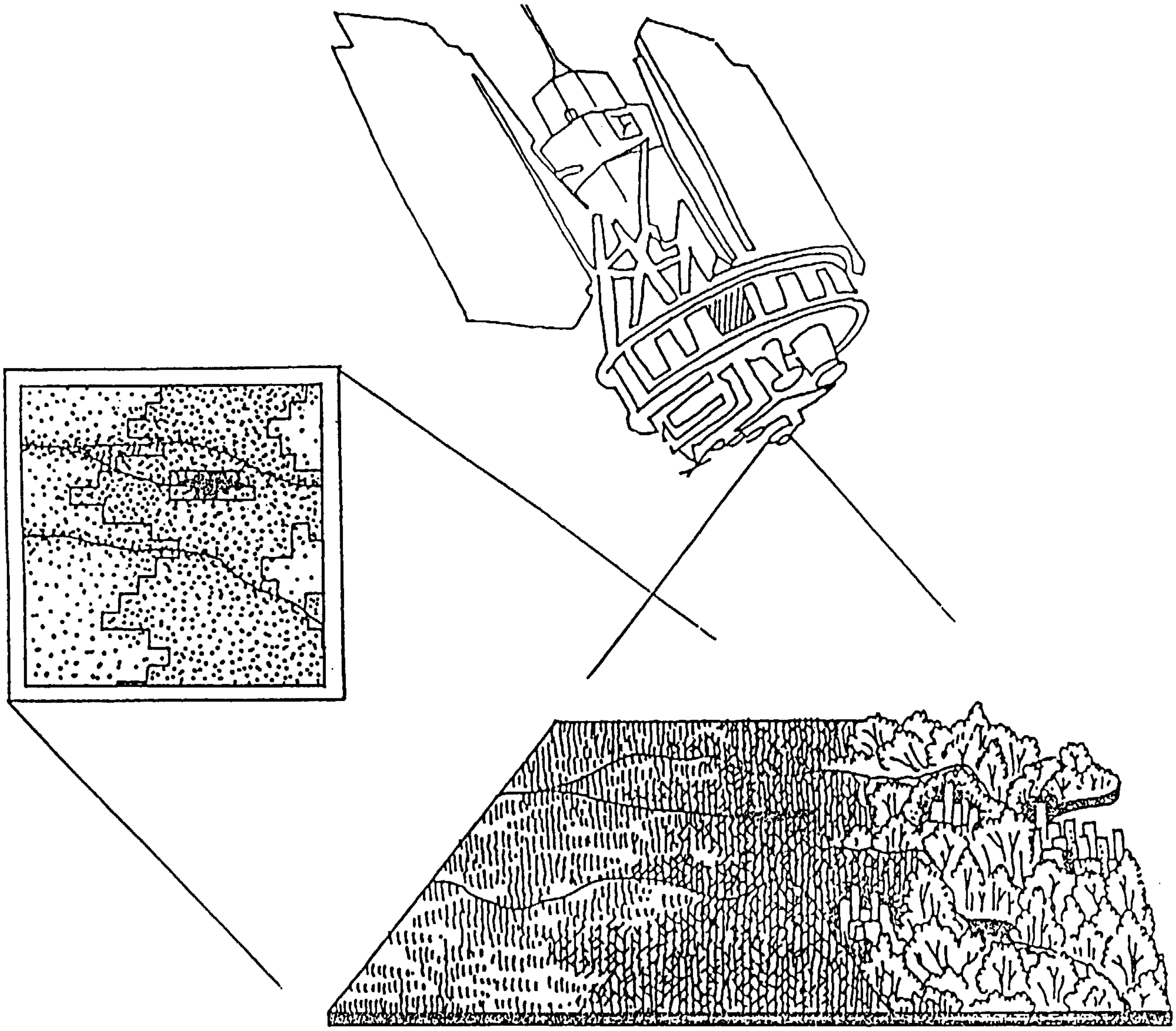


TITLE: REMOTE SENSING BRANCH HANDOUTS

for

STATISTICAL RESEARCH DIVISION REVIEW
JANUARY 29, 1986

Handout	Title
1	Methodology paper
2	Multitemporal vs. unitemporal study
3	Estimator differences: JES vs. DCLC
4	Summary of Lundgren simulation study
5	Possible bias explanation
6	Supporting research
7	SRS/RSB publications
8	NASA publications
9	University publications
10	New York orchard inventory
11	Digital frame development
12	Crop-condition research

THE ROLE OF LANDSAT DATA IN IMPROVING U.S. CROP STATISTICS*

Martin L. Holko

Richard S. Sigman

U.S. Department of Agriculture
Statistical Reporting Service
Washington, D.C., USA

ABSTRACT

Landsat data are used in two ways to improve U.S. crop statistics. Landsat color-composite images are used to stratify areas of land with regard to land use. This stratification is used as a technique to improve the efficiency of an area sampling frame. Also, Landsat digital data are classified and the classified results are used as supplementary information to an agricultural survey. The combination of Landsat classification results and survey data improves the precision of the estimates made.

1.0 Introduction

The Statistical Reporting Service (SRS) is the agency of the U.S. Department of Agriculture responsible for current statistics describing domestic crop and livestock production. For the most part, these statistics are estimates based on sample surveys conducted by SRS personnel.

A major source of data for SRS is its nationwide June Enumerative Survey (JES). It is in conjunction with the JES that SRS uses data from the Landsat satellites. Landsat data are used to improve the precision of the estimates obtained from the JES in two different ways. One use of Landsat data is in the development of an area sampling frame from which the JES sample is selected. A second use is as current, supplemental information that, when combined with the data collected during the JES, increases the precision of calculated area estimates.

*Presented at the Eighteenth International Symposium on Remote Sensing of Environment, Paris, France, October 1-5, 1984.

2.0 Use of Landsat Imagery in Area Frame Construction

2.1 Concepts

In area-frame sampling the sample units are pieces of land called segments. The boundaries of segments are well-defined, physical features -- such as roads, footpaths, rivers, and railways -- that can be both delineated on maps and aerial photographs and also readily identified by data collection personnel in the field. An area-sampling frame is a complete list (or more frequently a set of specifications that would generate a complete list) of segments that cover a geographical area of interest, such as a state or province. This geographical area of interest is called a population.

An area sampling frame is a basic tool for collecting agricultural statistics. It is used in a number of countries to estimate acreage and yield of agricultural products as well as farm-economics parameters such as prices and labor for the current year. Area frame sampling provides accurate information by taking representative samples from only a small portion of the total land area. Estimates can be available five to six weeks after the beginning of data collection.

The construction of an area sampling frame consists of several steps [Houseman, 1975]. The first step is the delineation on a base map of stratum blocks. These are large contiguous areas of homogeneous land use. In addition to the mapping symbols on the base map, information from satellite imagery, aerial photography, and other maps are used in this stratification step. All of the stratum blocks of the same land use constitute a stratum. Like segment boundaries, the delineated strata boundaries must be identifiable in the field. The purpose of stratification is to increase the precision of sample survey estimates.

The next step is to divide the strata blocks into smaller areas called primary sampling units (PSU's). The PSU's vary in size depending on the stratum but generally contain from 5 to 20 potential segments. Out of each stratum a suitable number of PSU's will be randomly chosen with probability of selection proportional to the area of the PSU.

The purpose of the PSU's is to serve as an intermediate delineation between the large strata blocks and the individual segments. By delineating PSU's all of the segments in the population need not be delineated. Instead, only the segments in the randomly selected PSU's are delineated by subdividing the PSU into the appropriate number of segments based on the area of the PSU and the target segment size. In strata that are predominantly cultivated land, the target segment size is typically one square mile. After the selected PSU has been subdivided, one segment is randomly selected from the PSU for field enumeration.

Desired data are then collected from the sample segments by interviewing farmers who operate land inside the segment. Since the segments within each stratum are statistically representative of the stratum, the data collected from the segments can be expanded to the total area of the stratum. The desired estimate for the entire population is then obtained by summing the results for each stratum.

2.2 SRS Experience

SRS has constructed and maintains an area frame for each of the 48 contiguous states. Since the construction of an area frame for a state is a major effort, SRS is only able to construct approximately three new area frames per year. Once an area frame for a state is constructed, it is used annually for anywhere from 10 to 20 years before it is revised or replaced.

The majority of SRS's area frames contain five basic strata: cultivated land, range and pasture, water, nonagricultural land, and cities and towns. The cultivated land in most states is further stratified by separating "intensively" cultivated land from "extensively" cultivated land. (In Nebraska there are two intensively-cultivated-land strata.) In addition to the five basic land-use strata, the area frames in California and Texas each contain one or more "crop specific" strata. The SRS area frames in Washington, Oregon, and Idaho have strata for dryland grain. [Geuder, 1984]

The use of Landsat imagery to stratify SRS area sampling frames was first demonstrated by Hanuschak and Morrisey [1977]. In this study, county maps at a scale of 1:126,720 were photographically reduced to a scale of 1:250,000 on mylar and overlaid on 1:250,000-scale, color Landsat imagery produced on paper by the EROS Data Center. The Landsat image was photo-interpreted to provide land-use information, whereas the overlaid county map provided physical features for delineating stratum blocks and PSU's. This procedure was then used by SRS in 1979 to construct a new area frame for the state of California [Fecso and Johnson, 1981]. Since 1979, SRS has photo-interpreted Landsat images for constructing new area frames in Arizona, Colorado, Florida, Idaho, New Mexico, Oregon, Texas, Washington, and Wyoming. The majority of these new frames have been in the western United States where much of the cultivated land is irrigated and can thus be readily identified on Landsat images.

In 1982, SRS updated the Nebraska area frame by restratifying the urban stratum and areas where rangeland had been converted to cropland. Used in this restratification effort were plots giving the location of all pivot irrigation in 58 counties. These plots were developed by the University of Nebraska from Landsat data, administrative records for well permits, and field observations by county agents. [Hale, 1983]

Burns [1983] has demonstrated the use of digital Landsat data for updating SRS sampling frames in an area in Louisiana. In this study, unsupervised clustering of the Landsat data was performed, and then stratum labels were assigned to the clusters by an analyst using an interactive image processing system. SRS is further evaluating this procedure for stratifying area sample frames in Wyoming and Florida [Geuder, Blackwood, and Radenz; 1983].

3.0 Landsat Data as Supplemental Information

3.1 Background

SRS conducts the JES annually in late May and early June. The JES survey procedure requires that information be obtained for all the land within each of the sampled segments. To insure that all the land is accounted for, aerial photographs, at a scale of 1:8,000, are used as an enumeration aid. The boundaries for each segment are drawn on individual non-current photographic prints. These segment photographs and corresponding questionnaires are sent to field enumerators for data collection. As part of the data collection procedure, each enumerator is instructed to draw the boundaries of all fields, within each segment, on the segment photograph (a field is defined as a continuous block of land containing the same crop or land cover). On the corresponding questionnaire the enumerator records the cover and size of each field, as well as livestock numbers and other agricultural information obtained from the operator. The information collected during the JES is aggregated to the segment level and direct expansion estimates are then calculated to obtain state level estimates for crop hectares. The formulas for the direct-expansion estimator and its variance are as follows:

Let $\hat{Y}_c$ = the direct expansion estimate for the hectares of crop c

$$\hat{Y}_c = \sum_{s=1}^{S} \frac{N_s}{n_s} \sum_{j=1}^{n_s} y_{jsc}$$

where:

y_{jsc} = the hectares reported to crop **c**, in segment **j**, for strata **s**

n_s = number of segments sampled in strata **s**

N_s = the total number of potential segments in stratum **s**

S = the total number of strata

The estimated variance is:

$$V(\hat{Y}) = \sum_{s=1}^{S} \frac{(N_s - n_s)\ N_s}{n_s\ (n_s - 1)} \sum_{j=1}^{n_s} (y_{jsc} - y_{.sc})^2$$

where:

$$y_{.sc} = \sum_{j=1}^{n_s} \frac{y_{jsc}}{n_s}$$

In 1972 SRS personnel started to investigate the potential of using <u>digital</u> Landsat data to improve the precision of the estimates obtained from the JES. The procedure developed consists of the following steps:

- Analysis District Selection: Landsat data are selected and boundaries of Landsat analysis districts defined.

- Signature Development: Data collected during the JES and corresponding Landsat data are used to develop a maximum likelihood classifier for each analysis district.

- Small Scale Processing: The Landsat pixels representing the area within each segment contained in an analysis district are classified. A relationship is developed between the number of pixels classified to a crop and the hectares recorded for that crop on the JES.

- Full Frame Processing: All of the Landsat pixels within the analysis district are classified. Estimates are calculated at the analysis district level by applying each crop regression relationship to the all-pixel classification results.

- State Level Accumulation: The estimates for all analysis districts are combined to create a state level estimate for each crop of interest.

3.2 Analysis District Selection

An analysis district is an area of land covered by Landsat imagery of the same overpass date. A separate Landsat analysis is done for each analysis district. Depending on the location and availability of Landsat data, each state is divided into a number of analysis districts. The Landsat analysis district location is treated as a geographical post-stratification imposed on the original area frame. As a result of this post-stratification, SRS personnel must determine the number of frame units and the sampled segments which fall into each post-stratum. This results in two types of strata categories:

1) The first stratum category corresponds to the area of the state for which there is no Landsat coverage. This area may be non-contiguous. The portion of each land-use stratum within these geographical areas makes up the post-strata. We let

M_s = the total number of segments in the non-Landsat area in land use strata **s**, and

m_s = the number of sampled segments in the non-Landsat area in land use strata **s**.

2) The second stratum category corresponds to the areas of the state where the land-use strata and the analysis districts are defined. In these areas each stratum consists of the area of intersection between the land use strata and a Landsat analysis district. Here, we let

M'_{as} = the number of frame units in analysis district **a**, land use strata **s**, and

m'_{as} = the number of sampled segments in analysis district **a**, land use strata **s**.

3.3 Signature Development

Signature development is done independently for each analysis district and consists of four phases. The first phase is segment calibration and digitization. Segment calibration is a first-order linear transformation which maps points on the segment photograph to a map base (in our application this map base is the U.S. Geological Surveys quadrangle map series, which uses the latitude/longitude coordinate system of reference). Segment digitization is the process by which field boundaries drawn on the segment photograph are recorded in computer-compatible form. The combined process of calibration and digitization gives us the capability of digitally locating every JES field relative to a map base.

The next phase in signature development is the registration of each Landsat scene. SRS's Landsat registration process is a third-order linear transformation that maps each Landsat pixel within a scene to a map base [Cook, 1982]. Corresponding points selected on a two-degree map and a 1:250,000 Landsat image are used to generate this mathematical transformation. The combination of segment calibration, digitization and Landsat registration provides the capability to locate each JES segment in its corresponding Landsat scene (to within about 5 pixels of the correct location). Since this registration is not accurate enough for selecting training data, line plots of segment field boundaries and corresponding greyscale prints are overlaid and each segment is manually located to within 1/2 pixel of the correct location. With this process we are able to accurately identify all of the pixels associated with any JES field. The result of this is a set of pixels labeled by JES cover.

The third phase of signature development is supervised clustering. In supervised clustering all of the pixels for each cover are processed through one of two available clustering algorithms: Classy or Ordinary Clustering. Classy is a maximum likelihood clustering algorithm developed at Johnson Space Center in Houston, Texas [Lennington and Rassback, 1972]. Ordinary Clustering is an algorithm derived from the ISODATA algorithm of Ball and Hall [1967]. Each clustering algorithm generates several spectral signatures (categories) for each cover. Each spectral signature consists of a mean vector and the covariance matrix for the reflectance values for that category.

In the fourth phase, the statistics for all categories from all covers are reviewed and combined to form the discriminant functions of the maximum likelihood classifier. The formulas for the discriminant functions are as follows:

The maximum likelihood classifier with equal priors:

Classify pixel **k** to category **c** if $D_{ck} \geq D_{ik}$ for all $i \neq c$

The maximum likelihood classifier with priors:

Classify pixel **k** to category **c** if $D^p_{ck} \geq D^p_{ik}$ for all $i \neq c$

where:

$D_{ik} = -\log_e(|Z_i|) - (X_k - U_i)' Z_i^{-1} (X_k - U_i)$

$D^p_{ik} = D_{ik} + \log(p_i)$

U_i = the mean vector for category **i**

Z_i = the covariance matrix for category **i**

p_i = the prior probability for category **i**

X_k = the reflectance value for pixel **k**

3.4 Small Scale Processing

In small-scale processing each pixel associated with a JES segment is classified to a category. This classification is usually done using both the classifier with priors and the equal priors classifier. For each classifier, pixels classified to each category are summed to segment totals. The category totals corresponding to crops of interest are summed to segment crop totals. These crop totals are used as the independent variable in a regression estimator. Correspondingly, the hectares reported on the JES for each crop 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 a linear regression. Two sets of regression equations are developed for each crop within each stratum (one for the classification with priors, one for the classification with equal priors).

The linear regression equations for analysis district **a**, strata **s**, and crop **c** are of the form:

$$y_{jasc} = b_{0asc} + b_{1asc}\, x_{jasc}$$

where:

y_{jasc} = the reported hectares of crop **c**, from segment **j**, analysis district **a**, land use stratum **s**

x_{jasc} = the crop total classification for segment **j**, analysis district **a**, land use strata **s**

b_{0asc}, b_{1asc} = least squared estimates of the regression parameters for crop **c**, analysis district **a**, land use strata **s**

3.5 Full Frame Processing

The regression equations developed in small-scale processing are evaluated and the classifier giving the best overall regression relationship is selected. This classifier is used to classify every pixel in the analysis district. The classified results are tabulated by category and land-use stratum. For each crop of interest the category totals are summed to stratum crop totals. From these totals the population averages per segment are calculated. Using the population average, a stratum-level regression estimate is made for that analysis district for each crop.

Let $\hat{Y}'_{asc}$ be the analysis district level regression estimator for crop **c** and stratum **s**.

Then:

$$\hat{Y}'_{asc} = M'_{as}\,[y_{.asc} + b_{isc}(X_{.asc} - x_{.asc})]$$

where:

$$y_{.asc} = \sum_{j=1}^{m'_{as}} \frac{y_{jasc}}{m'_{as}} \quad \text{and} \quad x_{.asc} = \sum_{j=1}^{m'_{as}} \frac{x_{jasc}}{m'_{as}}$$

M'_{as} = previously defined (3.2)

m'_{as} = previously defined (3.2)

x_{jasc} = previously defined (3.4)

y_{jasc} = previously defined (3.4)

$X_{.asc}$ = the population average for crop **c** in analysis district **a** land use stratum **s**

The estimated variance is:

$$V(\hat{Y}'_{asc}) = \frac{(m'_{as}-1)}{(m'_{as}-2)}\,(1-r_{asc}{}^2)\,\frac{(M'_{as}-m'_{as})M'_{as}}{m'_{as}(m'_{as}-1)} \sum_{j=1}^{m'_{as}} (y_{jasc}-y_{.asc})^2$$

where:

r^2_{asc} = the sample correlation between y_{jasc} and x_{jasc}

3.6 State Level Accumulation

The final step of our Landsat analysis is the combining of all of the estimates (one for each post strata) into a state-level estimate of the area of the desired crop.

Let $\hat{Y}'_c$ be the final state level estimate for the hectares of crop **c**.

Then:

$$\hat{Y}'_c = \sum_{a=1}^{A} \sum_{s=1}^{S'_a} \hat{Y}'_{asc} + \sum_{l=1}^{L} M_l \, y_{.lc}$$

where:

$$y_{.lc} = \sum_{j=1}^{m_l} \frac{y_{jlc}}{m_l}$$

M_l, m_l previously defined (3.2)

$\hat{Y}'_{asc}$ is as defined earlier (3.5)

y_{jlc} = the hectares reported to crop **c** for segment **j** in the non-Landsat post strata l

S'_a = The number of land use strata in analysis district **a**

A = The number of analysis districts

L = The number of land use strata that exist in the area where wedo not have Landsat coverage

The estimated variance is:

$$V(\hat{Y}'_c) = \sum_{a=1}^{A} \sum_{s=1}^{S'} V(\hat{Y}'_{asc}) + \sum_{l=1}^{L} \frac{(M_l - m_l)M_l}{m_l(m_l - 1)} \sum_{j=1}^{m_l} (y_{jlc} - y_{.lc})^2$$

3.7 Evaluation of the Landsat Estimate

Landsat data are used as supplemental information to improve the precision of the area estimates obtained from the JES. Unlike area frame construction, the effectiveness of this use of Landsat data can be measured. The measure used is the efficiency of the Landsat estimator relative to the JES direct expansion estimator. This relative efficiency (RE) is defined as the ratio of the variance of the direct expansion to the variance of the Landsat estimate. Equivalently, this is the factor by which the sample size would have to be increased to produce a direct expansion estimate with the same precision as the Landsat estimate.

$$RE = \frac{V(\hat{Y}_c)}{V(\hat{Y}'_c)}$$

3.8 Implementation

The basic concepts of SRS's Landsat analysis were developed during the 1972-1979 time period. In 1980 as part of the AgRISTARS Domestic Crop and Land

Cover Project, SRS's Remote Sensing Branch began making current-year, state-level area estimates for winter wheat, corn and soybeans in selected states. This move to a pseudo-operational mode meant that current year Landsat data (May for winter wheat, August for corn and soybeans) had to be processed to produce estimates by late-November and late-December for winter wheat and corn/soybeans respectively. The original implementation plan called for including two states in 1980 and adding two more states each year to a total of 10 states by 1984. In 1980 winter wheat estimates were produced for Kansas, corn and soybean estimates for Iowa. Table 1 shows the states included in the project, the crops for which estimates were made, and the number of Landsat scenes needed to cover each state. In 1983, SRS deviated from the original plan by adding only one state to the project. No new states were added in 1984. These modifications were necessary due to personnel ceilings and limitations of current processing capabilities. In 1984, under the modified plan, SRS expects to process about 2,000 JES segments contained in 66 Landsat scenes covering most of seven states (Table I).

3.9 Results

The JES direct expansion and Landsat estimates are two of many indications used to determine the official USDA area estimates. For most major crops the JES direct expansion is the key indication used for setting the preliminary area estimates in July. The Landsat estimates for the states in the project (available at the end of the crop year) are reviewed when the final end-of-season estimates are made.

Tables II through VI show the JES direct expansion, the Landsat estimates and the final USDA estimates. The relative efficiencies of the Landsat estimates are mostly in the range from 1.2 to 2.0 for the major crops of winter wheat, corn and soybeans. The relative efficiencies for crops with fewer hectares such as cotton and rice are considerably better. The level of some of the estimates for cotton and rice, however, differ considerably from other data sources used to make the official estimate. Part of the variability in the relative efficiencies for the major crops can be explained by the amount of Landsat coverage available to do each estimate. Figure 1 shows three graphs comparing the percent of each crop covered by Landsat data with the relative efficiency obtained. If the trend apparent in these graphs can be extended, one would expect that the best we could do is relative efficiencies of about 2.5. These results, although promising, are not as good as originally expected. However the continued personnel limitation and the increasing respondent burden being placed on our farm sector may make our Landsat estimator one of few techniques feasible for improving crop statistics in the U.S.

References

Ball, G.H. and D.J. Hall. A Clustering Technique for Summarizing Multivariate data. Behavioral Science, 12 153-155, March 1967.

Burns, G.S. Land Cover Change Monitoring Within the East Central Louisiana Study Site -- A Case for Large Scale Area Surveys with Landsat Multispectral Scanner Data. Technical report, DC-Y3-04418, NSTL/ERL-221. NASA, NSTL Station, Mississippi, June 1983.

Cook, P.W. LANDSAT Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service 1972-1982. 1982 American Statistical Association Meetings, Cincinnati, Ohio, August 1982.

Fecso, Ron and Van Johnson. The New California Area Frame: A Statistical Study. Report SRS-22. United States Department of Agriculture, Statistical Reporting Service, Washington, D.C., September 1981.

Geuder, Jeff; Paul Blackwood, and Ron Radenz. Using Landsat for Sampling Frame Development. Research proposal. Statistical Reporting Service, October 1983.

Geuder, Jeff. Paper stratification in SRS area sampling frames. SF & SRB Staff Report No. 79. Statistical Reporting Service, Washington, D.C., February 1984.

Hale, Bob. An Evaluation of the Nebraska Area Frame. SF & SRB working paper. Statistical Reporting Service, November 1983.

Hanuschak, George and Kathleen Morrissey. Pilot Study of the Potential Contributions of Landsat Data in the Construction of Area Sampling Frames. Staff report. Statistical Reporting Service, Washington, D.C., October 1977.

Houseman, Earl. Area Frame Sampling in Agriculture. Report SRS-20. Statistical Reporting Service, Washington, D.C., November 1975.

Table I: States and Crops for Which Landsat Area Estimates Have Been Made

State	Years in Project	Area Estimates Produced for:	Number of Landsat Scenes Needed:
Kansas	1980, '81, '82, '83, '84	winter wheat	16
Iowa	1980, '81, '82, '83, '84	corn, soybeans	12
Oklahoma*	1981, '82, '83, '84	winter wheat	7
Missouri*	1981, 1/, '83, '84	winter wheat, corn, soybeans, cotton, rice	12
Colorado*	1982, '83, '84	winter wheat	14
Illinois	1982, '83, '84	corn, soybeans	10
Arkansas*	1983, '84	soybeans, rice, cotton	5
TOTAL			66

* major producing areas

Table II: Area Estimates for Winter Wheat Harvested by State and Year

State/Year	JES Direct Expansion Estimate	JES Direct Expansion Standard Error	Landsat Regression Estimate	Landsat Regression Standard Error	Relative Efficiency	USDA Estimate
	(1,000 hectares)		(1,000 hectares)			(1,000 hectares)
Kansas						
1980	5,214	162	5,051	136	1.3	4,856
1981	5,452	158	5,298	104	2.3	4,897
1982	5,677	167	5,611	120	1.9	5,301
1983	4,652	153	4,477	124	1.5	4,371
Oklahoma						
1981	2,612	117	2,483	101	1.4	2,590
1982	2,914	119	2,660	90	1.8	2,792
1983	1,725	85	1,688	74	1.3	1,740
Colorado						
1982	1,276	91	1,132	49	3.4	1,178
1983	1,193	115	1,110	81	2.0	1,214
Missouri						
1983	830	66	866	49	1.9	749

Table III: Area Estimates for Corn by State and Year

State/Year	JES Direct Expansion Estimate	JES Direct Expansion Standard Error	Landsat Regression Estimate	Landsat Regression Standard Error	Relative Efficiency	USDA Estimate
	(1,000 hectares)		(1,000 hectares)			(1,000 hectares)
Iowa						
1980	5,735	115	5,801	93	**1.9**	5,666
1981	5,828	128	5,820	103	**1.6**	5,828
1982	5,601	118	5,568	113	**1.1**	5,565
1983	3,708	111	3,666	81	**1.8**	3,683
Missouri						
1981	870	75	775	51	**2.2**	850
1982 1/	-	-	-	-	-	-
1983	758	60	629	45	**1.8**	688
Illinois						
1982	4,809	115	4,677	106	**1.2**	4,735
1983	3,482	113	3,380	102	**1.2**	3,318

Table IV: Area Estimates for Soybeans by State and Year

State/Year	JES Direct Expansion Estimate	JES Direct Expansion Standard Error	Landsat Regression Estimate	Landsat Regression Standard Error	Relative Efficiency	USDA Estimate
	(1,000 hectares)		(1,000 hectares)			(1,000 hectares)
Iowa						
1980	3,395	112	3,290	96	**1.5**	3,359
1981	3,260	104	3,275	82	**1.6**	3,278
1982	3,539	106	3,433	99	**1.2**	3,428
1983	3,155	98	3,200	88	**1.3**	3,238
Missouri						
1981	2,306	115	1,964	86	**2.1**	2,072
1982 1/	-	-	-	-	-	-
1983	2,275	124	2,008	97	**1.6**	2,104
Illinois						
1982	3,866	120	3,767	109	**1.2**	3,743
1983	3,696	107	3,669	99	**1.2**	3,602
Arkansas						
1983	1,661	78	1,565	70	**1.3**	1,578

Table V: Area Estimates for Rice by State and Year

	JES Direct Expansion		Landsat Regression			
State/Year	Estimate	Standard Error	Estimate	Standard Error	Relative Efficiency	USDA Estimate
	(1,000 hectares)		(1,000 hectares)			(1,000 hectares)
Missouri						
1981	47	20	31	10	6.8	31
1982[1/]	-	-	-	-	-	-
1983	51	21	46	10	3.9	25
Arkansas						
1983	419	48	376	32	2.2	374

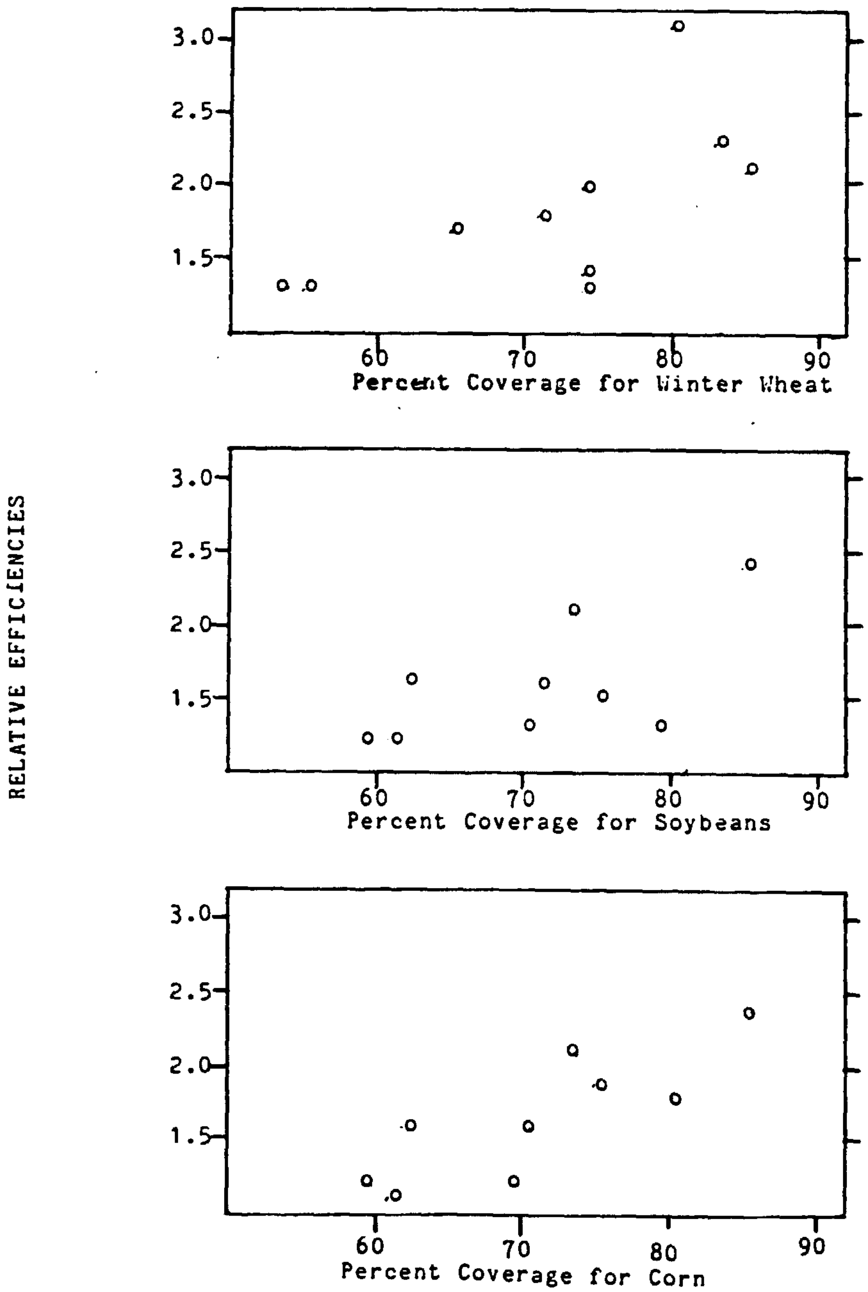
Table VI: Area Estimates for Cotton by State and Year

	JES Direct Expansion		Landsat Regression			
State/Year	Estimate	Standard Error	Estimate	Standard Error	Relative Efficiency	USDA Estimate
	(1,000 hectares)		(1,000 hectares)			(1,000 hectares)
Missouri						
1983	26	15	30	4	11.1	44
Arkansas						
1983	144	33	103[2/]	19	2.9	138

[1/] No Landsat estimates were made for Missouri during 1982 due to insufficient Landsat coverage.

[2/] Arkansas had a lot of cotton that was planted and abandoned prior to the satellite overpass. This area was not included in the Landsat regression estimate.

Figure 1: Plot of Percent of Each Crop Covered by Landsat Data Versus the Relative Efficiency of the Landsat Estimate.



Multitemporal vs. Unitemporal Analysis of MSS Landsat Data on a Full State Basis

by

Sherm Winings

INTRODUCTION

The Statistical Reporting Service (SRS) in its Domestic Crops and Land Cover (DCLC) project is currently using Landsat data combined with ground truth data to provide acreage estimates in seven mid-western states. The ground truth data are collected during the June Enumerative Survey (JES) conducted by SRS in early June. As a part of the DCLC project, Landsat regression acreage estimates for corn, cotton, rice, sorghum, and soybeans in Missouri were presented to the Crop Reporting Board of the Statistical Reporting Service, United States Department of Agriculture on December 15, 1984. Similar estimates for planted and harvested winter wheat acreages were presented on December 5, 1984 to the same board.

Use of Landsat data to produce these estimates implied that both spring (April-May) and summer (July-August) Landsat MSS scenes be analyzed to produce estimates for winter wheat and spring planted crops. A unitemporal approach requires two full analyses; one on the spring data to produce Landsat estimates for fall planted crops and a second on the summer data for spring planted crops. A multitemporal analysis allows a single analysis on the combined spring and summer data. A combination of unitemporal and multitemporal analysis was used for the 1984 crop year because of the earlier due date for the winter wheat estimates, some doubts as to the software efficiency for multitemporal processing, and a desire to make a comparison between unitemporal and multitemporal processing. Unitemporal analysis for winter wheat estimates could be started much earlier than a multitemporal analysis since analysis could begin as soon as the spring Landsat scenes were acquired and the ground truth data were edited in late June. Multitemporal analysis requires that scenes for both dates be in-house before processing can begin. Since many of the spring planted crops were not planted at the time of the primary ground data collection effort in June, an intentions follow-up survey must also be conducted and edited before analysis can proceed to assure accurate ground truth data for estimating acreages of spring planted crops.

The 1984 analysis was done as follows:

1. Unitemporal analysis for winter wheat
2. Multitemporal analysis for all crops
3. Unitemporal analysis for all spring planted crops

The third analysis was done in January after the Crop Reporting Board request was met. Except for registration of scenes and the number of data channels, the analysis procedures for unitemporal and multitemporal data were the same.(1)

REGISTRATION OF LANDSAT DATA

Scene to Map. The spring scenes were designated as the primary scenes and were registered in the usual unitemporal manner.(2) This method has been presented many times and will not be discussed here.

Scene to Scene. When the summer scenes were acquired, 12 to 24 corresponding points were digitized on each scene using features clearly identifiable on both scenes. Using these points, blocks of pixels from each scene were correlated on the CRAY computer at NASA-Ames. Two channels from each scene were used. This procedure is fully explained by Ozga and Sigman.(3) The output was then used to create an eight channel data set. The coordinates of the pixels in this data set were the same as for the primary scene.

Underscored numbers in parenthesis refer to literature cited at the end of this report.

Multitemporal Data Set. The eight channel data set, for use in the multitemporal analysis, was generated by combining the spring and summer scenes using the coordinates of the spring scenes. Data channels 1, 2, 3, and 4 were created from the spring scene and data channels 5, 6, 7, and 8 from the summer scene. Therefore, it was not necessary to recalibrate the ground truth data to the summer scenes when doing the unitemporal analysis of the summer data. By picking channels 5, 6, 7, and 8 to be read from the eight channel data set, four channel output for a unitemporal data set were obtained representing the summer data.

SOME COST CONSIDERATIONS

Because of time and money constraints, it was not possible to completely separate the unitemporal and multitemporal processing to evaluate the cost for each analysis. However, we did observe that processing the generated eight channel data through the clustering and classification algorithms used approximately four times the computer resources that four channel data used. SRS uses a supervised clustering algorithm which clusters Landsat pixels within known crop covers. It is assumed that pixels from a given cover type come from a number of multivariate normals. The clustering algorithm is designed to find the means and covariances of the matrices representing these normals. The classification procedure used to assign a category to each pixel in the data set uses the statistics developed in clustering and a maximum likelihood algorithm to make the category to pixel assignment. Processing that reads and/or writes the eight channel data (window creation, packing, greyscales, and scattergramming) used twice the resources as the corresponding four channel data. Window creation is the extraction of Landsat data around each sample unit. Packing is the assignment of the window data (pixels) from all training units within the analysis area to the covers identified to be in the training area. Greyscales are black and white representations of a window for a single channel. Scattergramming is the process of displaying a packed file by plotting two channels; one on the horizontal axis and one on the vertical axis. Processing that did not involve raw Landsat MSS data was not impacted. Affected costs of scene-to-scene registration and creating the eight channel data set were offset somewhat by eliminating registration and calibration procedures for the secondary scene. We estimate that multitemporal analysis processing would cost about 125 percent of a single unitemporal analysis. However, the reduced professional labor in developing the classifier would offset part of this increased cost.

In states like Missouri, where both fall and spring planted crops are to be estimated, multitemporal analysis has a cost advantage since two unitemporal analyses are otherwise required.

RESULTS

For all crops, multitemporal analysis reduced the standard error of the estimate from the standard error of the unitemporal estimate. Standard errors of the unitemporal and multitemporal estimates are shown in Table 1. Unitemporal analysis achieved the greatest reduction in standard errors for rice, with a 47 percent reduction over the standard error of the JES direct expansion estimate. Additional reductions in standard errors of multitemporal over unitemporal analysis were greatest for sorghum with a 27 percent decrease. The overall reduction of standard errors for multitemporal analysis over the standard errors for the JES direct expansion were greatest for rice with a 49 percent reduction. The smaller standard errors for multitemporal verses unitemporal analysis for corn, sorghum, and soybeans, translates into a 30 to 40 thousand acre reduction. For winter wheat, this reduction was in the 10 to 12 thousand acre range. It was expected that the improvement for winter wheat might be minimal since a preferred pairing of scenes for winter wheat multitemporal analysis would be previous fall and spring scenes.

TABLE 1.
Comparison of Results of Multitemporal vs. Unitemporal Analysis - in Missouri 1984

CROP	Direct Expansion[1/] TOTAL (000)Ac.	S.E.	C.V. (%)	Unitemporal TOTAL (000)Ac.	S.E.	C.V. (%)	R.E.	Multitemporal TOTAL (000)Ac.	S.E.	C.V. (%)	R.E.
CORN	2,107	183	8.7	1,782	148	8.2	1.5	2,019	110	5.5	2.8
COTTON	122	45	37.1	115	30	26.0	2.2	204	28	13.8	2.6
RICE	140	47	48.5	105	25	23.3	3.5	63	24	38.7	6.8
SORGHUM	1,552	175	11.3	1,364	147	10.8	1.4	1,361	108	7.9	2.6
SOYBEANS	6,006	298	5.0	5,395	195	3.6	2.3	5,655	165	2.9	3.2
WW-PL	2,403	172	7.2	2,137	129	6.0	1.8	2,348	118	5.0	2.1
WW-HV	2,246	165	7.3	2,045	126	6.2	1.7	2,024	114	5.6	2.1

1/ The JES Direct Expansion (D.E.), Standard Error (S.E.), and Coefficient of Variation (C.V.), are before the DCLC Field Level Edit.

The attached charts by crop show the relationship of the estimates and their 95 percent confidence intervals. For soybeans, the unitemporal estimate was outside the 95 percent confidence limit of the direct expansion estimation. For cotton, the multitemporal estimate was outside the 95 percent confidence interval for the unitemporal estimate. With 21 comparisons between the seven crop estimates, this has a high likelihood of being due to chance.

1985 ANALYSIS PLANS

SRS plans to make further evaluations of the benefits of using multitemporal Landsat data for making crop acreage estimates. Oklahoma will be done with multitemporal scenes for winter wheat using 1984 fall scenes and 1985 spring scenes. Arkansas and Missouri will be done with the spring-summer pairs.

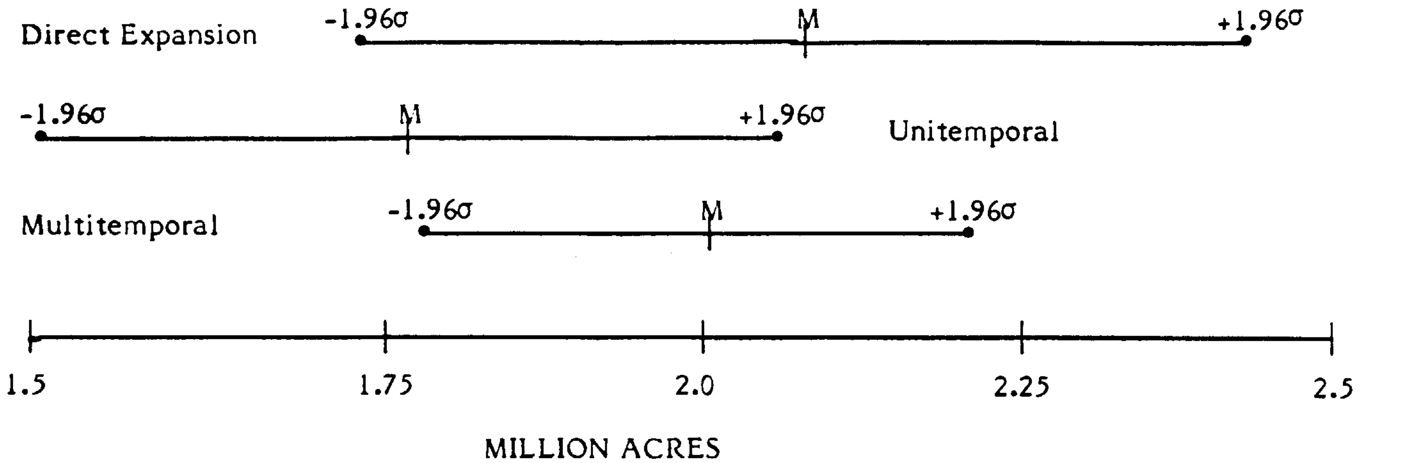
CONCLUSION

Multitemporal analysis saves both time and money over two separate unitemporal analyses when both spring and summer scenes are required for the same path-row combinations. The reduction in variance of the estimates, easier training of the classifier, and the shifting of workload to an earlier date make it attractive even where both spring and summer scenes are not required. However, if both spring and summer scenes are not required, there would be additional costs over a unitemporal analysis. Cloud cover can be more of a problem in multitemporal analysis, especially if only one satellite is operational since it may be difficult to obtain complete cloud free coverage for both dates chosen for the multitemporal analysis.(4)

Confidence Intervals at the 95 Percent Level of Estimates by Direct Expansion, Unitemporal, and Multitemporal Analysis

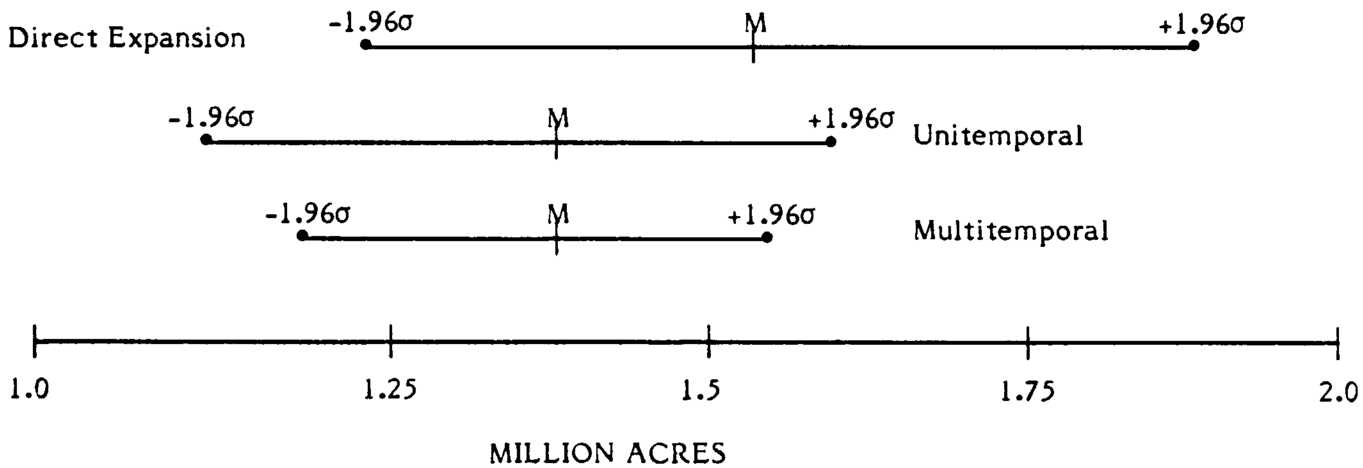
Missouri 1984

CORN



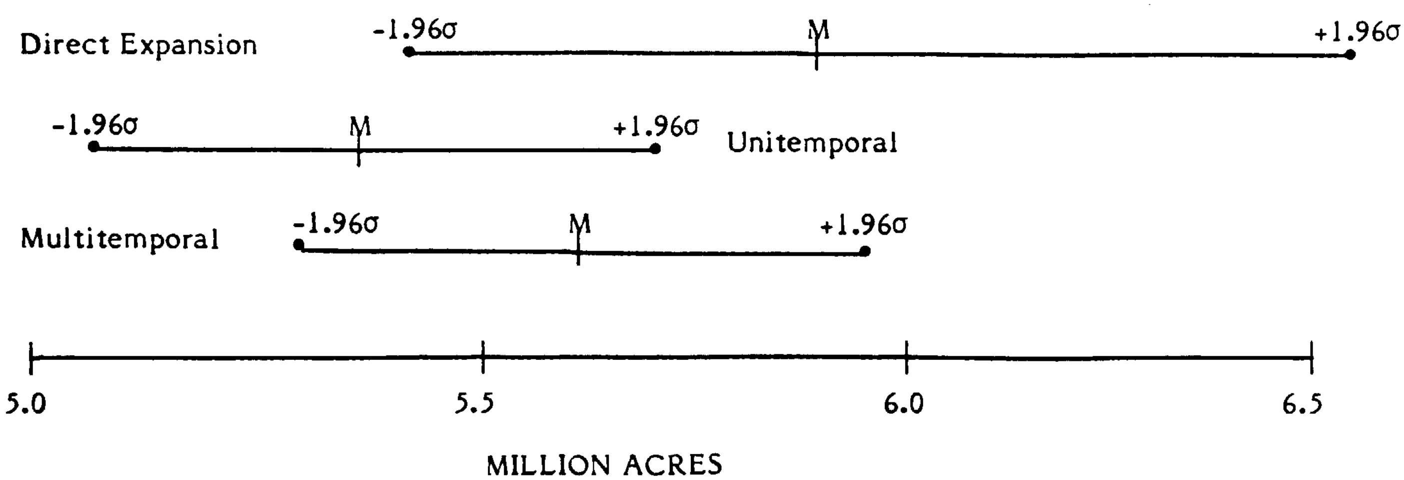
****

SORGHUM



****

SOYBEANS

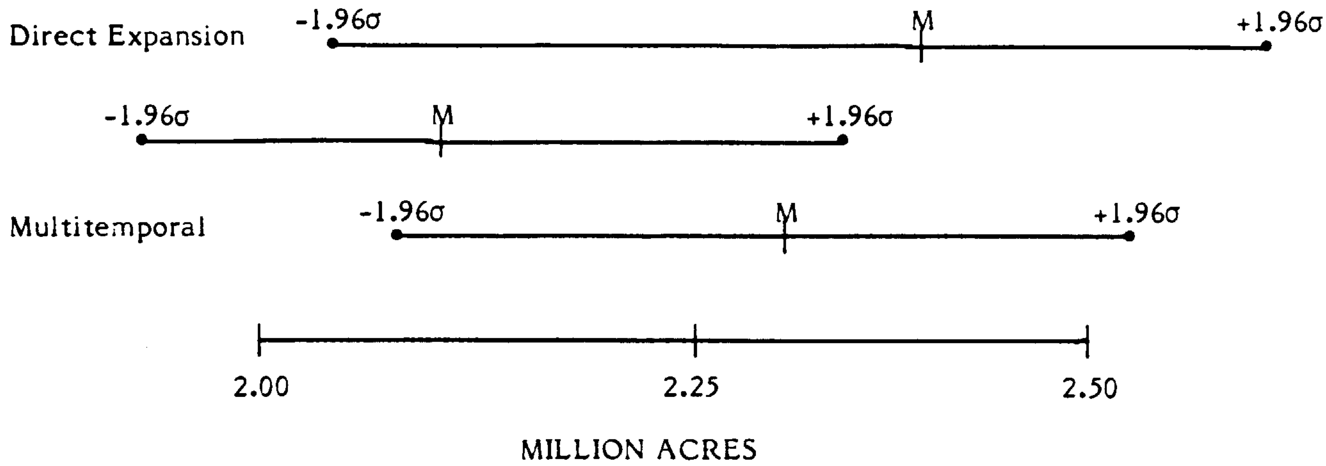


M = Mean

Confidence Intervals at the 95 Percent Level of
Estimates by Direct Expansion, Unitemporal, and
Multitemporal Analysis

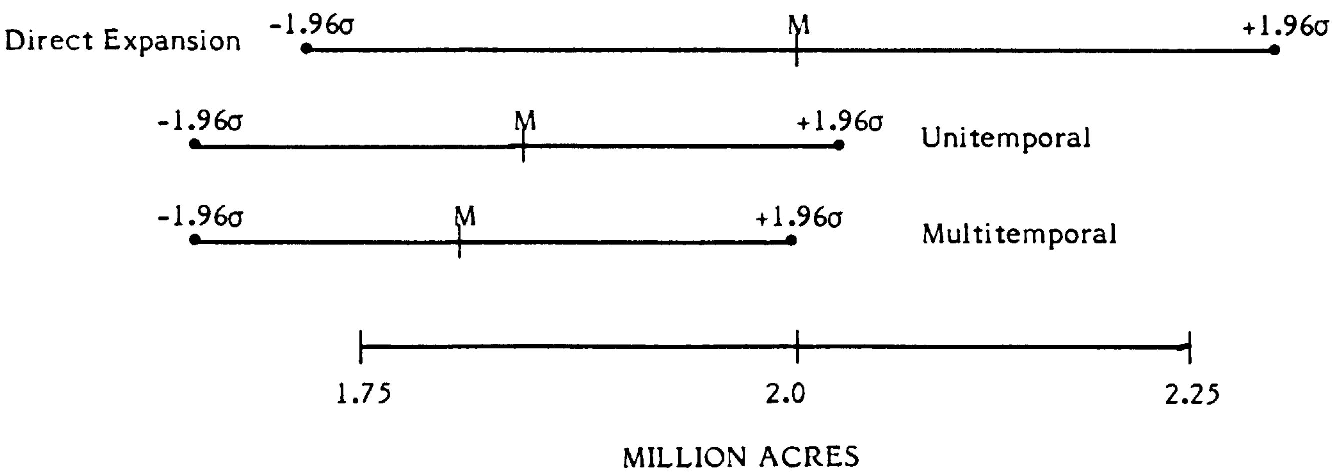
Missouri 1984

WINTER WHEAT PLANTED ACREAGE



****

WINTER WHEAT HARVESTED

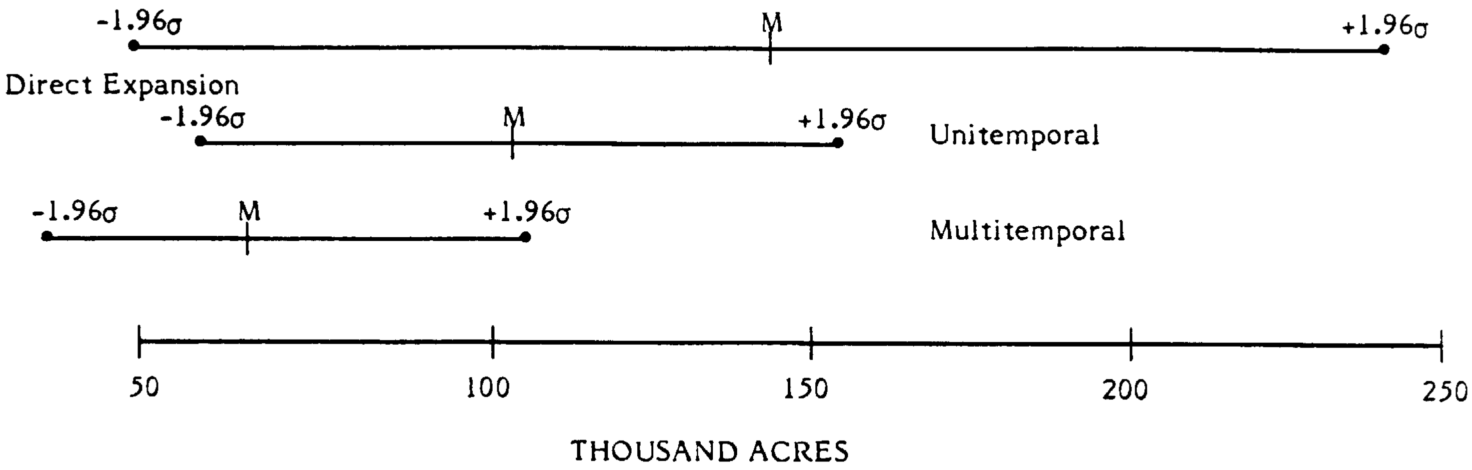


M = Mean

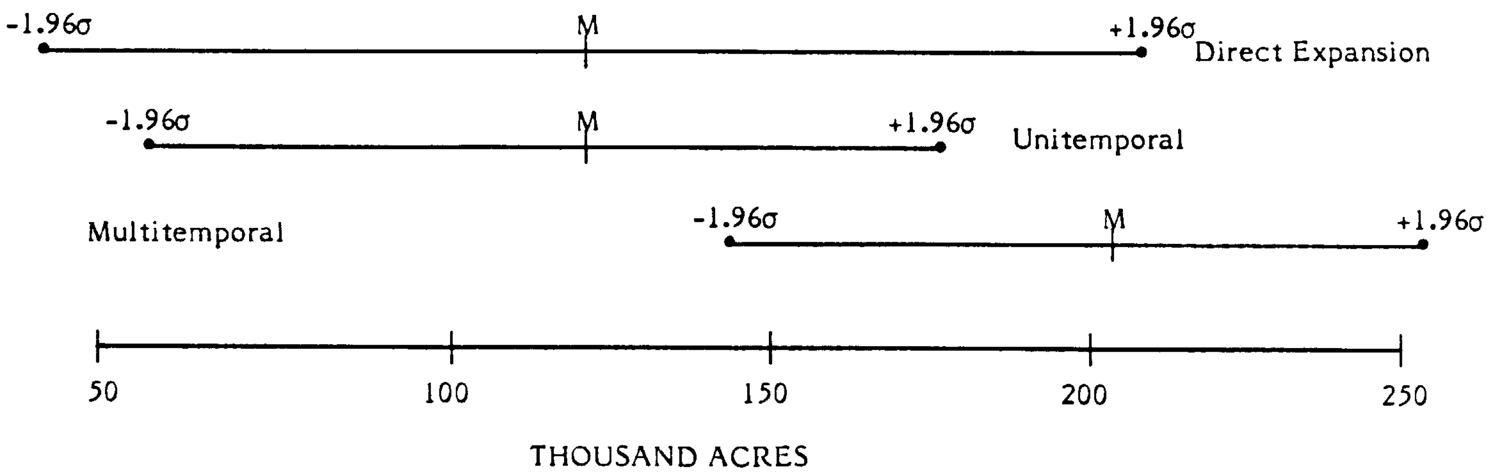
Confidence Intervals at the 95 Percent Level of Estimates by Direct Expansion, Unitemporal, and Multitemporal Analysis

Missouri 1984

RICE



COTTON



M = Mean

BIBLIOGRAPHY

(1) Hanuschak, G., R. Sigman, M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno, C. Miller; Obtaining Timely Crop Area Estimates Using Ground Gathered and Landsat Data. Economics, Statistics, and Cooperatives Service, U.S. Department of Agriculture, August 1979.

(2) Cook, Paul W., Landsat Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service, U.S. Department of Agriculture, May 1982.

(3) Ozga, M., and Sigman, R.S., Editor Multitemporal System, Statistical Reporting Service, U.S. Department of Agriculture, November 1979.

(4) Winings, S.B., Landsat Image Availability for Crop Area Estimation, Statistical Reporting Service, U.S Department of Agriculture, July 1982.

CORN
DIFF.= DCLC-JES

STATE/YEAR	DIFF. (1,000 ACRES)	DIFF. AS % OF JES	DIFF. AS % OF JES ST. ERR.
IOWA			
1978	-218	- 1.6	- 69.4
1980	162	1.1	57.2
1981	- 18	- 0.1	- 5.7
1982	- 82	- 0.6	- 28.2
1983	-104	- 1.1	- 37.8
1984	-110	- 0.8	- 36.4
1985	352	2.5	116.2
MISSOURI			
1981	-235	-10.9	-127.0
1983	-318	-17.0	-214.9
1984	- 88	- 4.2	- 48.1
1985	-244	- 8.8	-127.9
ILLINOIS			
1982	-326	- 2.7	-114.4
1983	-251	- 2.9	- 91.3
1984	-381	- 3.5	-139.6
1985	-307	- 2.7	-110.1
INDIANA			
1985	- 56	- 0.9	- 29.9
ALL STATES			
1978	-218	- 1.6	- 69.4
1980	162	1.1	57.2
1981	-253	- 1.5	- 68.9
1982	-408	- 1.6	-100.2
1983	-673	- 3.4	-161.8
1984	-579	- 2.2	-129.8
1985	-56	- 0.9	- 51.8

SOYBEANS
DIFF.= DCLC-JES

STATE/YEAR	DIFF. (1,000 ACRES)	DIFF. AS % OF JES	DIFF. AS % OF JES ST. ERR.
IOWA			
1978	94	1.2	30.4
1980	- 259	- 3.1	- 93.5
1981	37	0.5	14.3
1982	- 264	- 3.0	-100.8
1983	112	1.4	46.3
1984	- 287	- 3.3	-104.7
1985	- 54	- 0.7	- 20.5
MISSOURI			
1981	- 306	- 5.9	-107.4
1983	- 661	-11.8	-209.8
1984	- 351	- 5.8	-117.8
1985	- 547	- 9.6	-189.5
ILLINOIS			
1982	- 244	- 2.6	- 82.4
1983	- 68	- 0.7	- 25.7
1984	- 462	- 4.9	-168.0
1985	167	1.9	64.1
ARKANSAS			
1983	- 238	- 5.8	-123.3
1984	- 135	- 3.3	- 66.2
1985	- 201	- 5.4	-100.4
INDIANA			
1985	- 147	- 3.3	- 80.0
ALL STATES			
1978	94	1.2	30.4
1980	- 259	- 3.1	- 93.5
1981	- 269	- 2.0	- 70.1
1982	- 508	- 2.8	-128.6
1983	- 855	- 3.2	-166.0
1984	-1235	- 4.4	-233.0
1985	- 782	- 2.5	-144.2

WINTER WHEAT
DIFF.= DCLC-JES

STATE/YEAR	DIFF. (1,000 ACRES)	DIFF. AS % OF JES	DIFF. AS % OF JES ST. ERR.
KANSAS			
1980	- 403	-3.1	-101.0
1981	- 382	-2.8	- 98.2
1982	- 164	-1.2	- 39.7
1983	- 433	-3.8	-114.2
1984	- 91	-0.8	- 25.7
1985	- 764	-6.0	-201.7
OKLAHOMA			
1981	- 319	-4.9	-110.4
1982	- 629	-8.7	-213.2
1983	- 90	-2.1	- 43.1
1984	- 325	-6.1	-131.0
1985	71	0.9	21.9
COLORADO			
1982	- 356	-11.3	-158.9
1983	- 205	- 7.0	- 71.9
1984	- 14	- 0.4	- 6.5
1985	198	5.5	85.6
MISSOURI			
1983	89	4.3	54.6
1984	- 201	- 8.9	-121.8
1985	- 252	-16.6	-171.1
ALL STATES			
1980	- 403	- 3.1	-101.0
1981	- 701	- 3.5	-144.5
1982	-1149	- 4.7	-207.0
1983	- 639	- 3.1	-117.7
1984	- 631	- 2.9	-123.5
1985	- 747	- 2.9	-131.0

A Summary of
A Study of Bias and Variance in Landsat Data Based Regression Estimated for Crop Surveys Using Simulated Data
by James C. Lundgren

James C. Lundgren of Lockeheed Engineering and Management Services Company has completed an evaluation of the bias and variance of the crop area regression estimator. The evaluation was completed using a Landsat-data simulation algorithm which generates random segments with random pixel-level spectral values. These simulated segments were similar to 33 Missouri segments in terms of segment crop proportions; distribution of field size; distribution of segment sizes; proportion of edge pixels; and variance components between pixels (within fields), between fields (within segments), and between segments for each of four channels and for eight crops. The simulated ground covers and percentage of area were as follows:

Cover	Percent of Area
pasture	30
soybeans	25
corn	12
waste	13
woods	9
hay	7
winter wheat	3
alfafa	1

The simulated segments had an expected area of one square mile. DCLC procedures were simulated as follows:

1) A random sample of segments was selected from the simulated population
2) The sample was used to train a classifier
3) The simulated population was classified
4) Regression estimates of the area were calculated for each of the seven crops. (Alfalfa frequently had too few pixels to calculate a regression estimate.)

This procedure was then replicated 20 times for the given population. Various statistics which summarize the entire process were produced.

The number of evaluations using the classification and regression procedure were as follows:

		Population Size		
		25	50	100
	4	6	6	2
sample size	6	6	6	2
	10		6	2
	15			2

Bias

Three of the seven ground covers had significant biases in their regression estimates. The same three ground covers also had significant biases in their ratio estimates. The covers with significantly biased estimates were corn, pasture, and waste.

The study of bias concentrated on the difference between the sample and population regression equations. The two primary findings were as follows:

1) Five of the seven ground covers had significant differences between the slopes of the sample and population regression equations.
3) Three of the seven ground covers had a significant difference between the sample and the population mean number of pixels classified per segment. (These were the same covers which had significant biases in their regression and ratio estimates.)

Lundgren suggests that the slope of the regression equations for the sample tends to be larger than the slope for the population. He points out that this difference in sample and population slopes is supported by theory developed by Chhikara and Houston.

According to Lundgren the difference between the sample mean and the population mean for the number of classified pixels gives further evidence that the training sample is different from the rest of the population. Consequently, its use in the regression analysis may bias the estimates.

In a small study using a population of 25 segments, six segments were repeatedly selected at random after an independent sample of six segments had been used to train the classifier. Both the number of crops with significant biases and the magnitudes of these biases were unexpected. Significant biases were found in the regression estimates for soybeans, waste, woods and hay. Only one population was used in the analysis (thus the bias may be population specific). Two results of this study that were not surprising were as follows:

1) There was not a significant overall positive bias in the slope of the sample regression equations

2) The mean number of classified pixels for each ground cover per sample segment was not significantly different from that of the population.

Lundgren concludes that these results could be traced to the independence of the sample used to train the classifier from the sample used to calculate the regression equations.

Variance

Lundgren also examined the ratio of the mean Cochran variance to the true variance computed from the variability of the regression estimates among 20 replicates. He concluded that the Cochran formula was a suitable estimate of variance when there were enough pixels to train the classifier and where the classifier was not sensitive to small shifts in the distribution of the spectral data.

The large-sample estimate of variance was not an unbiased estimate of the variance of the regression estimator.

Relative Efficiency

Lundgren concludes that there is a trend for the relative efficiency to be larger for higher values of percent-correctly-classified and for higher values of separability; however, there does not appear to be a point above which one could safely predict high relative efficiency. Conversely Lundgren indicated if the percent-correctly-classified is less than 50 or if the separability distance is small, the use of Landsat data will not improve the efficiency of the estimate.

Prepared by NED JONES

Possible Explanation for Predominantly Negative Differences between JES and DCLC Estimates

I. Introduction

In the six-year series of DCLC estimates, the DCLC-minus-JES estimate differences are predominantly negative. The following sections discuss various phenomena that may (or may not) explain why the DCLC-minus-JES estimate differences are predominantly negative.

II. A Non-Explanation

The following has been suggested as a possible explanation for the predominantly negative differences between JES and DCLC estimates:

> There may be a systematic over-reporting in the JES. For example, non-productive areas of crop fields may be incorrectly included by either respondents or enumerators in reported crop acreages. Thus, the JES is over-estimating the true crop acreage. If the DCLC estimate is correcting for this over-reporting error, then the DCLC estimate would be estimating the true crop acreage. Hence, the DCLC-minus-JES estimate difference should be negative.

This is an appealing explanation, but unfortunately a regression estimator is mathematically incapable of correcting for this kind of data collection error. To see this, suppose systematic over-reporting is occurring. Then the reported acreages will be higher than actual acreages and on the average the direct expansion estimate will be higher than the population total. Not only does this over-reporting shift the direct expansion estimate upward, however, it also shift the regression line upward (See Figure E1). Moreover, the regression estimate will be higher than what it would be without over-reporting by the same amount as the direct expansion estimate is upwardly affected.

The only way that the regression estimate could correct for over-reporting would be if the regression line was calucated with actual acres instead of reported acres. But the actual acres are not known if there is a reporting error present. Thus, the regression estimator is not capable of correcting for a systematic reporting error in the JES. Consequently, this is not a possible explanation of why the two estimators are apparently estimating different levels.

III. A Possible-Explanation

Instead of "the JES is wrong and DCLC is right" as an explanation of the negative DCLC-minus-JES estimate differences, a more likely explanation is that there is a procedural bias in the DCLC estimates. The following discussion illustrates that this is likely by means of an example.

In this example, suppose that we have a one-Landsat-scene analysis district that is 100 miles on a side. Thus, the total amount of land in the analysis district is 10,000 square miles. Further, assume that all of the analysis district is in one land use stratum. Hence, the area-frame population size for the analysis district is also 10,000.

The JES sample size for the analysis district is 50, and one of the sampled segments, called segment S (for sampled), has 400 acres of corn in it. It so happens that one of the non-sampled segments, called segment N (for non-sampled), also has 400 acres of corn in it. Consequently, if JES enumerators were to visit segment N, then segments S and N would have identical JES data--both would have 400 acres of corn and 240 acres of non-corn.

We assume in this example that the Landsat data, on the other hand, are very different for segments S and N (See Figure E2). This is very possible because Landsat data responds not only to type of crop but also to the condition of the crop. Figure E3 displays (in two dimensions) the corn signatures for all 10,000 segments in the analysis district. We see in this figure that the corn signatures for segment N are on the outer fringe of all the corn signatures in the scene.

Figure E4a and E4b illustrate for two different situations the corn signatures in the 50 JES segments plus segment N. These two figures illustrate two undesirable situations that can occur when the JES segments are used to develop a classifier. In Figure E4a the JES segments poorly represent the spectral variability of the entire scene. In this case, when a decision is defined on the basis of the JES segments, the resulting classifier will perform well only on the JES segments. In segments that were not sampled, many corn pixels will be outside the decision boundary and called non-corn. In the case of segment N, for example, all the corn pixels will be called non-corn.

In Figure E4b, on the other hand, the JES segments do adequately represent the spectral variability of the Landsat scene. This figure illustrates, however, that an undesirable situation can occur if the procedure for developing the decision boundary puts too many wiggles in the boundary. The problem with too many wiggles in the corn decision boundary is that such wiggles tend to jog out to include corn pixels in the JES segments and then jog back in to avoid including non-corn pixels in the JES segments. This excessive wiggling again causes the developed classifier to perform better on the sampled segments, than on the non-sampled segments. For example, in Figure E4b only a portion of segment N is included inside the decision boundary. If a much simpler decision boundary had been developed that had fewer wiggles, then more of segment N would have been inside the boundary.

This excessive wiggling of decision boundaries can occur if Landsat analysts extensively tune their classifiers by trying out a large number of trial classifications resulting from different clustering parameters, cluster edting procedures and prior probabilities. Both of the situations illustrated by figures E4a and E4b produce what is called an overfitted classifier; that is, a classifier that performs much better on data that it was developed on, than data it was not developed on.

If classifier overfitting is present in DCLC procedures, then the resulting degraded performance over the non-sampled segments results directly in degraded regression-line prediction ability over non-sampled segments. For example, since segments S and N both have 400 acres of corn, they will have the same y-values (that is, segment corn acres) for regression estimation (See Figure E5). Because the corn signature for segment N were mostly or even completely outside the decision boundary whereas those for segment S were completely inside, it follows that the x-value (that is, pixels classified as corn) for segment N will be very small whereas the x-value for segment S will be large. The relationship between classification results and reported acres for segment S will be well predicted by the developed regression line but this will not be the case for segment N.

If this is happening for segments S and N, which each contain 400 acres of corn, it will also occur for other pairs of sampled and non-sampled segments having 300 acres, 200 acres, 100 acres, etc. Thus, the regresion line required to successfully predict non-sampled segments will be some distance to the left of the regression line calculated from the JES segments and overfitted classification results (See Figure E6). Moreover, simulation and theoretical results indicate that the required regression line will have a shallower slope than the slope of the regression line developed using an overfitted classifier.

The key point in Figure E6 is that the regression line which one should be using for regression estimation is the one that has the best predictive ability for non-sampled segments. The reason for this is that the purpose of the regression line is to make a prediction at the value of $\bar{X}$ (the population mean-per-segment of pixels classified as corn), and 99.5% of the segments (10,000-50 = 9,950 out of 10,000) are non-sampled segments. But, what is happening in the DCLC procedures is that if classifier overfitting is present, then we are using a regression line that is too far to the right. Thus, as is shown in Figure E6 when this incorrect regression line is evaluated at $\bar{X}$ it results in a regression estimate that is too low.

. Prepared by RICHARD SIGMAN

FIGURE E1

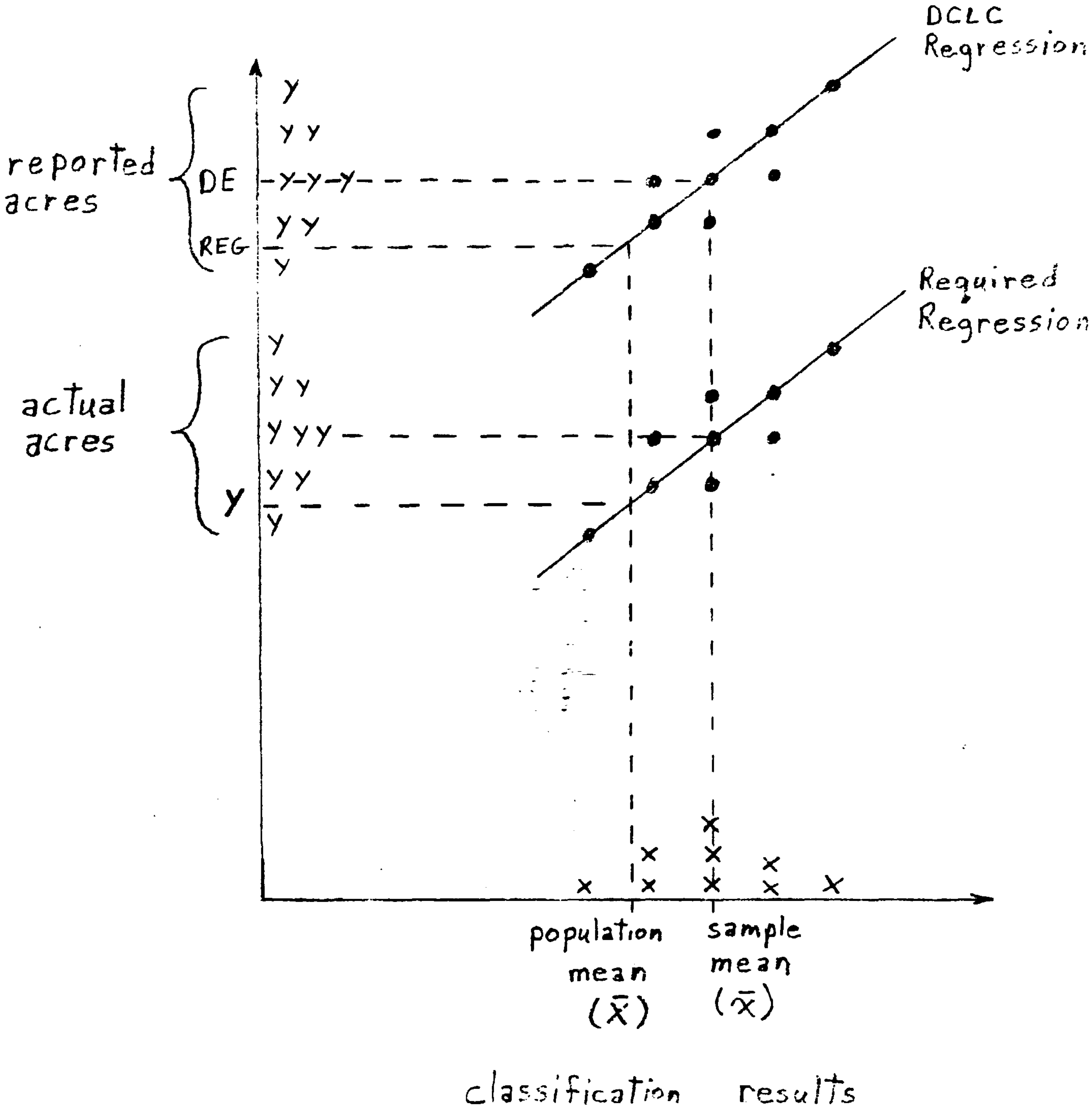
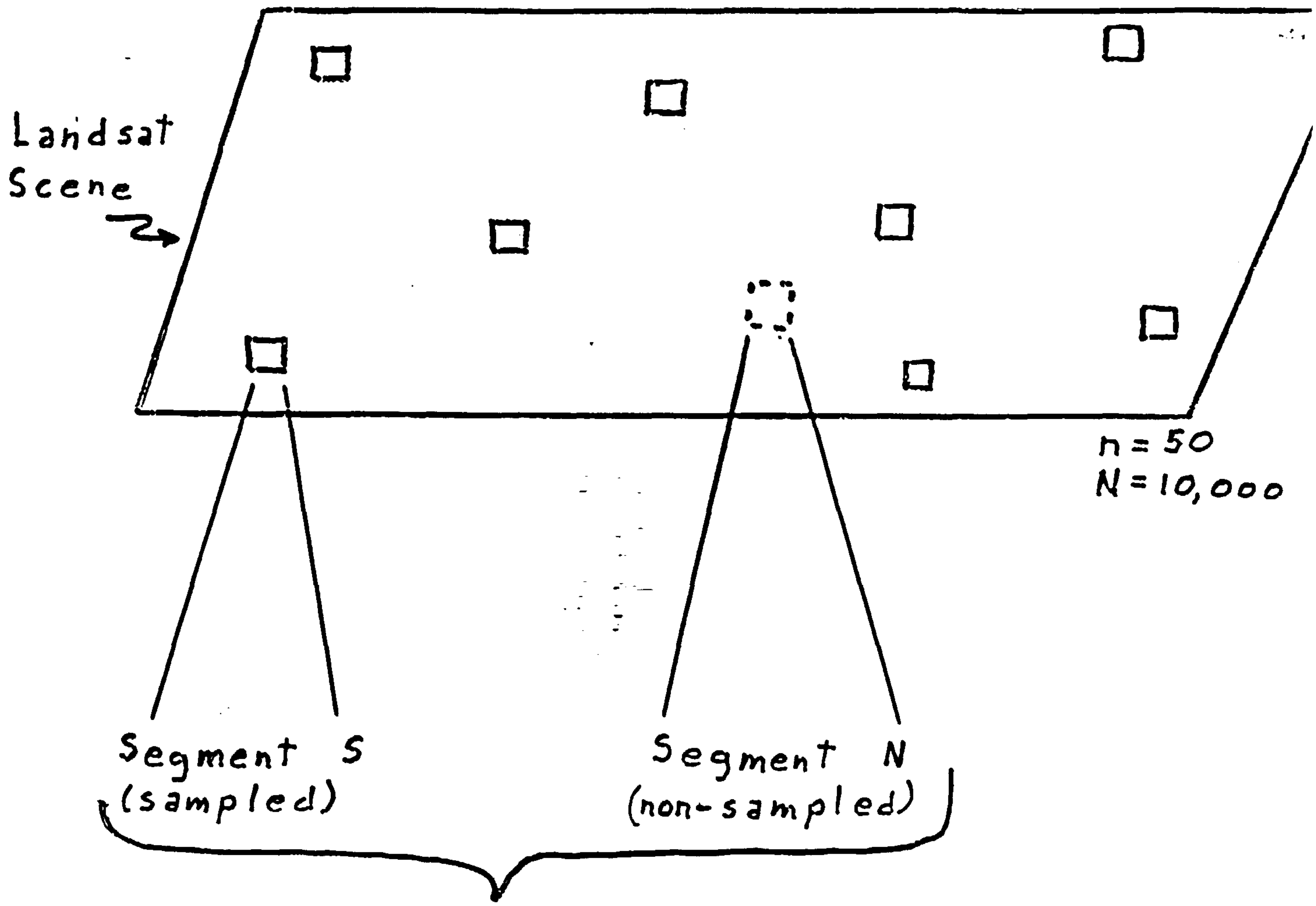

DCLC
Regression
Required
Regression
reported
acres
DE
REG
actual
acres
population
mean
($\bar{X}$)
sample
mean
($\bar{x}$)
classification results

FIGURE E2: Classifier Overfitting Example



- Identical JES data:
  - 400 acres corn, 240 acres other

- Dissimilar Landsat data
  - E.g. corn signatures (band 1):

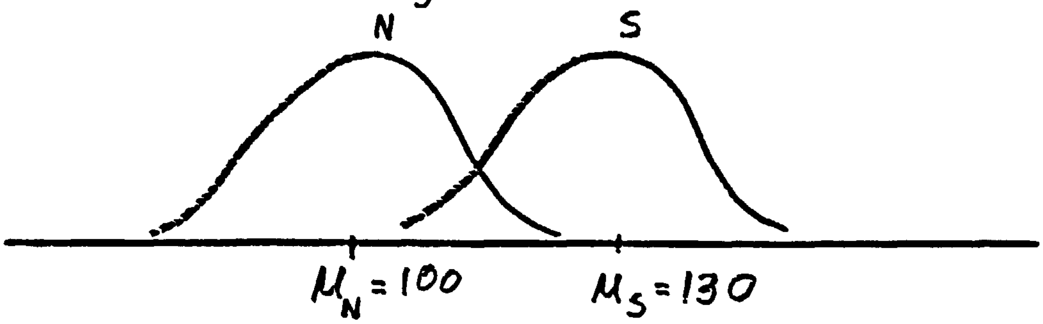


FIGURE E3: All Corn in Landsat Scene

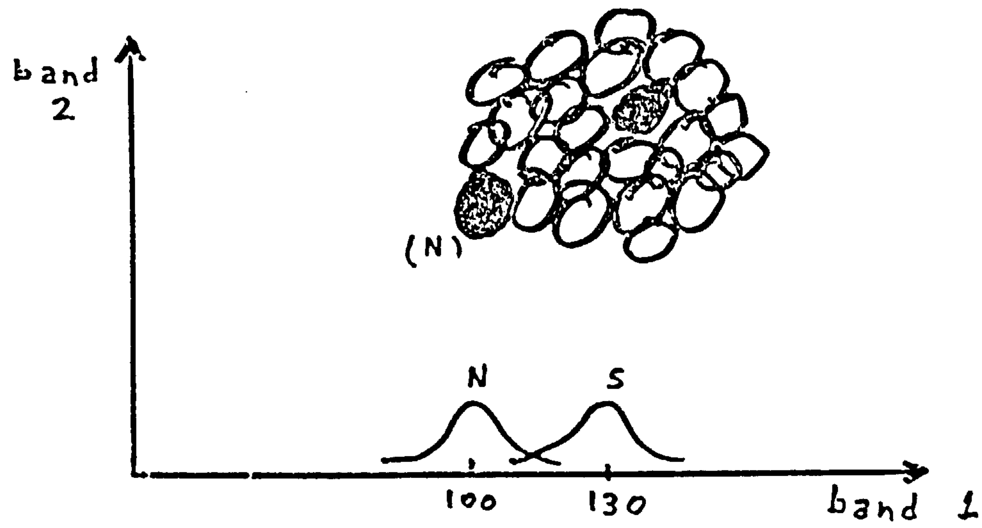
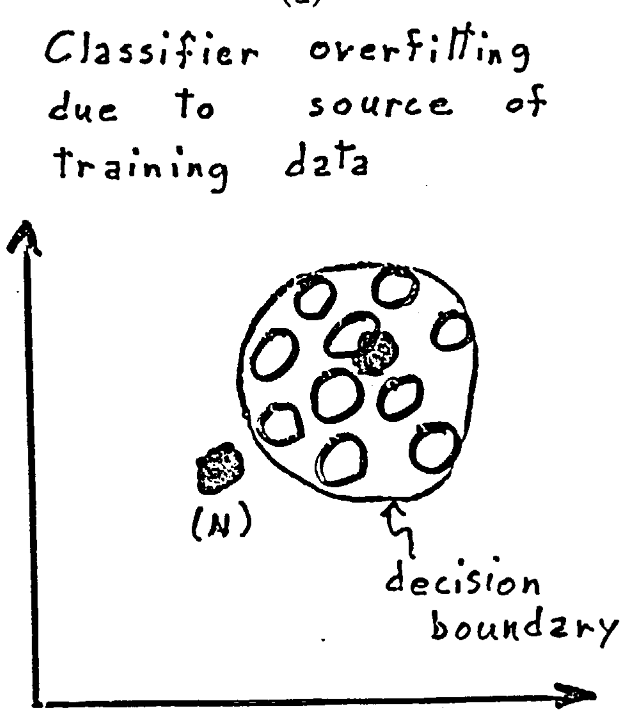


FIGURE E4: Corn in JES Segments

(a)



(b)

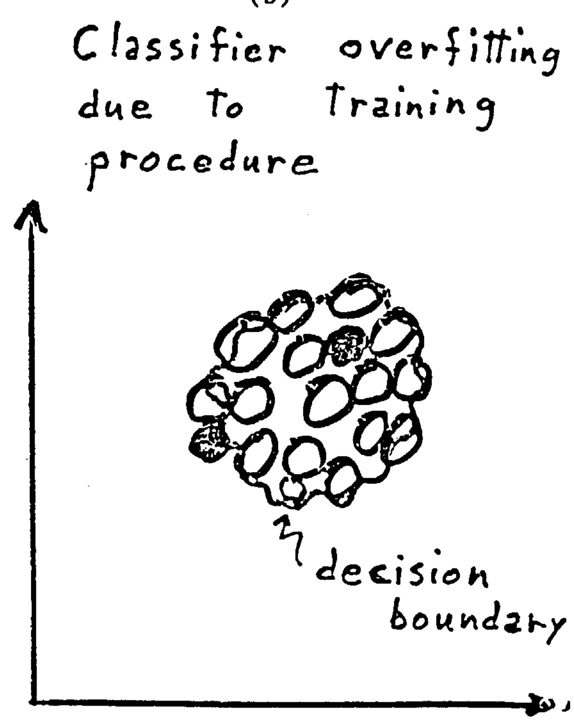


FIGURE E5

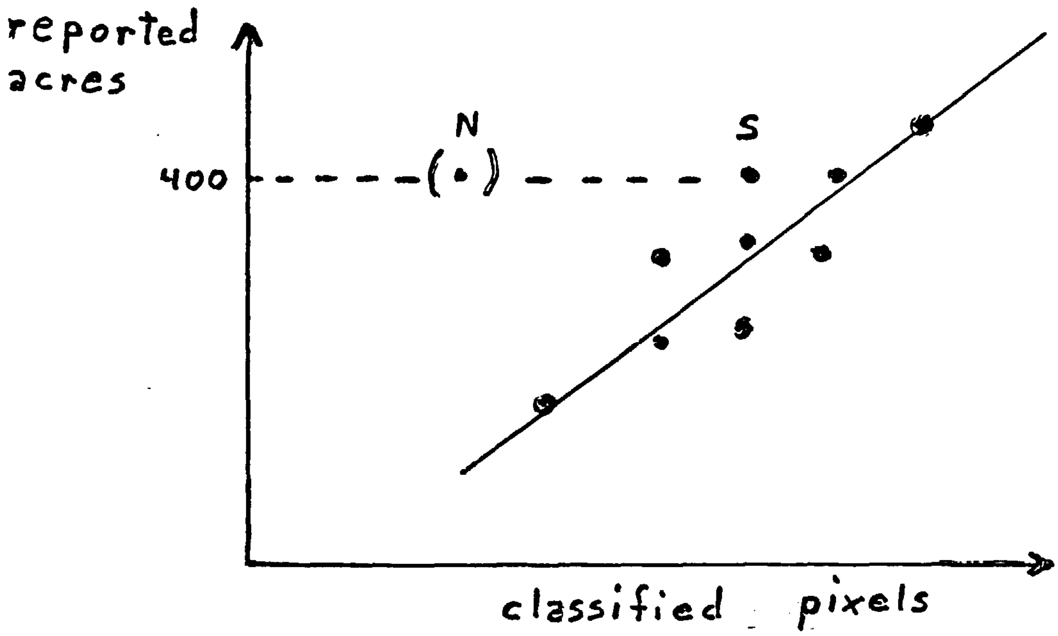
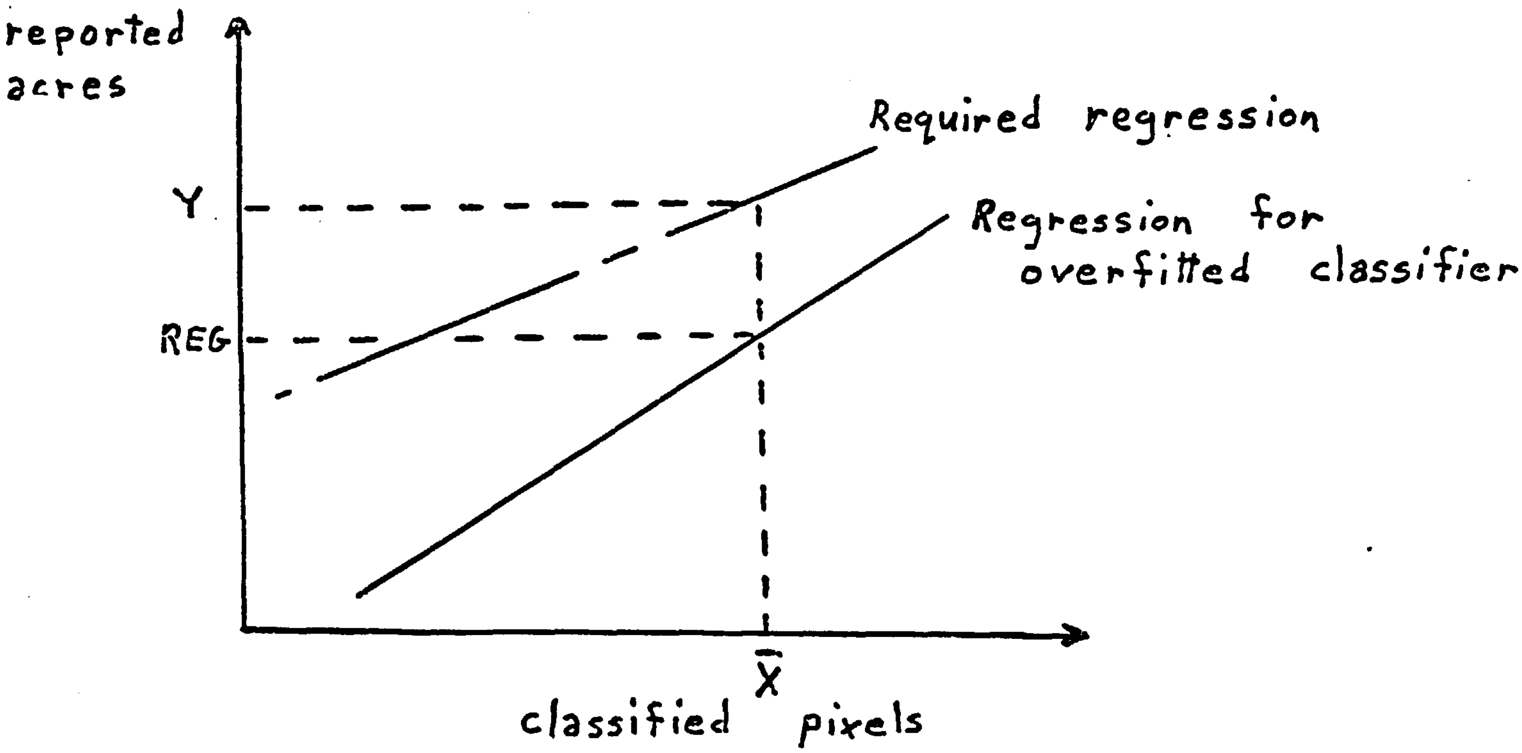


FIGURE E6



AEROSPACE REMOTE SENSING: RESEARCH RESULTS

I. INTRODUCTION

The topic of "Remote Sensing Research Results" is not a new national conference topic. At the 1977 conference, members of the then New Techniques Section spoke on this same topic. The presentation seven years ago reported on the 1975 Illinois project--SRS's first attempt to analyze Landsat data for an entire state. That research result in 1977 led in 1978 to a timely application--Landsat data for all of Iowa was analyzed to obtain end-of-year acreage estimates for corn and soybeans. In 1979 the New Techniques Section was replaced by the Remote Sensing Branch consisting of an Applications Section and a Research Section. One of the reasons for this organization was that timely, multi-state projects conducted by the Applications Section would be an important customer plus provide a large-scale test for enhanced procedures developed by the Research Section.

This talk is primarily about the activities of the Research Section since 1979, plus results from outside groups that have worked with the Remote Sensing Branch under AgRISTARS. A major driver for these activities, however, has been the large, multi-state Landsat projects conducted during this time for the most part by the Applications Section. The original AgRISTARS plan called for increasing the number of Landsat states by two each year from 1980 through 1985. This rate of growth, however, has slowed in the last two years. In 1984 crop-acreage estimates will be calculated for seven states.

II. BENEFIT-TO-COST RATIO OF REMOTE SENSING

The Remote Sensing Branch uses remotely sensed data to calculate regression estimates of crop areas. This estimation uses data from the June Enumerative Survey (JES) and Landsat satellite. The relative efficiency of the regression estimator is given by

$$\text{RE} = \text{relative efficiency}$$

$$= \text{variance (JES-est) / variance (Regression-est)} .$$

Equivalently, relative efficiency is the factor by which the sample size of the JES would have to be multiplied in order to achieve the same precision as the regression estimate. This permits the definition of the following benefit-to-cost ratio:

$$\frac{\text{benefit}}{\text{cost}} = \frac{\text{(RE)(JES cost)}}{\text{(JES cost) + (R.S. cost)}} ,$$

where (R.S. cost) = all remote sensing costs. The numerator is the cost of an enlarged JES with precision equal to the regression estimation. The denominator is the cost of the inputs to the regression estimator.

Prepared by Richard Sigman for presentation at the SRS National Conference, May 1984

III. R&D FOR ENHANCING "BACK ROOM" ACTIVITIES

The statistical theory for the regression estimator is straight forward. A very large "back room" of support activities is required, however, to process the inputs for the Landsat regression estimator.

This "back room" of activities includes a field-level edit of the JES, digitization of JES photos, scene-to-map registration of Landsat images, development of spectral signatures, computer classifications of Landsat data, plus accompanying software development and hardware maintenance. These "back room" activities have been the focus of a number of research and development studies.

A. "Winners"

A number of enhancements of "back room" activities have been "Winners" in that they have resulted in large time reductions or cost savings. In 1978 it required an average of two weeks to register a Landsat scene. In 1983 the average was four hours per scene. Though this time savings is largely attributable to a change in imagery format on the part of NASA, Branch software changes and the development of an efficient method for indexing and storing maps also contributed to this time savings.

For computer classification of Landsat data, in 1981 the cost was over $1000 per scene on the ILLIAC, whereas in 1983 the cost was between $35 and $150 (depending on the number of categories) on the CRAY XMP. This large savings will be short-lived, however, because in 1984 we will be assessed for the use of the ARPANET, which will average $300 per scene.

Another success has been our use of the Northstar microcomputer for local digitization and plotting. The use of the Northstar for local digitizing reduced our TELENET connect time from 400 hours to 200 hours per state. In 1984 the use of two Northstars for digitizing and plotting should reduce TELENET connect hours to 50 hours per state. This represents a savings of $8400 per state when comparing 1981 costs versus 1984 costs.

B. "Losers"

In addition to "Winners" we've also had "Losers", in the sense that suggested changes have not been improvements and, in some cases, have made things worse. One of these "Losers" is the use of raw Landsat data instead of our current use of resampled Landsat data. In a comparison study we found no difference--at least, for crop-acreage estimation. Another "Loser" was the use of a calibration estimator instead of a regression estimator. The difference is that calibration regresses Landsat results on the JES, whereas the regression estimator does the opposite. The calibration estimator was proposed by NASA/JSC and Lockheed. Lockheed has recently shown, however, that the calibration estimator has larger mean-square-error.

Another suggestion by an outside group has been the Canadian procedure, in which segment digitization and signature development are performed on a video display. Though this procedure may work in Canada, we found that we were unable to easily locate JES segments when evaluating the procedure on Kansas Landsat data.

Finally, another suggestion--this one by Iowa State University--has been the use of probability instead of classification as our Landsat variable. Both Iowa State and ourselves have recently shown that this does not offer any improvement for crop-acreage estimation.

Though all of these negative results may seem like research conducted for nought, they are reassuring in the sense that they indicate that our current procedures are near optimum.

C. "Jury Still Out"

In addition to the "Winners" and "Losers" we have a number of enhancements in which the "jury is still out", in the sense that there is some type of trade-off involved or evaluation is still in progress.

Two enhancements--maximum likelihood clustering (called CLASSY) and the Automatic Segment Movement Algorithm (or ASMA) have greatly increased our computer costs with, in some cases, only marginally improving estimation performance. We have not yet written these enhancements off as "Losers", but they are very expensive guests whose admission to the "procedure family" has not been decided.

The jury is still out on video digitization. In 1983 we had a first-year large-scale test in which JES segments for three Landsat states were successfully video digitized. A second-year test will be conducted this summer.

A final area where potential improvement is being evaluated is in the use of Thematic Mapper (TM) data. The TM is an improvement over the Multispectral Scanner (MSS), which we are currently using. Specifically, the TM has seven spectral bands compared to four bands for MSS. Moreover, the TM has 30 meter resolution compared to MSS's 57 meters.

The Remote Sensing Branch has conducted two studies of TM. One such study used simulated data acquired from an airplane. This study was conducted in Missouri in 1979. Relative Efficiency (RE) for corn increased from 2.0 for MSS to 6.0 for TM. Also, RE for soybeans increased from 14.3 for MSS to 20.0 for TM.

The second TM study is still in progress. It is examining real TM data acquired over Iowa on September 3, 1982. In the first phase of this study in which no spectral or spatial sampling is being performed, corn RE increased from 2.0 (MSS) to 8.3 (TM) and soybean RE from 9.1 (MSS) to 11.1 (TM).

Though TM increases relative efficiency it also increases remote sensing data and processing costs. An MSS tape costs $650 whereas a TM tape for the same area costs $3400, a more than five-fold increase. For processing costs the increase was eleven-fold in the first phase of the Iowa-TM study.

Thus TM increases both the numerator and denominator of the benefit-to-cost ratio. In the first-phase of the Iowa-TM study, the benefit-to-cost ratio increases from 0.7 for MSS to 0.8 for TM but is still less than 1.0. For soybeans, on the other hand, the benefit-to-cost ratio decreases from 3.1 for MSS to 1.1 for TM. In the second and later phases of the Iowa-TM study, subsampling either spatially or spectrally will be used. It is conjectured that this will increase the TM benefit-to-cost ratios.

IV. NEW PRODUCT STUDIES

The interest in new products is that their creation can increase the benefit-to-cost ratio. This can occur by one of two methods. In the first method additional products are generated which have some value to SRS and thus increases the numerator. In the second method, byproducts are sold outside of SRS and the resulting revenue decreases the denominator.

A. County Estimates

County estimates are an example of the first method for increasing the benefit-to-cost ratio. One way to calculate the Landsat county estimates is to calculate a regression estimate for each county. We have done this in Arizona and Idaho where the counties are large and contain many segments. This does not work, however, in the Midwest where there is an average of approximately three segments per county with some counties having no segments at all.

For situations like in the Midwest, a number of Landsat county estimators have been proposed. The Huddleston-Ray estimator uses the segment prediction equation to predict the county mean. The Cardenas, Blanchard, Craig estimator is a synthetic estimator which uses local adjustments to the mean of a large area to predict the mean of a small area. The Battese-Fuller estimator is also a prediction estimator but is based on a nested-error structure consisting of within-county and between-county variance components. It was developed by Iowa State University under a research agreement with SRS.

Two evaluation studies of these various estimators have been performed--one by NASA/JSC and the other by SRS. Both of these studies used a South Dakota data set which, because of an accompanying special soils study, had 200 area sample units distributed throughout a six-county area. The results of these two studies were that Huddleston-Ray has the smallest variance whereas Battese-Fuller has smallest bias and overall mean-square-error.

B. Land Cover Information

Land cover information is an example of the second method for increasing the benefit-to-cost ratio--that is, a processing byproduct of minor interest to SRS that is sold (through cost sharing) to an outside agency. In 1981 a land cover study was conducted in Kansas followed in 1983 by a land cover study in Missouri. In the Missouri study, 67 rotated-out, non-agricultural segments were used. These were flown by NASA/NSTL and enumeration was by photo-interpretation. A report on the Missouri study is currently being written. Also in 1983 ground data was collected in New Jersey for use with TM in producing land cover mapping products. The New Jersey data analysis is just now getting started. In 1984 a land cover study will be conducted in Arkansas. The Soil Conservation Service and the Forest Service are each paying $35,000 as customers for resulting Landsat classification tapes.

Results for the 1981 Kansas Land Cover Study were encouraging. Covers with regression estimate C.V.'s less than 10% were cropland, rangeland, farmstead, forest (not grazed), and residential. Very rare items such as stripmines and sand dunes, had very large C.V.'s. The focus of the Missouri Land Cover Study was forest categories. Only hardwoods, however, had a C.V. of less than 10%. The relative efficiencies in the Missouri study were not very high. Grazed forest and mixed conifers-and-hardwoods had a relative efficiency of 1.0, indicating no estimation improvement from the use of Landsat data. The other covers which had low C.V.'s were agricultural categories. Covers with high relative efficiencies were hardwoods, commercial, rivers, and row crops.

Additional costs result from producing land-cover information as a byproduct of crop-acreage estimation. The increased enumeration effort increases JES costs approximately 11%. The increase in time to perform the field-level edit results in a 42% increase, whereas the increase in manual digitization time is less than 7%. Because winter wheat, corn, and soybeans were being estimated in Missouri in 1983, there was no increase in Landsat data costs. BBN costs for winter wheat were approximately $10K, with an additional $12K for corn and soybeans, and an additional $11.5K for land cover. If the land cover work had not been performed, then the corn and soybeans increment would have been smaller because of fewer categories in the multitemporal classifications. Thus, the resulting BBN cost increase from doing wheat, corn, and soybeans to doing crops plus land cover is approximately 100%.

C. Cooperative Projects

Cooperative projects in California and Idaho are underway because the departments of water resources in these states also want to use Landsat data for their inventory needs.

In California, SRS is funding accompanying research studies by the University of California at Berkeley and by NASA/Ames. A 1982 data set for the Central Valley was created. SRS has completed its study of the 1982 data set and a report has been recently published. The University of California at Berkeley is supposed to complete their analysis of the 1982 data set this summer and make recommendations toward a large-scale test in 1985.

In Idaho, SRS has provided funds to the Idaho Department of Water Resources, who then sub-contracted NASA/Ames. NASA/Ames has recently completed 1983 Landsat estimates for potatoes in a four-county area in Idaho. At the request of the Idaho SSO, the Remote Sensing Branch has reviewed these estimates and some questions about the estimates have been raised.

V. PRODUCT CHARACTERISTICS STUDIES

A. Variance-Underestimation Studies

One of our concerns is if we are estimating the variance of the regression estimator correctly. The reason for this concern is that we use the JES ground data twice. Once for developing the Landsat classifier, and a second time in calculating the regression estimate. This is a departure from the standard textbook procedures and would suggest that the large-sample variance formula underestimates the variance of the Landsat regression estimator.

This question has been studied by jackknifing studies. In these studies, classifier development and estimation are performed on different portions of the same data set. Jackknifing studies have been performed by SRS in Illinois in 1975 and in California in 1982. Also studies have been performed by NASA/JSC and Iowa State University using a 1979 Missouri data set.

The conclusions from all these studies is that in the Midwest we are underestimating the variance of the regression estimator by less than 10% for major crops and from 20% to 30% for minor crops. In California, on the other hand, this problem is potentially very serious, suggesting that the JES segments may not be adequately representing the spectral variability of the population.

B. Simulation Studies

Lockheed researchers are currently performing for SRS a simulation study for the Landsat regression estimator. Such simulation permits the characterization of small-sample properties not determinable from sampling theory or from a single JES sample. For example, if these simulations show large biases in very small samples, then we would not calculate Landsat regression estimates in areas where we have very few segments. Two simulations are being performed: a simplified simulation and a realistic simulation. The simplified simulation study was partially funded by NASA and is almost finished. It assumes equal size segments and two crops—that is, the target crop and everything else. The simplified simulation simulates segment crop proportions and the variabity of classification performance from segment to segment.

The realistic simulation is just getting started. The price for realism, however, is limited scope: it can only simulate Missouri. The ground data module simulates crop proportions, field sizes, segment sizes, and percentage of edge pixels for eight crops. The Landsat module simulates the segment, field, and pixel effects on Landsat reflectance values and also simulates mixed pixels, or pixels falling on the edge of a field.

An item of major interest is if the relationship between the ground data and Landsat classification results is really linear over the entire population. The reason for this interest is if the population relationship is linear then the Landsat regression is unbiased. The results from the simplified simulation are similar in appearance to observed data and appear to be very linear.

The preliminary results from the simplified simulation are encouraging. One of the simulation runs consisted of 500 samples of size 10 from population with a crop proportion of 0.25. The relative bias was only 1%, which was 0.12 of the standard error. The underestimation of the variance was 18% for the large-sample variance formula but only 7% for the "small sample" variance formula, which is valid only when the population relationship is linear.

VI. CONCLUSIONS

The research activities described above have resulted in enhancements to many of our "back room" activities, have studied possible new output products, and have provided increased understanding of the characteristics of Landsat-based crop-area estimates. Some of these activities are now in progress. The results from the current work plus follow-on studies will be the subject of future reports.

REMOTE SENSING BRANCH

PUBLICATIONS

Files Maintained in Room 4833 under
6000 SRS PROGRAMS RECORDS/6300 PUBLICATIONS

6000 SRS PROGRAM RECORDS
6300 Publications

1975

75-1 Houseman, Earl E. "Area Frame Sampling in Agriculture," SRS No. 20, Nov. 1975.

75-2 Wigton, William H. "The Technology of Landsat Imagery and Its Value in Crop Estimation for the U.S. Department of Agriculture." December 1975.

1976

76-1 Von Steen, Donald H. and William H. Wigton. "Crop Identification and Acreage Measurement Utilizing Landsat Imagery," Final Report for ERTS 1 Investigation, Goddard Space Flight Center, Greenbelt, MD. March 1976.

76-2 Caudill, Charles E. "Current Methods and Policies of the Statistical Reporting Service," 1976 LARS Symposium, Purdue University. June 1976.

76-3 Wigton, William H. "Use of Landsat Technology by Statistical Reporting Service," 1976 LARS Smposium, Purdue University. June 1976.

76-4 Wigton, William H. "La Utilization de la Technologia Landsat por el SRS," 1976 LARS Symposium, Purdue University. Junio 1976. (SP)

76-5 Hanuschak, George. "Landsat Estimation with Cloud Cover," 1976 LARS Symposium, Purdue University. June 1976.

76-6 Caudill, Charles E. "Normas y Metodos Corrientes del Servicio de Estadisticas Agropecuarias," 1976, LARS Symposium, Purdue University. June 1976. (SP)

76-7 Cook, Paul. "Multi-Date Analysis Results Using ERL's Registration-Illinois 1976," 1976.

76-8 Cook, Paul. "The 1976 Illinois Project Follow-On Study." 1976.

76-9 Huddleston, Harold F., "A Training Course in Sampling Concepts for Agricultural Surveys," USDA, SRS No. 21, 1976.

1977

77-1 Gleason, Chapman, Robert R. Starbuck, Richard S. Sigman, and George A. Hanuschak. "Stratified Acreage Estimates in The Illinois Crop Acreage Experiment," Fourth Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana. June 1977.

77-2 Hanuschak, George. "The Effect of the Landsat Cloud Cover Domain on Winter Wheat Acreage Estimation in Kansas During 1976." June 1977.

77-3 Ozga, Martin, Walter E. Donovan, and Chapman Gleason. "An Interactive System for Agricultural Acreage Estimates Using Landsat Data." June 1977.

77-4 Cook, Paul. "Texas 1974 Remote Sensing Project." August 1977.

77-5 Hanuschak, George, and others. "Crop Reporting by Satellite." August 1977.

77-6 Gleason, Chapman and others. "The Auxiliary Use of Landsat Data in Estimating Crop Acreages: Results of the 1975 Illinois Crop-Acreage Experiment." October 1977.

77-7 Hanuschak, George and Kathleen Morrissey. "Pilot Study of the Potential Contributions of Landsat Data in the Construction of Area Sampling frames." October 1977.

77-8 Caudill, Charles E. "Estimation of Agricultural Production by Sample Surveys, the U.S. Experience," Invited paper at International Statistical Institute, New Dehli, India. 1977.

77-9 Wigton, William H. "World Food Survey." 1977.

1978

78-1 Wigton, William H. and Peter Bormann. "A Guide to Area Sampling Frame Construction Utilizing Satellite Imagery." March 1978.

78-2 Wigton, William H. and Peter Bormann. "Guia Para la Construccion de un Marco de Muestreo por Areas Utilizando Imagenes por Satelite." Mareo 1978. (SP)

78-3 Wigton, William H. and Harold F. Huddleston. "A Land Use Information System Based on Statistical Inference." April 1978.

78-4 Hanuschak, George and Manuel Cardenas. "Multiple Regression Estimation Using Classified Landsat Data." April 1978.

78-5 Wigton, William H. and Harold F. Huddleston. "Un Sistema Para Recoger Informacion Sobreel Uso de la Tierra, Basado en la Inferencia Estadistics." April 1978. (SP)

78-6 Craig, Michael E. and Manuel Cardenas. "Kansas Wheat Nonsampling Error Analysis." April 1978.

78-7 Hart, Galen F., Richard Sigman, and George Hanuschak. "Improved Crop Acreage Estimation Utilizing Landsat," Meeting of American Economics Association in Washington, D.C. May 1978.

78-8 Cardenas, Manuel and George Hanuschak. "An Exploratory Study on the Effects of Field Size and Field Boundary Pixels on Crop Spectral Signatures." June 1978.

78-9 Sigman, Richard S. and others. "The Use of Regression Estimation with Landsat and Probability Ground Sample Data," American Statistical Association. July 1978.

78-10 Cardenas, Manuel, Mark Blanchard, and Michael Craig. "On the Development of Small Area Estimators Using Landsat Data as Auxiliary Information." August 1978.

78-11 Hart, Galen F. and others. "Integrated Use of Landsat with Ground Data," XXIXth Congress of the International Astronautical Federation. October 1978.

78-12 Craig, Michael, Richard Sigman, and Manuel Cardenas. "Area Estimates by Landsat: Kansas 1976 Winter Wheat." August 1978.

1979

79-1 Huddleston, Harold F. and Roy Russell. "Evaluacion Agricola Y de Recursos en Jamaica Utilizando un Marco de Muestreo por Areas." April 1979. (SP)

79-2 Craig, Michael, Richard Sigman, and Manuel Cardenas. "Area Estimates by Landsat: Kansas 1976 Winter Wheat," Thirteenth International Symposium of Remote Sensing of Environment Proceedings (Poster Session). April 1979.

79-3 Hanuschak, George. "Precision of Crop-Area Estimates," Thirteenth International Symposium of Remote Sensing of Environment Proceedings. April 1979.

79-4 Hanuschak, G. and others. "Crop-Area Estimates from Landsat: Transition from Research and Development to Timely Results," Proceedings, 1979 Machine Processing of Remotely Sensed Data Symposium, LARS, Purdue University. June 1979.

79-5 Hanuschak, George and others. "Obtaining Timely Crop-Area Estimates Using Ground-Gathered and Landsat Data," Technical Bulletin No. 1609. August 1979.

79-6 Hanuschak, George and others. "La Obtencion Oportuna de Estimaciones de Areas Cultivados Mediante el Uso de Datos Recolectadoe en Tierra e Informacion Recibida por Landsat." August 1979. (SP)

79-7 Caudill, Charles E. and Richard McArdle. "Research Evaluation Considerations for AgRISTARS," Third Conference on Economics of Remote Sensing, Lake Tahoe. November 1979.

79-8 Cardenas, Manuel, Michael Craig, and Mark Blanchard. "Small Area Estimators: County Crop Acreage Estimates Using Landsat Data," Public Data Use Vol. 7, #5/6,pp. 23-28, Elsevier North Holland. December 1979.

79-9 Hart, Galen F. and William H Wigton. "Applications of Landsat in Area Frame Design." 1979.

79-10 Houseman, Earl E. "L'Echantillonnage D'Aires dan l'Agriculture," SRS No. 20-F. November 1979. (FR)

1980

80-1 Craig, Michael. "Using the EDITOR System for Crop Area Estimation." February 1980.

80-2 Sigman, Richard. "Crop Area Estimates from Landsat and Ground Survey Data," Canadian Federal-Provincial Committee on Agricultural Statistics, Ottawa, Ontario. March 1980.

80-3 Russell, Roy and Harold Huddleston. "Surveys of Crop Production in Jamaica," UN Seminar Proceedings of the 14th International Symposium on Remote Sensing of Environment, San Jose, Costa Rica. April 1980.

80-4 Hanuschak, George and others. "Crop Area Estimates from Landsat: Transition from Research and Development to Timely Results," IEEE Transaction on Geo-Science and Remote Sensing, Vol. GE-18, No. 2. April 1980.

80-5 Huddleston, Harold F. "A Training Course in Sampling Concepts for Agricultural Surveys." November 1980.

80-6 Huddleston, Harold F. "Cours D"Instruction En Concepts D'Echantillonnage Pour Etudes Agricoles," SRS # 21-F. November 1980. (FR)

80-7 Sigman, Richard and Gregory Larson. "The Auxiliary Use of Landsat Data in Estimating Crop Yields: Results of the 1978 Iowa Feasibility Study." Dec. 1980.

80-8 Craig, Michael and Charles E. Miller. "Area Estimates By Landsat: Arizona 1979." January 1980.

80-9 Russel, Roy and Harold Huddleston. "Encuestas Solere Produccion Agricola En Jamaica." April 1980. (SP)

80-10 Ozga, Martin; Seth Steinberg, Luebbe, Raymond. "A Systems Guide to EDITOR." October 1980.

80-11 Sigman, Richard. "Estimation Des Surfaces Cultivees Grace Aux Observations Landsat Et Terrestres." 1980.

1981

81-1 Mergerson, James. "Crop Area Estimates Using Ground-Gathered and Landsat Data, A Multitemporal Approach: Missouri 1979," ESS Staff Report No. AGESS810223. February 1981.

81-2 Kleweno, D. and C. Miller. "1980 AgRISTARS DCLC Project Summary-Crop Area Estimates for Kansas and Iowa," ESS Staff Report No. AGESS810414. March 1981.

81-3 Willett, Joseph. "Area Sampling Frames for Agriculture in Developing Countries," USDA/ESS report to USAID. May 1981.

81-4 Ozga, M. and Richard Sigman. "An Autodigitizing Procedure for Ground Data Labelling of Landsat Pixels," Fifteenth International Symposium on Remote Sensing of Environment. May 1981.

81-5 Sigman, Richard and Michael Craig. "Potential Utility of Thematic Mapper Data in Estimating Crop Areas," Fifteenth International Symposium of Remote Sensing of Environment. May 1981

81-6 Winings, Sherman B. "Estimating Potatoes and Other Crops in the Red River Valley of North Dakota and Minnesota Using 1980 Landsat Imagery." May 1981.

81-7 May, George. "Evolution of Land Cover Definitions and Survey for the Economics and Statistics Service." June 1981.

81-8 Hart, Galen F. and William H. Wigton. "Application of Landsat in Area Frame Design," Interactive Symposium and Workshop on Applications of Remote Sensing for Rice Production, Hyderabad, India. September 9-13, 1981.

81-9 Hart, Galen F., Paul W. Cook and William H. Wigton. "Application of Remote Sensing for Rice: U.S. Experience," Interactive Symposium and Workshop on Applications of Remote Sensing for Rice Production, Hyderabad, India. September 9-13, 1981.

81-10 Ryerson, Robert A., Richard S. Sigman, and Ronald J. Brown. "Satellite Remote Sensing for Domestic Crop Reporting in the United States and Canada: A look to the Future," Presented as a Plenary Session Paper, 7th Canadian Symposium on Remote Sensing, Winnipeg, Manitoba. September 9, 1981.

81-11 May, George and Rich Allen. "Non-Sampling Errors in Non-Agricultural Strata of the 1980 Kansas June Enumerative Survey." October 1981.

81-12 Mergerson, James. "Crop Area Estimation Using Ground-Gathered and Sampled Landsat Data," ESS Staff Report No. AGESS810408. May 1981.

81-13 "Crop - Area Estimates Using Landsat and Point Sampling Survey Data." (1). 1981.

1982

82-1 May, George and Michael Craig. "Man Induced and Natural Resource Analysis of an Arid Region in California, USA," ERIM Symposium. January 1982.

82-2 AgRISTARS Program Support Staff, "FY-81 AgRISTARS", Annual Report. January 1982.

82-3 Cook, Paul W. "Landsat Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service 1972-1982," Presented at the 1982 ASA Meetings in Cincinnati, Ohio, August, 1982. May 1982.

82-4 Mergerson, James W.; Hanuschak, George A.; Cook, Paul W. "Applications of Satellite Remote Sensing for U.S. Crop Acreage Estimation, 1980-81 Results." May 1982.

82-5 Hanuschak, George. "Agriculture and Resource Inventory Survey Through Aerospace Remote Sensing (AgRISTARS) Domestic Crops and Land Cover (DCLC), Satellite Resolution Comparison Via Airplane Simulation." June 1982.

82-6 Hanuschak, George A.; Allen, Richard D.; Wigton, William H. "Integration of Landsat Data Into The Crop Estimation Program of USDA's Statistical Reporting Service, 1972-1982." Invited paper at the 1982 Machine Processing of Remotely Sensed Data Symposium, Purdue University, West Lafayette, Indiana. July 1982.

82-7 Winings, Sherman B. "Landsat Image Availability of Crop Area Estimation," Presented at 1982 Machine Processing of Remotely Sensed Data Symposium Laboratory for Applications of Remote Sensing. Lafayette, Indiana. July 1982.

82-8 Mergerson, James W.; Miller, Charles E.; Ozga, Martin; Holko, Martin; Winings, Sherman; Cook, Paul; Hanuschak, George A. "1981 AgRISTARS DCLC Four State Project," Presented as an invited paper, at the Eighth International Symposium on Machine Processing of Remotely Sensed Data. July 1982.

82-9 Walker, Gail; Sigman, Richard. "The Use of Landsat for County Estimates of Crop Areas: Evaluation of the Huddleston-Ray and Battese-Fuller Estimators," SRS Staff Report No. AGES 820909. September 1982.

82-10 Walker, Gail; Sigman, Richard. "County-Level Crop Area Estimates from Landsat: Evaluation of the Battese-Fuller Estimator," Proceedings of the Survey Section, American Statistical Association, 1982 Annual Meeting.

82-11 May, George and Dave Kleweno. "1981 Kansas Land Cover Survey Manual."

82-12 Fecso, Ron; Wayne Gardner; Bob Hale; Van Johnson, and Steve Pavlasek. "Construction of a Remotely Sensed Area Sampling Frame for Southern Brazil." June 1982.

82-13 Allen, Rich; "Proceedings of the NASA Workshop on Registration and Rectification." June 1982.

82-14 Allen, Rich; "The Utility of Remote Sensing in Agricultural Statistics." July 1982.

1983

83-1 USDA, NASA, Commerce, Interior. "FY 1982 AgRISTARS Research Report," AP-J2-0393. January 1983.

83-2 Johnson, Van; Kestle, Rick. "1982 AgRISTARS DCLC Project, Winter Wheat Estimates, GLOSSARY OF TERMS." May 1983.

83-3 Winings, Sherman; Cook, Paul; Hanuschak, George. "AgRISTARS DCLC Applications Project - 1982 Corn and Soybeans Area Estimates for Iowa and Illinois," Presented at the Seventeenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, May 9-13, 1983. May 1983.

83-4 Mergerson, J.W.; Johnson, V.B., Kestle, R.A. (REPRINT FROM PROCEEDINGS) "Machine Processing of Remotely Sensed Data with emphasis on Natural Resources Evaluation," 1982 Winter Wheat Area Estimates for Colorado, Kansas, Oklahoma. Purdue University, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana 47907 USA. June 1983.

83-5 Federal Survey Plan

83-6 Sigman, R; G. Walker. Comparison of Auxiliary of Variables for Satellite-Data Regression Estimators in Crop Surveys. September 1983 (Presented at ASA Mtg. in Toronto, Canada in August '83).

83-7 Sigman, R; G. Walker. The Use of Landsat for County Estimates of Crop Areas: Evaluation of the Huddleston-Ray and the Battese-Fuller Estimators for the Case of Stratified Sampling.

83-8 Hanuschak, G; C. Caudill. "Management Issues of Integrating Earth Resources Satellite Data Into The United States Department of Agriculture's Domestic Crop-Area Estimation Program," (Presented at ORSA/TIMS in Orlando, Florida). November 1983.

83-9 May, G.A.; Holko, M.L., Anderson, J.E. "Classification and Area Estimation of Land Covers in Kansas Using Ground-Gathered and Landsat Digital Data." October 1983.

83-10 Zuttermeister, Paul., Sigman, R. "The Effect of Cubic Convolution Resampling on Crop Area Estimates." December 1983.

1984

84-1 Ozga, Martin. "Experience With The Use of Supercomputers to Process Landsat Data." March 1984.

84-2 Holko, Martin L. "1982 Results for the California Cooperative Remote Sensing Project." March 1984.

84-3 Sigman, Richard. "Aerospace Remote Sensing: Research Results." May 1984.

84-4 Holko, Martin L. "The Results of Integration of the North Star Microcomputers into the Remote Sensing Branch's Data Processing System." June 1984.

84-5 Holko, Martin L., R. Sigman. "The Role of Landsat Data in Improving U.S. Crop Statistics." Cook, Paul W., George May, Richard Kestle. "A Continuing Development of an Application for Landsat Data: 1983 DCLC Winter Wheat Acreage Estimates for Four States." These two papers were presented at the Eighteenth International Symposium Remote Sensing of Environment, Paris, France. October 1984.

1985

85-1 Winings, Sherman . "Multitemporal vs. Unitemporal Analysis of MSS Landsat Data on a Full State Basis." May 1985.

85-2 Zuttermeister, Paul. "Evaluating TM Data for SRS Crop Acreage and Production Estimates." July 1985.

85-3 Ozga, Martin. "USDA/SRS Software for Landsat MSS-Based Crop-Acreage Estimation." September 1985.

85-4 May, George, Ned Jones and Martin Holko. "Landsat Large-Area Estimates for Land Cover." September 1985.

85-5

85 6

85-7

85-8

AGRISTARS REPORTS

* = NASA

6. DC/LC

Reports

* 6-01. Evaluation of Large Area Crop Estimation Techniques Using Landsat and Ground-Derived Data. DC-L1-04051, JSC-17116, LEMSCO-15763, Mar. 1981.

* 6-02. Graham, M. H., and R. Luebbe: An Evaluation of MSS P-Format Data Registration. DC-Y1-04069, NSTL/ERL-197, Apr. 1981.

* 6-03. Stoner, E. R., G. A. May, and M. T. Kalcic: Evaluation of Multiband, Multitemporal and Transformed Landsat MSS Data for Land Cover Area Estimation. DC-Y1-04089, NSTL/ERL-196, Apr. 1981.

* 6-04. An Analysis of Landsat MSS Scene-to-Scene Registration Accuracy. DC-Y1-04156, Dec. 1981. NTIS: 82N24563.

* 6-05. An Algorithm for Automating the Registration of USDA Segment Ground Data to Landsat MSS Data. DC-Y1-04211, Dec. 1981. NTIS: 82N24564.

* 6-06. Thematic Mapper Simulator Data Collected Over Eastern North Dakota. DC-Y1-04232, Apr. 1982.

* 6-07. Evaluation of Small Area Crop Estimation Techniques Using Landsat- and Ground-Derived Data. DC-L2-04264, JSC-17829, LEMSCO-17597, Apr. 1982.

6-08. Semi-Annual Program Review Presentation to Level 1, Interagency Coordination Committee. DC-J2-04273, JSC-18223, Apr. 20, 1982.

* 6-09. Automated Segment Matching Algorithm Theory, Test and Evaluation. DC-Y2-04325, June 1982.

6-10. SCS Urban Curve Numbers from a Landsat Data Base. DC-R2-04353, Oct. 1981.

* 6-11. Analysis of Data Acquired by Synthetic Aperture Radar and Landsat Multispectral Scanner Over Kershaw County, South Carolina, During the Summer Season. DC-Y3-04374, NSTL/ERL-213, Mar. 1983.

* 6-12. Analysis of Data Acquired by Synthetic Aperture Radar Over Dade County, Florida, and Acadia Parish, Louisiana. DC-Y2-04396, NSTL/ERL-218, Dec. 1983.

* 6-13. Land Cover Change Monitoring Within the East-Central Louisiana Study Site - A Case for Large-Area Surveys With Landsat Multispectral Scanner Data. DC-Y3-04418, NSTL/ERL-221, June 1983.

* 6-14. Stratification of Sampled Land Cover by Soils for Landsat-Based Area Estimation and Mapping. DC-Y3-04439, NSTL/ERL-224, Sept. 1983.

* 6-15. A Correlation Analysis of Percent Canopy Closure vs. TMS Spectral Response for Selected Forest Sites in the San Juan National Forest, Colorado. DC-Y3-04440, NSTL/ERL-212, Nov. 1983.

* 6-16. Classification and Area Estimation of Land Covers in Kansas Using Ground-Gathered and Landsat Digital Data. DC-Y3-04441, NSTL/ERL-225, Dec. 1983.

Plans

6-17. Domestic Crops and Land Cover Implementation Plan. DC-J1-C0619, JSC-17109, 1981.

6-18. Domestic Crops and Land Cover Implementation Plan. DC-J1-C0626, JSC-17149, 1980.

6-19. Domestic Crops and Land Cover Project Implementation Plan for Fiscal Years 1982 and 1983. DC-U1-C0648, JSC-17799, Nov. 1981.

6-20. Domestic Crops and Land Cover Project Implementation Plan for Fiscal Year 1983. DC-U2-C0660, JSC-18585, Nov. 1983.

Unnumbered Documents

6-21. Allen, R.: USDA Registration and Rectification Requirements. NASA Workshop on Registration and Rectification, Nov. 17, 1981.

6-22. Colwell, R. N., R. W. Thomas, L. H. Beck, C. E. Brown, S. L. Wall: California Landsat-Aided Crop Estimation. Annual Report, 1 Jan. 1982 - 15 Jan. 1983, Space Sciences Laboratory, Univ. of Calif. (Berkeley), 1983.

6-23. Cook, P.: Landsat Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service, 1972-1982. 1982 American Statistical Association Meetings (Cincinnati, Ohio), Aug. 1982.

6-24. Craig, M. E.: Area Estimates by Landsat, Arizona 1979. Statistical Research Division; Economics, Statistics and Cooperatives Service, USDA (Washington, D.C.), 1980.

* 6-25. Erickson, W.: The Microcomputer Work Station: An Alternative Hardware Architecture for Digital Image Processing. Eighth William T. Pecora Memorial Remote Sensing Symposium (Sioux Falls, S. Dak.), Oct. 1983.

6-26. Hanuschak, G., and C. Caudill: Management Issues of Integrating Earth Resource Satellite Data Into the United States Department of Agriculture's Domestic Crop-Area Estimation Program. 1983 Meeting of the Institute of Management Science/Operations Research Society of America (Orlando, Fla.), Nov. 1983.

6-27. Hanuschak, G., R. Allen, and W. Wigton: Integration of Landsat Data Into the Crop Estimation Program of USDA's Statistical Reporting Service, 1972-1982. Eighth Int. Symposium on Machine Processing of Remotely Sensed Data, Purdue Univ. (W. Lafayette, Ind.), July 8, 1982.

* 6-28. Hofman, L. B., W. K. Erickson, and W. E. Donovan: MIDAS: A Microcomputer-Based Image Display and Analysis System With Full Landsat Frame Processing Capabilities. Seventeenth Int. Symposium on Remote Sensing of the Environment (Ann Arbor, Mich.), May 1983.

6-29. Holko, M.: 1982 USDA Results for the California Cooperative Remote Sensing Project. Statistical Reporting Service, USDA (Washington, D.C.), 1984.

6-30. Kleweno, D. D., and C. E. Miller: 1980 AgRISTARS DC/LC Project Summary Crop Area Estimates for Kansas and Iowa. ESS Staff Rep. AGESS 810414, Mar. 1981.

6-31. May, G.: Evolution of Land Cover Definitions and Survey for the Economics and Statistics Service. June 1981.

6-32. May, G., and R. Allen: Non-Sampling Errors in Non-Agricultural Strata of the 1980 Kansas June Enumerative Survey. Oct. 1981.

6-33. May, G., and D. Kleweno: 1981 Kansas Land Cover Survey Manual. ESS Research Division (Washington, D.C.).

6-34. May, G., M. Holko, and E. Jones: State Level Landsat Classification and Area Estimates of Major Crops and Land Covers in Missouri. 1983.

6-35. Mergerson, J. W.: Crop Area Estimates Using Ground-Gathered and Landsat Data. ESS Staff Rep. AGESS 810223, Feb. 1981.

6-36. Mergerson, J. W.: Crop Area Estimation Using Ground-Gathered and Sampled Landsat Data. ESS Staff Rep. AGESS 810408, May 1981.

6-37. Mergerson, J., G. Hanuschak, and P. Cook: Applications of Satellite Remote Sensing for U.S. Crop Acreage Estimation, 1980-81 Results. Sixteenth Int. Symposium on Remote Sensing of the Environment (Ann Arbor, Mich.), May 1982.

6-38. Mergerson, J., V. B. Johnson, and R. A. Kestle: AgRISTARS DC/LC Applications Project: 1982 Winter Wheat Area Estimates for Colorado, Kansas and Oklahoma. Ninth Int. Symposium on Machine Processing of Remotely Sensed Data, Purdue Univ. (W. Lafayette, Ind.), June 1983.

6-39. Mergerson, J., C. Miller, M. Ozga, M. Holko, S. Winings, P. Cook, and G. Hanuschak: 1981 AgRISTARS DC/LC Four State Project. Eighth Int. Symposium on Machine Processing of Remotely Sensed Data, Purdue Univ. (W. Lafayette, Ind.), July 1982.

6-40. Mergerson, J., S. Winings, P. Cook, V. Johnson, G. May, M. Holko, and R. Kestle: 1983 AgRISTARS DC/LC Large Scale Crop Acreage Estimation Project. Statistical Reporting Service, USDA (Washington, D.C.), 1984.

6-41. Ozga, M., and R. Sigman: An Autodigitizing Procedure for Ground-Data Labeling of Landsat Pixels. Fifteenth Int. Symposium on Remote Sensing of the Environment (Ann Arbor, Mich.), May 1981.

6-42. Ryerson, R. A., R. S. Sigman, and R. J. Brown: Satellite Remote Sensing for Domestic Crop Reporting in the United States and Canada: A Look to the Future. Seventh Canadian Symposium on Remote Sensing (Winnipeg, Manitoba), Sept. 9, 1981.

6-43. Sigman, R.: Crop Area Estimate From Landsat and Ground Survey Data. Canadian Federal Provincial Committee on Agriculture Statistics (Ottawa, Canada), Mar. 1980.

6-44. Sigman, R., and M. Craig: Potential Utility of Thematic Mapper Data in Estimating Crop Areas. Fifteenth Int. Symposium on Remote Sensing of the Environment (Ann Arbor, Mich.), May 1981.

6-45. Sigman, R., and G. Walker: Comparison of Auxiliary Variables for Satellite-Data Regression Estimators in Crop Surveys. Proc. Section on Survey Research Methods, 1983 American Statistical Association Meetings (Washington, D.C.), Sept. 1983.

6-46. Walker, G., and R. Sigman: County-Level Crop Area Estimates From Landsat: Evaluation of the Battese-Fuller Estimator. Proc. Section on Survey Research Methods, 1982 American Statistical Association Meetings (Washington, D.C.), 1982.

6-47. Walker, G., and R. Sigman: Evaluation of the Huddleston-Ray and Battese-Fuller Estimators. ESS Staff Rep. AGESS-820909, Statistical Reporting Service, USDA (Washington, D.C.), Sept. 1982.

6-48. Winings, S. B.: Estimating Potatoes and Other Crops in the Red River Valley of North Dakota and Minnesota Using 1980 Landsat Imagery. ESS Staff Rep. AGESS 810519, May 1981.

6-49. Winings, S.: Landsat Image Availability of Crop Area Estimation. Eighth Int. Symposium on Machine Processing of Remotely Sensed Data, Purdue Univ. (W. Lafayette, Ind.), Statistical Reporting Service, USDA (Washington, D. C.), July 1982.

6-50. Winings, S., P. Cook, and G. Hanuschak: AgRISTARS DC/LC Applications Project: 1982 Corn and Soybeans Area Estimates for Iowa and Illinois. Seventeenth Int. Symposium on Remote Sensing of the Environment (Ann Arbor, Mich.), May 1983.

6-51. Zuttermeister, J. P., and R. Sigman: The Effect of Cubic Convolution Resampling on Crop Area Estimates. ESS Staff Rep. AGESS-831209, Statistical Reporting Service, USDA (Washington, D. C.), Dec. 1983.

REMOTE SENSING PUBLICATIONS
BY COOPERATING INSTITUTIONS

Cornell University:
Cornell Laboratory for Environmental Applications of Remote Sensing

1984

Philipson, W.R., V.L. Williams, D.K. Gordan and W.D. Philpot. Vegetable and fruit tree inventory with Landsat TM data: Preliminary results. pp. 39-46. In: Proc. 51st Ann. Mtg. Amer. Soc. Photogrammetry, ASP, Falls Church, Va.

Philipson, W.R., D.K. Gordan, W.D. Philpot and V.L. Williams,. Field reflectance calibration with white, grey and block standard reflectors. pp. 47-54. In: Proc 51st Ann. Mtg. Amer. Soc. Photogrammetry, ASP, Falls Church, Va.

1985

Gordan, D.K. An investigation of thematic mapper satellite imagery for inventorying fruit trees in New York. M.S. thesis. Cornell University.

1986

Williams, V.L., W.R. Philipson and W.D. Philpot. Scaling field reflectance measurements to satellite pixels. (submitted for publication)

Williams, V.L. M.S. thesis. (in progress)

Iowa State University:
Statistics Department

1981

Battese, G.E. and W.A. Fuller. Prediction of county crop areas using survey and satellite data. Report, ISU Statistical Laboratory, April 1981 and In: Proc 1981 Annual Meeting of the American Statistical Association's Survey Section, ASA, Washington, D.C.

1982

Fuller, W.A. Regression estimators for crop area estimation. Report, ISU Statistical Laboratory, January 1982.

Battese, G.E. and Fuller, W.A. An error components model for prediction of county crop areas using survey and satellite data. Report No. 15, ISU Statistical Laboratory, February 1982.

Amemiya, Y. and W.A. Fuller. Maximum likelihood estimators of covariance componets in the multivariate balanced one-way random model. Report, ISU Statistical Laboratory, October 1982.

1983
Hung, H.M. Use of transformed Landsat data in regression estimation of crop acreages. Ph. D. dissertation. ISU.

Harter, R.M. Small area estimation using nested-error models and auxiliary data. Ph. D. dissertation. ISU.

University of Califoria at Berkeley:
Space Sciences Laboratory

1983
Thomas, R.W., L.H. Beck, C.E. Brown, S.L. Wall. California Landsat-aided crop estimation. Annual report, 1 January 1983 - 15 January 1983.

1984
Thomas, R.W., L.H. Beck, C.E. Brown, S.L. Wall. Annual Report, California Landsat-aided crop estimation. Report for the period March 1983 to February 1984.

Thomas, R.W., L.H. Beck, C.E. Brown, S.L. Wall. Development of a satellite-aided crop acreage estimation and mapping system for the state of California. In: Proc Eighteenth Internation Symposium on Remote Sensing of Environment, ERIM, Ann Arbor, MI

University of Houston at Clear Lake:
Mathematics Department

1985
Chhikara, R.S. and J.J. Mckeon. Regression estimators in sampe surveys. Preliminary report. University of Houston at Clear Lake.

McKeon, J.J., R.S. Chhikara, and T.L. Boullion. Linear regression estimators in sample surveys under calibration. In: Proc 1985 Annual Meeting of the American Statistical Association's Sample Survey Section. ASA. Washington, D.C.

1986
Chhikara, R.S., A.G. Houston, and J.C. Lundgren. Crop acreage estimation using a Landsat-based estimator as auxiliary variable. IEEE Journal of Geoscience and Remote Sensing. January 1986. (In press.)

NEW YORK ORCHARD INVENTORY

I] BACKGROUND

Once every five years the New York State Statistical Office (NYSSO) inventories the fruit orchards in the state. This inventory consists of obtaining the number of trees by fruit, variety, and age; the number of acres by fruit; and the number of orchards.

The procedure followed for the past two inventories (1975 and 1980) is the summary of a census of all the orchards known to the state office. The published numbers are the summary results with some imputation for incomplete reports. A snowballing question has been used to improve the coverage of the list, however no check was made to determine completeness.

II] PROBLEM

For the 1985 inventory the state office wanted to introduce some probabilistic method that would give them an idea of the incompleteness of the list used for the current inventory. First the June Enumerative Survey (JES) was reviewed to determine if this survey could be used. Only six of the 360 sampled units in the JES reported any type of fruit, therefore it seemed unlikely that the JES, or any sample based on the Master Area Frame, would be of use.

Concurrently the Remote Sensing Branch (RSB) was supporting research at Cornell University to investigate the potential use of Landsat Thematic Mapper data for identifying orchards in New York. Because of our efforts the NYSSO asked us to develop a completeness check for the current inventory.

III] FRAME DEVELOPMENT

An early attempt to use a Landsat classification procedure developed at Cornell was abandoned because of Cornell's inability to process sufficient Landsat data. Because of the urgent need for the data a New York Orchard Area Frame (NYOAF) was developed using 1983 and 1984 ASCS compliance photography. This frame, and the subsequent survey were confined to 5 counties (containing 65% of the states orchards) because of limited resources.

The frame was developed as follows:

1) Staff at Cornell University photointerpreted ASCS color slides and transferred boundaries of all identified orchards to 7.5 minute USGS quadrangle maps. These maps were then mailed to Washington.

2) RSB personnel reviewed the boundaries outlined at Cornell and grouped fields into polygons ranging from 1 to 110 acres in size. RSB personnel digitized these polygons and calibrated them to the USGS map base.

3) Areas for each polygon were determined and the polygons were listed by size. Polygons were then combined to generate a listing of frame units. Each frame unit consisted of one or more polygons such that each unit covered from 60 to 110 acres. The majority of these units ranged from 60 to 75 ares.

4) The frame units were stratified by county. The county stratification served two purposes. First the fruit types and distribution are highly variable across counties. Secondly the stratification would allow us to control sample size by county to ensure some measure of county level precision.

5) A sample of 100 units was allocated across counties. The minimum acceptable county level sample size was 10 units. The allocation of remaining units was determined by historic proportion of the aggregate orchard acreage.

The current New York Area Sample Survey (NYASS) was devised to serve two primary purposes.

1) To supply some measure of completeness for the current New York Orchard Inventory.

2) To supply a research data set to be used to evaluate and refine procedures developed at Cornell to identify orchards using Landsat Thematic Mapper data.

IV] NEW YORK ORCHARD SURVEY DESIGN

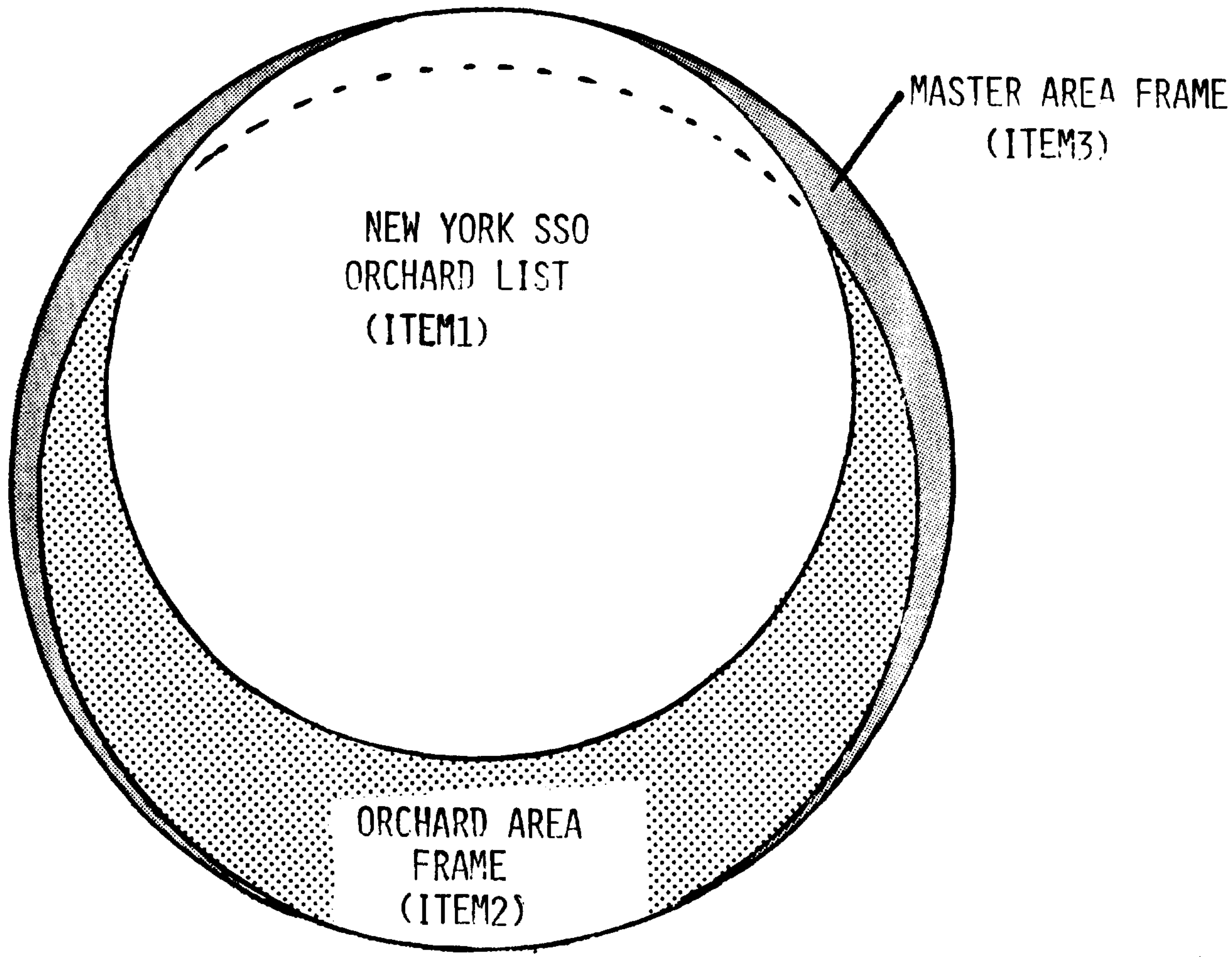
A] ESTIMATION OPTION ONE

The derived plan for this years orchard inventory is a simple multiple frame design. The population of orchards is divided into three subpopulations the aggregate of which guarantees complete coverage of all orchards. The three subpopulations are:

1) The orchards contained on the NYSSO's list of orchards.

2) The orchards contained in the NYOAF and not on the NYSSO's list.

3) Orchards contained in the New York Master Area Frame (NYMAF used for the JES) that are not in subpopulation 1 or 2 (Note that this frame is complete therefore the aggregate of subpopulations 1,2 and 3 is complete.



(ITEM)
ESTIMATION OPTION ONE

ESTIMATES: (Item can be acres of apples or number of Rome apple trees that are 1 to 3 years old etc.)

Item = Item1 + Item2 + Item3

$$\text{Item1} = \sum_{i=1}^{N1} y1_i$$

$y1_i$ = the data reported for Item for orchard i on the NYSSO's list.

N1 = The number of orchards on the NYSSO's list.

$$\text{Item2} = \sum_{c=1}^{5} \sum_{j=1}^{m_c} e_c \, y2_{jc}$$

$$e_c = \frac{M_c}{m_c}$$

M_c = the number of units in county c of the NYOAF.

m_c = the number of units sampled in county c of the NYOAF.

$y2_{jc}$ = the sum of all data reported for Item for orchards in subpopulation 2 for sample unit j in county c.

$$\text{Item3} = \sum_{s=1}^{8} \sum_{k=1}^{n_s} e_{sk} \; y3_{sk}$$

e_{sk} = the expansion factor for segment k in land use strata s of the NYMAF.

$y3_{sk}$ = the sum of all data reported for Item for orchards in subpopulation 3 for sample unit k in land use strata s of the NYMAF.

VARIANCES:

Var(Item) = Var(Item1) + Var(Item2) + Var(Item3)

Var(Item1) = 0

$$Var(Item2) = \sum_{c=1}^{5} e_c^2 f_c \sum_{j=1}^{m_c} \frac{(y2_{jc} - \overline{y2}_c)^2}{(m_c - 1)}$$

$$f_c = \frac{(M_c - m_c)}{M_c \; m_c}$$

$$Var(Item3) = \sum_{s=1}^{8} \sum_{k=1}^{n_s} f_s \frac{(y'_{sk} - \overline{y'_s})^2}{(n - 1)}$$

$$y'_{sk} = e_{sk} \, y3_{sk}$$

$$\overline{y'_s} = \frac{1}{n_s} \sum_{k=1}^{n_s} y'_{sk}$$

$$f_s = \frac{(N_s - n_s)}{N_s \; n_s}$$

N_s = the number of units in strata s on the Master Area Frame.

ESTIMATE OF COMPLETENESS

$$C = \frac{Item1}{Item}$$

VARIANCE:

Using Taylor's expansion the large sample aproximation to the variance is:

$$Var(f(x)) = (f'(x))^2 \; Var(x)$$

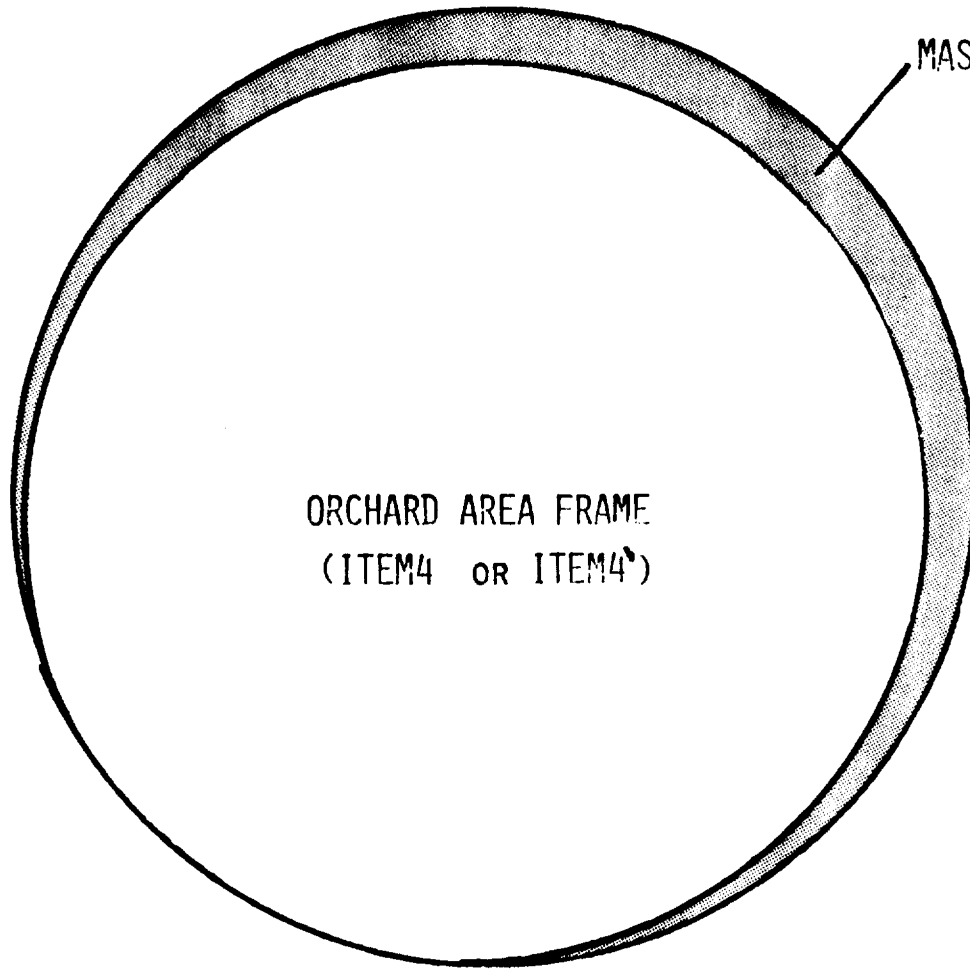
$$Var(C) = \frac{(Item1)^2}{(Item)^4} \; Var(Item) \quad = C^2(\text{relative variance of Item})$$

B] ESTIMATION OPTION TWO

The population of orchards is divided into two subpopulations the aggregate of which guarantees complete coverage of all orchards. The two subpopulations are:

1) The orchards contained in the NYOAF and not on the NYSSO's list.

2) Orchards contained in the New York Master Area Frame (NYMAF used for the JES) that are not in subpopulation 1 (Note that this frame is complete therefore the aggregate of subpopulations 1 and 2 is complete.



(ITEM)
ESTIMATION OPTION TWO

ESTIMATES: (Item can be acres of apples or number of Rome apple trees that are 1 to 3 years old etc.)

Item = Item4 + Item5

$$\text{Item4} = \sum_{c=1}^{5} \sum_{j=1}^{m_c} e_c \, y4_{jc}$$

$$e_c = \frac{M_c}{m_c}$$

M_c = the number of units in county c of the NYOAF.

m_c = the number of units sampled in county c of the NYOAF.

$y4_{jc}$ = the sum of all data reported for Item for orchards in subpopulation 1 for sample unit j in county c.

$$\text{Item5} = \sum_{s=1}^{8} \sum_{k=1}^{n_s} e_{sk} \; y5_{sk}$$

e_{sk} = the expansion factor for segment k in land use strata s of the NYMAF.

$y5_{sk}$ = the sum of all data reported for Item for orchards in subpopulation 2 for sample unit k in land use strata s of the NYMAF.

VARIANCES:

$$\mathrm{Var(Item)} = \mathrm{Var(Item4)} + \mathrm{Var(Item5)}$$

$$\mathrm{Var(Item4)} = \sum_{c=1}^{5} e_c^2 f_c \sum_{j=1}^{m_c} \frac{(y4_{jc} - \overline{y4}_c)^2}{(m_c - 1)}$$

$$f_c = \frac{(M_c - m_c)}{M_c \; m_c}$$

$$\mathrm{Var(Item5)} = \sum_{s=1}^{8} \sum_{k=1}^{n_s} \frac{(y'_{sk} - \overline{y'_s})^2}{(n_s - 1)}$$

$$y'_{sk} = e_{sk} \; y5_{sk}$$

$$\overline{y'_s} = \frac{1}{n_s} \sum_{k=1}^{n_s} y'_{sk}$$

$$f_s = \frac{(N_s - n_s)}{N_s \; n_s}$$

N_s = the number of units in strata s on the Master Area Frame.

ALTERNATIVE ESTIMATES: (Item can be acres of apples or number of Rome apple trees that are 1 to 3 years old etc.)

Item = Item4` + Item5

$$\text{Item4`} = \sum_{c=1}^{5} e_c \, z_c$$

$$e_c = \frac{M_c}{m_c}$$

$$z_c = \overline{y4}_c + b_c(\overline{X}_c - \overline{x}_c)$$

b_c = least squares regression coefficient from regressing $y4_{jc}$ on x_{jc} .

$y4_{jc}$ = the sum of all data reported for Item for orchards in subpopulation 1 for sample unit j in county c.

$$\overline{y4}_c = \frac{1}{m_c} \sum_{j=1}^{m_c} y4_{jc}$$

x_{jc} = the size of sample unit j in county c.

$$\overline{x}_c = \frac{1}{m_c} \sum_{j=1}^{m_c} x_{jc}$$

$\overline{X}_c$ = the average size of all the units in county c.

M_c = the number of units in county c of the NYOAF.

m_c = the number of units sampled in county c of the NYOAF.

VARIANCES:

$$\text{Var(Item)} = \text{Var(Item4')} + \text{Var(Item5)}$$

$$\text{Var(Item4')} = \sum_{c=1}^{5} f_c \frac{(1 - r_c^2)}{(m_c - 2)} \sum_{j=1}^{m_c} (y4_{jc} - \overline{y4}_c)^2$$

r_c^2 = sample level coefficient of determination between y4 and x .

$$f_c = \frac{(M_c - m_c)}{M_c \, m_c}$$

DIGITAL FRAME DEVELOPMENT

Remote Sensing Branch (RSB) personnel are presently starting the development of a prototype image display and analysis system to aid area frame development. This proposed system is being considered because of some three relatively recent developments.

a] The development of the super microcomputer technology which allows us to consider processing large quantities of data in an interactive environment.

b] The success of high quality, high resolution sensors such as the Landsat thematic mapper and the linear array on the SPOT satellite.

c] The soon to be available Digital Line Graph (DLG) data being developed by the U.S. Geological Survey (USGS).

The key ingredient to our proposed system is the availability of digital map data. USGS in cooperation with the Census Bureau will be making available digital data representing the hydrography and transportation information seen on the 1:100,000 topographical map series. The joint project is being done so that Census can use the digital data for the 1990 population census. This means that, unlike some DLG projects by USGS, 100 percent coverage of the U.S. should be available by 1988 or 1989. The use of the DLG data allows us to use computer graphics procedures to combine ground observable features with the remotely sensed data.

We expect to use current image processing techniques to combine the DLG data with the satellite data in order to generate (on an image display device) an aerial photograph like image with map attributes. We feel that this generated image can be used to interactively delineate count unit boundaries. Thereby replacing the current use of Landsat paper products, county highway maps, photo index sheets, and other data sources currently used.

If we are successful some of the benefits of digital frame construction could be to:

a] Eliminate the current requirement of a seperate activity for digitization count unit boundaries to determine area. (Areas will be calculated automatically by the computer during delineation.)

b] Eliminate the error-prone transfer of boundaries between cartographic materials of varying scale (such as the transfer of boundaries from photo index sheets to county highway maps).

c] Improve the frame developer's analysis of land covers through image enhancement or classification techniques.

d] Provide better control over the homogeneity of count units within strata and possibly permit more flexibility in determining stratification strategies.

e] Allow development of frame updating techniques rather than requiring reconstruction of entire state frames.

f] Allow easier exploration of specialized area frames such as the work being tried in New York. Some of this specialization could incorporate automated remote sensing techniques or it could simply put to use the classification results generated as part of the remote sensing estimation program.

Department of the Interior
U.S. Geological Survey
National Cartographic
Information Center (NCIC)

National Mapping Program

US GeoData

1:100,000-Scale Digital Line Graph Sampler Now Available

As part of its work building a National Digital Cartographic Data Base of machine-readable data, offered for sale to the public as US GeoData, the U.S. Geological Survey is preparing to make available to the public US GeoData tapes containing digitized planimetric cartographic data (called Digital Line Graphs) from its 1:100,000-scale, 30- by 60-minute topographic map series. The data include hydrography and transportation.

Because of the amount of interest in the technical specifications and characteristics of the Digital Line Graphs (DLG's) from 1:100,000-scale maps, the USGS has produced a DLG Sampler of digital planimetric data from the Chickamauga 30- by 60-minute, 1:100,000-scale topographic map quadrangle. The quadrangle covers parts of Georgia, Alabama, and Tennessee. This US GeoData DLG Sampler is intended to familiarize potential users of the 1:100,000-scale data with the new product and allow them to experiment on their equipment and with their processing systems.

This flier describes briefly the 1:100,000 DLG Sampler and the process used to produce the files. It also identifies differences between this DLG Sampler and the DLG's to be produced using production systems within the USGS. The Sampler is available from the USGS for $25.00 from the Eastern Mapping Center National Cartographic Information Center.

The DLG Sampler

The Sampler contains digital planimetric data on hydrography and transportation for the 30- by 60-minute Chickamauga, GA-AL-TN, 1:100,000-scale topographic map quadrangle. The 1:100,000-scale DLG's are being packaged in 30- by 30-minute blocks; therefore the DLG Sampler includes the digital data for two 30- by 30-minute blocks (Chickamauga East and Chickamauga West).

Within this Sampler, each DLG block contains two categories: hydrography and transportation. The transportation category is further subdivided into three data overlays: (1) roads and trails, (2) railroads, and (3) pipelines and power lines (or miscellaneous transportation). Therefore, each 30- by 30-minute unit is composed of four data overlays.

Additionally, each 30- by 30-minute block is composed of four files per data overlay. Each file contains the data for one overlay for a 15- by 15-minute portion of the 30- by 30-minute block. Each overlay (hydrography, roads and trails, railroads, and miscellaneous transportation) is file separated from the other overlays. Thus, in the DLG Sampler there are 16 files (four hydrography, four roads and trails, four railroads, and four miscellaneous transportation) in each 30- by 30-minute block.

The 1:100,000-scale DLG's are available in two formats—standard ASCII DLG and optional ASCII DLG—as described in the "Digital Line Graphs from 1:100,000-Scale Maps" Data Users Guide, available from the Survey's NCIC's.

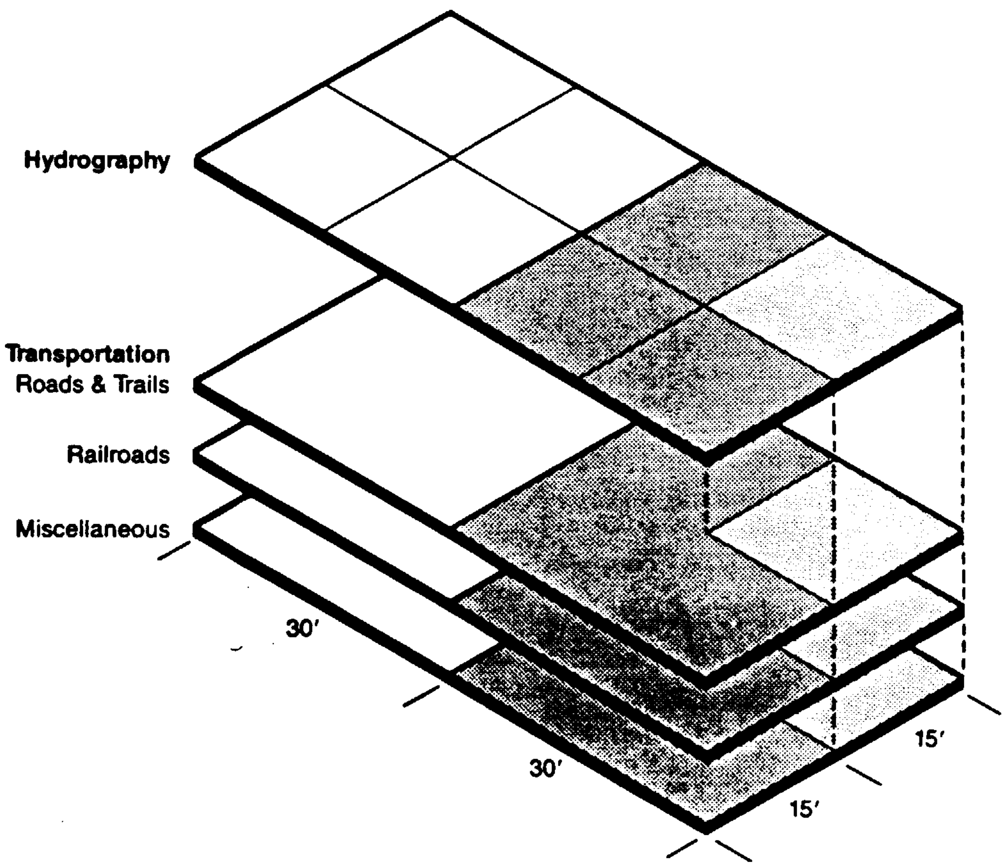


Diagram Showing Organization of Overlays and Files for the DLG Sampler

(Continued on back)

Production of the DLG Sampler

The hydrography and the road and trail data were originally digitized in 7.5-minute sections. The data sets were merged manually using interactive graphics equipment into 15-minute sections. The data were then processed to generate the DLG Sampler.

The railroad and miscellaneous transportation data were originally digitized in 30-minute sections. The files were divided into 15-minute sections by bisecting the bounding map edges with a straight line. The files were then processed to create the DLG Sampler.

After the 15-minute DLG 3 ASCII (144 byte/record) files were created, they were processed through a conversion routine to create DLG optional (80 byte/record) versions of the data.

Differences Between Sampler and Final Product

This DLG Sampler of 1:100,000-scale data has several differences from the final 1:100,000-scale DLG's to be made available:

- Software is being developed to perform edge checks against the edges of the four adjacent 15-minute sections in the 30-minute blocks. Flags in the header of the DLG file will be set by the software and indicate the status of the edge matching process. The flags in this Sampler have been manually entered but are representative of the type of codes that will be set with software for the final DLG's. These flags are described in more detail in the "Digital Line Graphs from 1:100,000-Scale Maps" Data Users Guide, available from the NCIC's.
- As described earlier, the 30-minute sections of railroad and miscellaneous transportation data files were divided by a manual process on an interactive graphics system. In the future, this division will be handled by partitioning software.
- The edges between this DLG Sampler and adjoining 1:100,000-scale maps were not checked. The edge flags in the header indicate this fact.
- In a 30-minute block, the adjacent 15-minute sections join within the restrictions of the file coordinate system. Lines will join each other within a tolerance of 3 mils. It is anticipated that the final data produced through software techniques will address this problem.

Orders and Further Information

Comments, suggestions, or orders for the DLG Sampler, as well as requests for further information, should be directed to

Eastern Mapping Center NCIC
U.S. Geological Survey
536 National Center
Reston, VA 22092
703-860-6336 FTS 928-~~6336~~ 6645

AVHRR

SRS PLAN FOR CROP CONDITION RESEARCH USING REMOTE SENSING

This research will be conducted under the Memorandum of Understanding signed by the Foreign Agricultural Service (FAS), Agricultural Stabilization and Conservation Service (ASCS), Agricultural Research Service (ARS) and the Statistical Reporting Service (SRS). The project involves joint use of the FAS remote sensing facilities. SRS's primary interest is in determining the feasibility of using weather data and digital satellite data from Landsat and weather satellites for crop condition assessment. In order for this methodology to be useful for domestic applications, the output must be statistically defensible, objective, repeatable and timely.

The first phase of the SRS research will be to develop some familiarization with the FAS image analysis system and data sources. This will involve determining the capabilities of the equipment and software. In addition, types of data available, frequency of coverage, timeliness of data acquisition and the amount of historic data available will be reviewed. This phase will have some continuing aspects, as new and unique procedures are explored and additional data are acquired.

Among the possible products that might be developed are maps of floods, freeze damage or other factors which affect crops in localized areas. These products have the potential for enhancing the Weekly Weather Crop Report. Possible relationships between indexes that can be generated from satellite data and indications that are available from SRS surveys will be investigated.

Vegetative indexes and indexes of soil moisture, plant stress, biomass and evapotranspiration will be compared with survey indications such as crop condition, yield and plant counts. Vegetative indexes will be related to crop condition from the SRS Weekly

---

Prepared for Planning Purposes, December 5, 1985

Weather Crop Survey and forecast yields from monthly SRS surveys. This approach might provide useful geographic information on crop variability. Relationships will be explored first to determine the usefullness of this approach at the state level, since SRS survey data is more precise at that level of aggregation. The research will concentrate on corn and soybeans in Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Nebraska, Ohio, South Dakota, Wisconsin and perhaps a few neighboring states where suitable data are available. If results at the state level are promising, an attempt will be made to develop products for crop reporting districts and counties since regular survey information is more limited at these lower levels of aggregation.

As a first step in exploring this approach, meterological satellite imagery will be sought in a restricted temporal window around August 1, 1985. The SRS yield forecasts for the August 1, 1985 reference date will be regressed on various state level vegetative index (and other functions of meteorological satellite channel data) means. If the relationship between forecast yields and imagery values for the same reference date are strong, then an attempt will be made to produce a calibrated image value for each meteorological satellite pixel. The calibration values would scale the original pixel values (using the regression relationship) so that they correspond with the ground data based state level yield forecasts. The product could be represented in a smoothed map of the region showing general areas where the crop was in better condition (imagery related to a higher yield forecast) or where the crop was in poorer condition. The approach allows the condition indication to correspond to the forecasted yield at the state level, but relies on the satellite imagery to define a further geographic distribution. Such a product would have potential in terms of the additional geographic information that could be provided to data users or, internally, as an aid in identifying unique areas for greater scrutiny in future surveys.

In exploring the kinds of relationships which can be developed, digital satellite data for various dates will be compared. Multi-temporal versus unitemporal coverage will also be investigated. It will be necessary to look at historic data for several years to determine the feasibility of this approach. Regression techniques will be used to develop mathematical relationships between condition or forecast yield and index values. To be most useful, it is important that baselines for condition or yield be established which provide a basis of gauging relative change from one period to another. It will be necessary to tie these indications to a map base containing state, crop reporting district and possibly county boundaries in order to provide a useful product in support of improved crop condition assessment and yield forecasts. These map based values may be of use with yield or other crop condition variables in producing more precise regression estimators or contributing as valuable supplemental or auxiliary variables in other ways.

A more sophisticated approach which may be explored is the possibility of obtaining actual harvested acreage, yield and production for June Enumerative Survey segments at harvest. These data could be used with vegetative index numbers calculated for the segments in a procedure similar to that now used by SRS for obtaining crop acreages from remote sensing. Segment crop yields would be regressed on the mean vegetative index for pixels classified in the crop, and segment production on an indication created by multiplying the number of pixels classified to a crop by the average vegetative index for the same pixels. The effort would attempt to add improved (more precise) estimates of acreage actually harvested, yield and production to the improved (planted or intended for harvest) acreage estimates currently provided by the combined use of remote sensing and statistically valid ground data. Multispectral scanner and/or thematic mapper data from Landsat would be used for this approach. This approach would involve collection of additional (farmer reported) ground data from operational SRS surveys. It would also involve additional data processing which would need to be accomplished in time for use in setting final domestic yield estimates. An initial effort in this area could involve obtaining the additional ground data for a single State and completing the required data processing steps on a retrospective basis.