a separate cover type. The pixels for each cover are then clustered using the Classy clustering algorithm (7). This produces several spectral signatures (categories) for each cover. Each spectral signature consists of the mean vector and the covariance matrix of the reflectance values for each category. The statistics for all categories and cover types are then reviewed and combined to form the discriminant functions for a Gaussian Maximum Likelihood classifier for each analysis district (8).

Classification Estimation

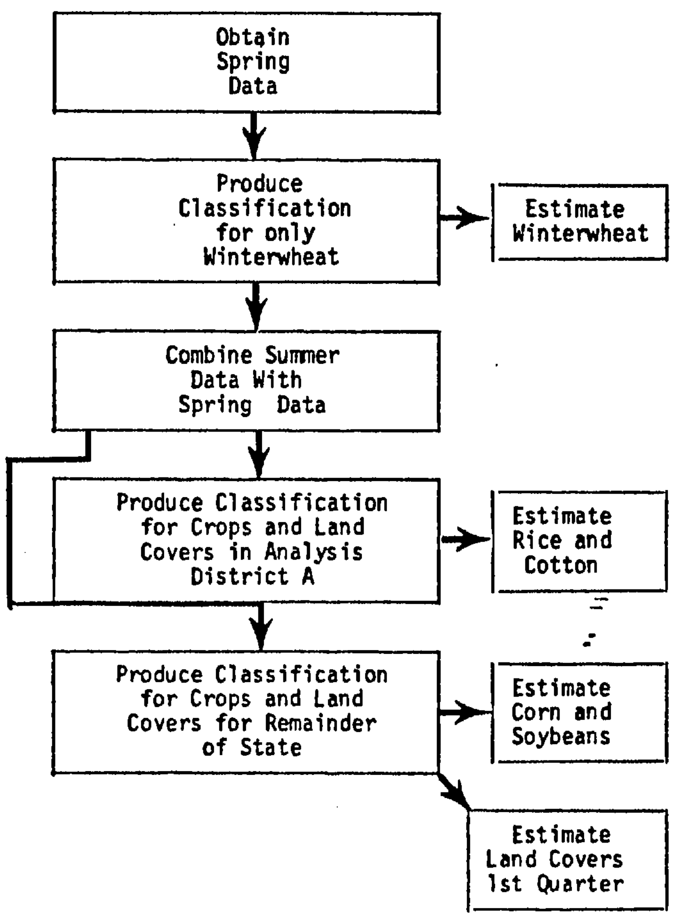
To reduce processing cost, the classification-estimation is done in two stages for each analysis district--small-scale and full-frame. In small-scale processing each pixel associated with a segment is classified to a category. The number of pixels classified to each category are summed to segment totals by cover type. These totals are used as the independent (auxiliary) variable in a regression estimator. Correspondingly, the reported acreages are summed to segment totals and used as the dependent variable. The segment totals are used to calculate least squares estimates for the parameters of the single-variable regression estimator. A separate regression estimator for reported acreage is developed for each crop or land cover for each stratum.

In full-frame processing each pixel in the Landsat scene is classified with the classifier selected from small-scale processsing. The classified results are then tabulated by category and stratum. For each cover used in small-scale processing, the category totals are sumed to stratum totals. From these tabulations, population averages of the number of pixels per segment by stratum are calculated. Using the population averages a regression estimate for the acreage of each crop or land cover is made for each stratum. The stratum estimates are then summed to an analysis district estimate.

Accumulation

In general accumulation is the process by which a direct expansion estimate is made for all areas for which a

Figure 2. Landsat Data Processing and Estimation Steps



regression estimate does not exist. This direct expansion estimate is, then summed with the analysis district regression estimates to obtrain a state-level acreage estimate.

Overall Approach

Two Landsat analyses were completed for the study (Figure 2). First, a complete analysis was conducted using spring Landsat data to estimate winter wheat acreages. Second, the summer crop analysis was conducted using the multitemporal or fall data. To reduce processing, crop analysis was conducted in a manner such that classifier development and classification, included all land cover categories. Analysis district A (Figure 1) was the first area analyzed since it contains nearly all of Missouri's rice and cotton. After this analysis was completed state-level estimates for corn and soybeans were produced. After the analysis of the areas used for crop acrage estimates were completed and estimates delivered to the CRB, analsyis districts B and H were analyzed for land covers only. Land cover estimates were then calculated for all analysis districts shown in Figure 1 and state-level estimates produced.

Table 3. Land Cover Direct Expansion Estimates

	Direct Expansion Estimates		
Cover	Estimate	Standard Deviation	C.V.%
Hardwood	10,449,754	529,061	5.0
Conifer	181,568	43,325	23.9
Conifer-Hardwood	1,149,738	247,934	21.6
Grazed Forest	2,884,732	297,743	10.3
Brushland	1,286,435	143,382	11.1
Row crops	8,539,851	361,734	4.2
Sown crops	2,391,119	175,337	7.3
Idle/cropland	2,100,277	163,574	7.8
Hay	3,110,286	197,393	6.4
Cropland/pasture	1,434,850	234,325	16.3
Other pasture	7,698,684	423,699	5.5
Idle grassland	1,403,300	140,411	10.0
Farmsteads	385,091	23,474	6.1
Residential	962,910	105,045	10.9
Commercial	328,253	81,590	24.9
Other urban	140,229	39,114	27.9
Transportation	296,577	53,422	18.0
Lakes	307,755	118,936	38.7
Ponds	84,270	17,563	20.8
Rivers	129,922	43,887	33.8
Disturbed land	44,223	17,741	40.1
Transitional	183,379	137,668	75.0
Wetlands	106,830	87,386	81.8

*Fields that are double cropped are included in the estimates for each crop

RESULTS

The direct expansion and regression estimates for land covers are listed in Tables 3 and 4. Two land covers have large acreage differences between their respective types of estimates. The regression estimate for row crops is 798,000 acres less than the direct expansion estimate. A major portion of the decrease in row crops came from stratum 10 which had 380,000 fewer acres in the regression estimate.

Relative efficiency (R.E.) measures of the improved precision of the regression estimate. The ratio of the variance of the direct expansion (which uses JES data only) to the variance of the Landsat regression estimate defines this R.E. (1) Equivalently, it is the factor by which the sample would have to be increased to produce direct expansion estimates with the same precision as the regression estimates.

Table 4. Land Cover Regression Estimates

	Regression Estimates			
Cover	Estimate	Standard Deviation	C.V.%	R.E.
Hardwood	11,139,532	443,461	4.0	1.4
Conifer	187,650	21,782	11.6	4.0
Conifer-Hardwood	1,148,447	245,461	21.4	1.0
Grazed Forest	2,705,512	299,958	11.1	1.0
Brushland	1,318,875	138,723	10.5	1.1
Row crops	7,742,383	246,344	3.2	2.2
Sown crops	2,547,815	127,349	5.0	1.9
Idle/cropland	2,015,582	139,389	6.9	1.4
Hay	2,980,606	171,303	5.7	1.3
Cropland/pasture	1,245,797	149,895	12.0	2.4
Other pasture	7,624,049	380,381	5.0	1.2
Idle grassland	1,331,205	133,127	10.0	1.1
Farmsteads	387,434	23,515	6.0	1.0
Residential	823,018	95,629	11.6	1.2
Commercial	305,556	41,463	13.6	3.9
Other urban	122,873	30,365	24,7	1.7
Transportation	288,724	53,398	18.5	1.0
Lakes	265,246	108,556	40.9	1.2
Ponds	84,438	13,130	15.6	1.8
Rivers	103,729	23,368	22.5	3.5
Disturbed land	42,455	16,020	37.7	1.2
Transitional	-	-	-	-
Wetlands	-	-	-	-

*Fields that are double cropped are included in the estimates for each crop

The hardwood regression estimate increased by 640,000 acres. Tables 5 compares the hardwoods direct expansion and regression estimates by analysis area. The major difference between the two estimates is found in analysis district B (Figure 1). A breakdown by strata for the area indicates that most of the increase is in stratum 45 which has a direct expansion estimate of 541,797 and a regression estimate of 1,134,508. The r-square for the regression is .80 and the stratum contains six segments of which five are heavily forested.

Table 5. Comparison of Direct Expansion and Regression Estimates for Hardwoods by Analysis District

	Direct Expansion		Regression	
Analysis Area	Acres	C.V%	Acres	C.V.%
A	698,852	18.4	655,439	19.2
B	3,604,987	8.0	4,219,075	4.0
C	632,631	18.7	744,267	6.3
D/F	1,063,890	20.2	1,063,890	20.2
E	2,917,765	9.1	2,872,497	6.9
G	574,552	17.7	666,814	11.2
H	816,498	13.2	752,548	10.7
I	208,136	21.2	164,998	19.0
State Total	10,499,754	5.0	11,139,532	4.0

REFERENCES

1 Hanuschak, George A., Michael Craig, Martin Ozga, Raymond Luebbe, Paul Cook, David Kleweno, and Charles Miller. "Obtaining Timely Crop Area Estimates Using Ground-Gathered and Landsat Data." U.S. Dept. of Agr. ESCS Technical Bulletin No. 1609, August 1979.

2 Winings, Sherman B., Paul W. Cook, George A. Hanuschak. "AgRISTARS DCLC Applications Project 1982 Corn and Soybeans Area Estimates for Iowa, Illinois." Seventh International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, May 1983.

3 Mergerson, J.W., V.B. Johnson, R.A. Kestle. "AgRISTARS DCLC Applications Project: 1982 Winter Wheat Area Estimates for Colorado, Kansas and Oklahoma." Ninth International Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Laboratory for Applications of Remote Sensing, June 1983.

4 May, G., M. Holko, and J. Anderson. 1983. Classification and area estimation of land covers in Kansas, using ground gathered and Landsat digital data. DC-Y3-0441,NSTL/ERL-225, NASA,NSTL,MS. 10 pp.

6 Ozga, Martin, Sue R. Faerman, Richard S. Sigman. "Editor Multitemporal System." U.S. Dept. of Agr., ESCS, November 1979.

7 Lennington, R.K., and M.E. Rassback. "Classy - An Adaptive Maximum Likelihood Clustering Algorithm." Proceeding of the LACIE Symposium, NASA Manned Spacecraft Center, Houston, Texas, Jan. 1972.

8 Swain, Philip H., and Shirley M. Davis. "Remote Sensing the Quantitative Approach." McGraw-Hill, Inc., New York, 1978.

9 Spencer, J., and B. Essex. Timber in Missouri 1972. pp. 108. USDA Forest Service Resource Bulletin NC-30. North Central Forest Experiment Station, St. Paul, MN, 1976.