INTEGRATION OF LANDSAT DATA INTO THE CROP ESTIMATION

PROGRAM OF USDA'S STATISTICAL REPORTING SERVICE

1972 - 1982

GEORGE A. HANUSCHAK
RICHARD D. ALLEN
WILLIAM H. WIGTON

U.S. DEPARTMENT OF AGRICULTURE
STATISTICAL REPORTING SERVICE
RESEARCH DIVISION
WASHINGTON, D.C. 20250

INVITED PAPER AT THE 1982 MACHINE PROCESSING OF
REMOTELY SENSED DATA SYMPOSIUM, PURDUE UNIVERSITY,
WEST LAFAYETTE, INDIANA, JULY 8, 1982
INTEGRATION OF LANDSAT DATA INTO THE CROP ESTIMATION PROGRAM OF USDA'S STATISTICAL REPORTING SERVICE 1972 - 1982

GEORGE A. HANUSCHAK
RICHARD D. ALLEN
WILLIAM H. WIGTON

U.S. DEPARTMENT OF AGRICULTURE STATISTICAL REPORTING SERVICE RESEARCH DIVISION REMOTE SENSING BRANCH WASHINGTON, D.C. 20250

JULY 1982
INTEGRATION OF LANDSAT DATA INTO THE CROP ESTIMATION PROGRAM OF USDA'S STATISTICAL REPORTING SERVICE

1972 - 1982

I. ABSTRACT

This report describes how NASA's LANDSAT data has been integrated into the USDA-SRS domestic crop estimation program in the last ten years. Since the launch of LANDSAT I in 1972, SRS has investigated the potential contribution of earth resources satellite data to its domestic crop estimation program. SRS use of area sampling frames provides a statistically sound and cost-effective crop estimation program even without LANDSAT data. Thus, the integration of LANDSAT data has had to show potential for improving the statistical reliability of an already sound program. In the last ten years, SRS has discovered two major uses of LANDSAT data for its program. The first is photo interpretation of LANDSAT MSS and RBV image products for broad land use stratification in construction of area sampling frames. The second major use of LANDSAT MSS data is to classify the digital data into crop types and regress SRS ground collected data results from the area frame sampled segments onto the classified LANDSAT data for each crop type. The degree of success of these improved regression estimates depends heavily on optimum timing of LANDSAT coverage, the extent of cloud cover during the optimum window, and the rapid delivery of LANDSAT data to SRS. This report is a summary of the SRS experience to date under various program structures such as joint NASA-SRS efforts, LACIE, and AgRISTARS plus a brief look to the future.

II. BACKGROUND

SRS has been an extensive user of remote sensing products since the 1950's when it began using aerial photography in the construction and use of area sampling frames. Thus, SRS since the 1950's has had an intense interest in being aware of and using the best and most cost-effective photographic products available for its purposes. The primary source of these products has been USDA's Agricultural Stabilization and Conservation Service (ASCS). SRS's primary uses have been the use of aerial photography mosaics for broad land use stratification and eight-inch equal one mile rectified aerial photographs for improving the quality control of data collection procedures. There were considerable research efforts conducted by SRS and other research institutions such as NASA, LARS, ERIM, and the University of California at Berkeley in the 1960's and early 1970's that investigated the potential of aerial photography and aerial multispectral scanner data to meet SRS information needs. Topics addressed were diverse and included tree counts, fruit counts, livestock counts, and measuring the effects of corn blight. Thus, the previous research by SRS and these other research institutions and the use of its area sampling frame procedures put SRS in a position to address the research needed after the launch of LANDSAT I in 1972 for domestic crop estimation.

III. METHODOLOGY

Methodology used by SRS in its domestic crop estimation program has been described in several papers.4-7,9 Major methods used include area frame construction and sampling. Approximately 15,700 stratified and randomly selected sample units about 0.7 of a square mile in size called segments have crop, livestock, and economic data collected for them each year in late May and early June during SRS's June Enumerative Survey (JES). The segment data is expanded to state, regional and national totals by using a direct expansion estimator as described in Cochran. Surface area of this segment is measured in the field. The degree of success of these improved regression estimates depends heavily on optimum timing of LANDSAT coverage, the extent of cloud cover during the optimum window, and the rapid delivery of LANDSAT data to SRS. This report is a summary of the SRS experience to date under various program structures such as joint NASA-SRS efforts, LACIE, and AgRISTARS plus a brief look to the future.

RE of the regression estimator is the multiplier that would be necessary to increase the sample size of the JES in order to get an estimate just as precise from the ground data alone (direct expansion) estimator as the ground data plus LANDSAT (regression) estimator. For example, there are 435 JES segments in Kansas. If the JES plus LANDSAT regression estimator has an RE of 2.0, then
it would have taken 870 JES segments to obtain the same precision from a ground data only (direct expansion) estimator.

IV. INTEGRATION OF LANDSAT DATA INTO SRS'S PROGRAM

A. 1972-1973 DEVELOPMENTS

In 1972, SRS was funded to undertake an ERTS 1 investigation. This funding was the first time SRS personnel had a chance to 'hands on' experience at satellite data digital processing. Ten full-time persons focused entirely on this effort. The proposal was to develop new ways to use LANDSAT type data to improve estimates of crop acreage and at the same time compare estimates from ground sampling, LANDSAT and aerial photography. One crop reporting district (CRD) in each of four states (South Dakota, Kansas, Missouri and Idaho) was selected as the test area. A CRD is usually one ninth of a state and contained up to fourteen counties for the selected CRD's in these four states.

There were several reasons why the ERTS 1 investigation was difficult. No one who worked on the project had studied remote sensing techniques, remote sensing principles or scanners, pattern recognition or photo interpretation. The assigned personnel were statisticians and computer programmers. A multi-disciplinary team with experience in remote sensing would have facilitated the research effort.

As previously stated, the JES provides timely, precise and cost effective statistical estimates. On the other hand, in order to separate corn from soybeans using LANDSAT MSS digital data, early August data was best. By the time the data is acquired by NASA, preprocessed and sent to EROS and mailed to USDA, often a month or more has passed. Once SRS gets it, the LANDSAT must be reformatted, registered to maps, and digitally processed. Thus, any LANDSAT based estimate couldn't be more timely than the JES.

Great care was taken to collect the best support data possible. All selected land parcels were flown to obtain high altitude aerial photography. Additional follow-up interviews were conducted and data on crop maturity collected. This extra data was useful and the results and findings were substantial. A product generated during this study was a table of field size estimates. Since the SRS sample was representative of all agriculture, the data was used to provide valid estimates of field sizes. With normal Agency procedures, field level data was not keypunched and thus, it was lost. These tables generated much discussion since that question is crucial in determining the spatial resolution of a sensor. The aerial photography that was collected was used to monitor ground data collection but this data was also scanned with a microdensitometer and processed on a minicomputer. This process took hours and the results were poor because of the unstable platform. This effort was judged to have little potential compared to MSS data and was phased out.

Probably the most substantial finding was that a methodology, the regression estimator, was developed that would improve late season acreage estimates if imagery was usable and available. This methodology, impacted the remote sensing community because it was one of the first applications where results were statistically quantitative and not simply a pixel counting algorithm or a best opinion. USDA-SRS was able to use classification results of data that contained a thirty percent misclassification rate with ground data and produce an estimate with a two to three percent coefficient of variation (CV) using the regression estimator. This model is now being used by several institutions both inside and outside the United States.10,11,16

Even though the regression estimator was solid, that first year's data didn't particularly give good results. Because of late season LANDSAT imagery collected in 1972, results were less than optimum. Methodology was developed during the original ERTS 1 investigation, but better timed data was needed.

B. 1974 DEVELOPMENTS

In 1974, SRS chose a Texas site covering thirteen counties for its LANDSAT related research.12 Most passes of LANDSAT data, however, were cloud covered. Only June 27 data was of sufficient quality to be useful. Thus, the entire thirteen county area was reduced to six counties. It was at this point that the simple truth about cloud cover on LANDSAT was beginning to be recognized. Often good agricultural producing areas are cloud covered enough of the time so that obtaining cloud free LANDSAT imagery can be a difficult problem.

C. 1975 ILLINOIS PROJECT AND EDITOR SOFTWARE DEVELOPMENT

Starting in 1974, SRS and the University of Illinois' Center for Advanced Computation (CAC) entered into a joint agreement to redesign a software system. The base software system was an interactive version of Purdue University's LARSYS developed by CAC and U.S. Department of Interior. The system (EDITOR) was redesigned to analyze SRS's ground gathered data, NASA's LANDSAT data and calculate the regression estimates. Data processing was to be done at several locations and used several systems and networks. SRS's June Enumerative Survey (JES) data was processed on a UNIVAC 1108 on the INFONET network and also on an IBM 370 at the USDA's Washington Computer Center (WCC). The LANDSAT data was processed first at CAC on an IBM 370 and then on PDP-10s at the Bolt, Beranek and Newman (BBN) Data Processing Center in Cambridge, Massachusetts on the ARPANET and finally on the LLILAC-IV at the NASA-Ames facility at Moffett Field, California also on the ARPANET. A detailed paper of the initial capabilities of the EDITOR system was published in June 1977.13

The first major project to develop and test the EDITOR system was SRS's project in Illinois using 1975 JES data and LANDSAT data. It took two years to design, develop and test the system to do regression estimates, digitize, plot and check SRS segment data, evaluate full scene classification estimates for all the LANDSAT data available for Illinois during the appropriate window (July 15 - Sept. 15).
Also, in this project many issues were addressed on the amount, quality and type of ground data necessary to do regression estimates for all areas in an entire state for which LANDSAT data was available. The summary of this extensive research was published in several papers in 1976 and 1977. Highlights of the project were studies of:

1. the statistical methodology and the software to do entire states,
2. the amount, quality and type of ground data needed in future efforts,
3. the effects of clustering, prior probabilities and jackknifing,
4. the registration methodology for LANDSAT data, and
5. the estimates and corresponding sampling errors of the regression estimates at the state and substate levels.

A more detailed description of highlights follows:

1. The area sampling frame was digitized so that regression estimates could be computed according to the stratified by land use/cover random sampling design used to select SRS area frame samples. Software was developed to digitize, plot and check SRS area frames and also selected sample segments. Registration software was developed to locate JES field boundaries on LANDSAT data tapes to within one-half to two pixels. EDITOR was further developed to extract a random sample of LANDSAT pixels to develop LANDSAT segment data. A statistical methodology to handle cloud cover and missing data problems of LANDSAT was also implemented. The methodology was to post stratify the universe (state) into two post strata (cloud free and cloud covered). The regression estimate was then used in the cloud free post stratum and a direct expansion estimate (JES ground gathered segment data only) was used for the cloud covered post stratum. Two detailed papers on this methodology were published in 1976 and 1977. SRS has since used this methodology in all full state LANDSAT projects.

2. The JES segments were visited in June, July, August, and September for any nonsampling error revisions and actual changes in farmer's intentions or fields. At the outset SRS realized that this would be a one time maximum effort to determine ground data needs for future LANDSAT crop acreage estimation projects. Regarding the four visits, it was determined that the August and September visits were not cost effective improvements over the July data. There were very few changes in the farmer's intentions and fields after the July visit.

3. The effects of clustering, prior probabilities and jackknifing were also studied. A clustering algorithm using Swain-Fu distance was used to find clusters within a known crop type. The major concern with its use was that there were many subjective decisions to be made by the data analyst. Analysts made decisions on the maximum and minimum number of clusters, the separability distance measure and separability threshold, and amount of thresholding to unknown clusters. The next decision was whether to use prior probabilities proportional to estimated crop area for the universe or equal probabilities of occurrence. Empirical results using both methods showed no consistent superiority of one method over the other based on \( R^2 \) (sample correlation coefficient between classified LANDSAT data and JES ground gathered data) or \( RE \) as the performance criterion, not percent correct classification. Thus, SRS still uses both methods in 1982 and chooses the best method (highest \( R^2 \) or \( RE \)) for the state and crop of interest. Another statistical concern was that training data should be independent from the test or estimation data. Thus, a LANDSAT pass in Western Illinois with 29 wholly contained counties and approximately 85 JES segments was used to test for bias in the \( R^2 \) or \( RE \) due to non-independent train and test data sets, a procedure commonly called jackknifing. Results showed for major crops and land covers, such as corn, soybeans and pasture, that the bias was negligible for this size of training data set. When the sample size is small, however, this assumption may not hold. As a result of this study SRS decided to use the same segment data to train the classifier and estimate crop area totals in future full state studies. In most of these studies, however, some partial holdout-jackknife testing was done to check for any significant biases in the \( R^2 \) or \( RE \).

4. Registration of LANDSAT data to a map base had changed considerably from the 1972 project. Third order polynomials yielded one-half pixel root mean square (RMS) errors for line location and two pixel RMS errors for column location for the 1975 Illinois Project. A detailed paper on the history of LANDSAT data registration procedures used by SRS from 1972 - 1981 is soon to be published.

5. The state was divided into six analysis districts. The reason for subdividing the state's land area is that LANDSAT data is quite sensitive to atmospheric changes (haze, clouds, wet ground etc.). Thus, LANDSAT data taken one day apart may have a different regression slope and intercept when JES data is regressed on the LANDSAT classified data. The reduction in variance when comparing the regression estimates to the direct expansion estimates were encouraging for the haze free August LANDSAT imagery, but showed little improvement for the July and September LANDSAT imagery. The August window success was consistent with previous results from the CITARS study. \( 1.00 \) to 6.11 for corn and from 1.26 to 2.97 for soybeans. County estimates were also calculated but the coefficients of variation ranging from ten to 100 percent were too large to warrant use by SRS.

D. 1976 DEVELOPMENTS

The next major studies planned were full state efforts for both Illinois and Kansas in 1976. Due to cloud cover, no usable LANDSAT data was acquired in August and thus the full state Illinois project was abandoned. However, a substate area was analyzed with multispectral LANDSAT data in March and September, and the results were encouraging even with the
September data as the second image. The LANDSAT data available for estimating the 1976 winter wheat crop area in Kansas was excellent and acquired in April and May. The project was described in detail in an SRS paper in August, 1978. Highlighs of the project were:

1. Only 40 percent of the regular JES sample was used to train the classifier and to estimate winter wheat acreage. Thus, an almost opposite direction in the amount of ground data acquired for the 1975 Illinois study was taken for the 1976 Kansas study. The effort was designed to answer the question as to what might be the minimum amount of ground data to arrive at an efficient LANDSAT regression estimator. LANDSAT coverage was good in terms of total area covered (87 out of 105 counties) but some wheat intensive areas were not available during the April-May window. Thus, SRS decided to use 100 percent of JES segments in future full state remote sensing studies for crop area estimation.

2. Several refinements in methodology were implemented in the 1976 Kansas project. Rather than a straightforward pooling of land use strata data, a combined regression estimator as described in Cochran was implemented. Since there were not enough sample segments in nonagricultural land use strata within individual analysis districts to calculate reliable contribution analysis districts were quite large, ranging from seven to twenty-five counties. Researchers felt that based on the 1976 Illinois project and the serious cloud cover problems encountered that the analysis district sizes might not always be that large. When analysis districts reach the one to five county size range serious problems arise with the lack of statistical degrees of freedom or in other terms too small of a sample of JES segments. Thus, SRS decided to use 100 percent of JES segments in future full state remote sensing studies for crop area estimation.

3. The results from the 1976 Kansas project were very encouraging. It was the first SRS project that had both excellent percent correct classification as well as high relative efficiencies. The average overall percent correct classification was 80 percent and the average relative efficiency for wheat acreage was 6.25. Also for the first time the precision of the SRS LANDSAT based regression county level estimates was encouraging. The coefficients of variation for the county level estimates ranged from 6.1 percent to 38.8 percent with an average of 16.7 percent.

E. 1977 DEVELOPMENTS

For the 1977 crop year SRS chose Kings County, California to test how quickly SRS's LANDSAT regression estimator could be calculated. The goal was to determine if an entire state could be processed in a timely fashion for the 1978 crop year. The results were very encouraging. Relative efficiencies for major crops ranged from 20.0 to 28.0. Even more encouraging was the time required to accomplish the task. This project provided the first evidence that SRS could perhaps calculate regression estimates for a full state prior to its Annual Crop Summary.

F. 1978 IOWA PROJECT

Based upon the success in California, SRS in the 1978 crop year decided to try to do an entire state in a timely fashion provide the 1978 Iowa estimates as input to their Crop Reporting Board's Annual Crop Summary. Iowa was the state chosen and corn and soybeans were the major crops of interest. Highlights of the 1978 Iowa project were:

1. All available LANDSAT data was processed in time with the SRS's JES data to be input to SRS's Annual Crop Summary issued January 16, 1979. The data for the Annual Crop Summary was barely set in this first effort as the regression estimates were available as of the first week of January.

2. The estimates at the state and analysis district levels were substantially more precise than those made from ground data alone JES estimates. Relative efficiencies at the state level were 2.43 and 2.38 for corn and soybeans respectively. Relative efficiencies at the analysis district level ranged from 0.93 to 5.98 for corn and from 2.73 to 7.59 for soybeans.

3. Although results were very encouraging, SRS decided not to plan for an operational program at that time due to problems with total project cost, delivery of LANDSAT data to SRS, and the potential extent of cloud cover. Total project cost for the 1978 Iowa project was approximately $300,000. The June Enumerative Survey cost was $162,500 which did not make the regression estimate a cost effective improvement at the time of the 1978 Iowa project. As cited in the detailed paper on the 1978 Iowa project, there were delays as long as three months in obtaining some of the LANDSAT data. Clouds covered 13 of the 99 counties in Iowa for the available LANDSAT data.

4. There were several first SRS research efforts in addition to crop area estimation associated with the 1978 Iowa project. One of these efforts was to determine if the classified LANDSAT MSS data from the crop area effort could provide any additional information to SRS's objective yield models. The results of the research were that at the full state level the relative efficiencies of the estimator that used objective yield plus LANDSAT data compared to an objective yield only estimator for both corn and soybeans were less than 1.10. That is, there was minimal additional information provided by the LANDSAT data for a state level SRS objective yield model. Perhaps the Thematic Mapper on SPOT will warrant future research efforts in the yield estimation area. The second new SRS research effort associated with the 1978 Iowa project was an attempt to determine if the regression estimator (ERS plus LANDSAT) could provide information for forest inventories. A joint effort by SRS and USDA's Forest...
Service personnel was conducted using the eastern third of the 1978 Iowa data set. Forest Service personnel photo interpreted current aerial photography of SRS's JES segments to classify the forested areas by variables such as type, size, and density classes. Relative efficiencies for total forestland for the two eastern Iowa regression estimates were 4.36 and 5.45, however the regression land use change was substantially larger than the 1974 Forest Service survey results. The major hypothesis concerning the difference is that the definition of forestland for the 1978 study was not the same as the 1974 survey definition.

G. USE OF LANDSAT IMAGERY IN AREA FRAME CONSTRUCTION

The second major use by SRS of LANDSAT was to photointerpret LANDSAT imagery for broad land use (cover) differentiation or stratification in the area sampling frame construction process. In a series of papers on the basis of the area sampling frame construction process first outlined the potential use of LANDSAT for this purpose.31,32 Then in 1976 a pilot study was set up for a portion of California to test the feasibility of using LANDSAT along with conventional tools (maps, aerial photography, historic agricultural data, etc.) to build an area sampling frame. The results of the pilot study showed a definite contribution of LANDSAT imagery in the area sampling frame construction process.33 The contribution was based on current LANDSAT imagery being a superior source for stratification over older aerial photography in areas where there was considerable land use change over a short period of time. The old aerial photography and the LANDSAT imagery. The LANDSAT imagery because of limited resolution however could not replace the aerial photography in the finer breakdown of land use strata into count units and sampling units. SRS decided to build a new area sampling frame for California using LANDSAT along with all the conventional tools. The new frame and sample for the 1979 JRS in California yielded about the same level of precision as the old frame for major items of interest but with 150 less sample segments.34 Based upon the California experience, new area sampling frames were built for Oregon, Washington, Michigan dry bean producing area), Idaho and Texas during 1980-82 using LANDSAT imagery along with all the conventional tools.

H. 1978 IDAHO AND ARKANSAS PROJECTS

Also in the 1978 crop year SRS conducted substrate research studies in Idaho and Arkansas aimed at potato and small grain area estimation in Idaho and rice, soybean, and cotton area estimation in Arkansas. These studies were not designed to be done on a timely basis but to investigate more crops in different geographical regions for the LANDSAT regression estimator. For the substrate areas analyzed in Idaho35, using July 18 and July 26 LANDSAT data, relative efficiencies for spring wheat, winter wheat, barley and potatoes respectively ranged from 3.29-4.95, 1.75-2.88, 1.45-1.67 and 1.23-5.62. For the one substrate analysis district in Arkansas, using June 30 LANDSAT data, the relative efficiencies for rice, soybeans, and cotton were 3.32, 2.46, and 1.70 respectively.

I. 1979 DEVELOPMENTS

For the 1979 crop year, SRS chose substrate areas of Arizona and South Dakota for crop area estimation with the LANDSAT regression estimator. In Arizona using all available July LANDSAT data, relative efficiencies for cotton and sorghum for the substrate analysis districts ranged from 2.02-6.07 and 1.0-5.07 respectively. In the LANDSAT MSS data backlog at NASA's Goddard Space Flight Center was quite large at the time of this project. Out of 91 possible LANDSAT scenes from July 1-September 7, 1979 only 16 had been received by the U.S. Department of Interior's EROS Data Center by mid August. No major or agricultural counties were eliminated from the project. For the 1979 South Dakota project the increasing acreages planted for sunflowers was the item of interest.36 Using August 25, LANDSAT data for a six county area the relative efficiency was 3.7. Another objective for the South Dakota project was to study the effects of different soil types on crop signatures. The soil studies was a joint project by the area's Remote Sensing Institute (RSI) of South Dakota State University. The results of the study were mixed and varied by crop type and the amount of biomass at the time of the LANDSAT data and further research on this topic is necessary.37

J. SRS ROLE IN LACIE

From 1974-1978, SRS also participated in a major interdepartmental effort called the Large Area Crop Inventory Experiment (LACIE). LACIE was designed to demonstrate the potential of applying space borne aerospace remote sensing data in meeting USDA's information needs.40 LACIE was planned by NASA as a joint effort with USDA, and the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce (USDC). SRS had several roles in LACIE. One role was to provide historic and current agricultural statistics for the U.S., state and agricultural counties. Another role was to provide several statisticians to be trained in remote sensing techniques. The third role was to evaluate the statistical methodology used in LACIE at the request of the U.S. Office of Management and Budget. Overall, the SRS management felt that the LACIE project highlighted the need for more basic research. Previous SRS and Forest Service research efforts along with LACIE provided the momentum for the continuation of the development of remote sensing techniques as related to agricultural statistics by USDA's NASA, USDC/NOAA, and USDI/ERS in the new AgRISTARS program.

K. AGRISTARS

AgRISTARS is a six year (1980-1985) interagency research program designed to evaluate the potential of aerospace remote sensing data in meeting USDA information needs.41 Multi-agency teams of scientists were assigned to LACIE at the request of the U.S. Department of Agriculture to address the following research topics: 1. early warning and crop condition assessment, 2. foreign commodity production forecasting, 3. yield model development, 4. domestic crops and land cover, 5. renewable resources inventory, 6. soil moisture, 7. conservation and pollution, and 8. supporting research.
SRS has overall management responsibility for the AgrISTARS interagency coordinating committee and program management team. Within the structure of the interagency AgrISTARS research program, SRS also assumed leadership for the Domestic Crops and Land Cover Project (DCLC). SRS also has a supporting role in several of the seven project areas other than DCLC.

L. 1980 AgrISTARS Domestic Crops and Land Cover

The DCLC project is an excellent prototype for the AgrISTARS program. Basic concepts had been developed in the 1977-77 period and proof of concept demonstrated in the 1978 Iowa project. However, the procedures used in 1978 which required almost all work to be done by the SRS Research Division and were labor intensive would not be feasible operational procedures. Thus, a basic conceptual model for improvement of crop area estimates for major crops provided a definite focus for the research effort.

The most visible research element of the DCLC project was the plan to make crop area estimates for two states in 1980 and add two more states to the effort in each succeeding year for a total of ten states in 1984. Kansas was selected in 1980 for winter wheat estimates along with Iowa for corn and soybeans. This aerial photography is acquired in most agricultural counties and Conservation Service (ASCS) Offices. This aerial photography was helpful for determining current field boundaries. The plots were helpful for determining crop area estimates for winter wheat estimates for Kansas for the end of the crop time to create more precise winter wheat acreage by ASCS for use in their planted photography. Similar slides of photography were able to perform all needed operations quite successfully.

1. The first procedures transferred to the SRS State Statistical Offices (SSO's) were the ground data processing procedures. Performing the detailed edits in the SSO's meant that all relative materials such as questionnaires and enumerator notes were readily available for solving any discrepancies and the time and cost expenses of preparing duplicate materials for performing the edit elsewhere were avoided. The two SSO's were provided with digitization equipment and plotters. All plotter segment and field boundaries were digitized within the respective offices in order to create the field location files needed by the Research Division for matching with LANDSAT data and for selection of training fields. These slides were helpful for determining current field boundaries.

2. A new editing tool available in 1980 and utilized in both states was current 35 mm aerial photography. Color slides of photography were obtained from the respective Agricultural Stabilization and Conservation Service (ASCs) Offices. This aerial photography is acquired in most agricultural counties across the country by ASCS for use in their planted acreage compliance certification programs. These slides were helpful for determining current field boundaries.

3. The DCLC efforts to process LANDSAT data in time to create more precise winter wheat acreage estimates for Kansas before the end of the crop estimate season meant that SRS was processing 1980 imagery before most other researchers. Thus, SRS identified geometric distortions in 1980 processed LANDSAT data which were not obvious in casual review of image products or were not obvious if LANDSAT data were res planned by users. SRS was able, through the AgrISTARS communications channel, to inform NASA's Goddard Space Flight Center personnel of

4. Improvements from the use of the LANDSAT data for acreage estimates were much less than expected (based on earlier years' studies) in both Kansas and Iowa. State level RE for winter wheat in Kansas was 1.33. The range of RE's at the substate level was 1.20 to 3.05. State level RE's were 1.65 and 1.51 for corn and soybeans in Iowa respectively. The range of RE's at the substate level was from 1.37 to 6.40 for soybeans and from 1.61 to 3.01 for corn. Much of the deterioration in results was due to LANDSAT data processing problems in 1980 which could be corrected.

The other important aspect of the 1980 work, involvement of the SSO's in editing and digitization, was successful and provided encouragement for expansion to four states in 1981.

A new feature of the DCLC project to SRS was the expansion of research capabilities through the work of three NASA research centers which were participants in DCLC. The centers were Johnson Space Center (JSC), National Space Technologies Lab (NSTL), and the Ames Research Center (ARC). In addition, SRS continued its strong internal research efforts.

Highlights of the output from these research efforts in 1980 were:

1. During 1980 the main DCLC effort at JSC was an evaluation of the CLASSY algorithm as a possible replacement for the existing SRS clustering approach. A parallel comparison study was carried out for an area of Missouri for which SRS had multitemporal clustering and classification results for 33 frame segments. The parallel study, although limited by having only 33 segments, indicated that CLASSY identified nearly the same number of clusters for each category of crop type or land cover as present procedures and the correlation for each category was the same or numerically higher than original results. Based on 1980 results, CLASSY was placed into the EDITOR system to be used as the main clustering procedure for 1981 projects.

2. At NSTL registration techniques had been developed which utilized an image processing system for selection of registration points. SRS's approach was similar except that points were selected from grey scales or from image products to match against known locations on map products. The major difference in NSTL/SRS multitemporal registration procedures was the resampling algorithms used (nearest neighbor in the SRS case and cubic convolution in the NSTL case). One of the important first steps in registration research was to reach agreement on accuracy definitions and calculations. During 1980 concepts which could lead to algorithms for automated registration were discussed between SRS and NSTL.

1982 Machine Processing of Remotely Sensed Data Symposium
3. Land Cover was a completely new area of interest for SRS. The concentration before had always been an estimate of crop acreage only. The DCLC project gave SRS a chance to evaluate and develop techniques which might answer information needs of other USDA or state level agencies. NSTL had prior experience in examining land cover and geographic information needs of other USDA or state level agencies.

The first technique that SRS wished to explore for land cover statistical estimation was an extension of the JES direct expansion and the JES plus LANDSAT regression estimator. It was expected that the present JES sample sizes would not be adequate since that sample is allocated mainly for crop and livestock estimation and the nonagricultural land covers might not have enough representation for training. In 1980 a special visit to the nonagricultural strata segments, plus a sampling of previously used segments in those strata, in Kansas was made to gather information on the basic land cover definitions of interest to USDA and others. Current color infrared aerial photography was available for use by enumerators in this special study.

The SRS requirements for land cover research were that any estimates created must be statistical estimates and any map products created must match corresponding point estimates. An SRS researcher suggested an approach of retaining classification probability information for each pixel in order to match the regression results rather than classifying each pixel only into the class for which it has the highest probability.

4. Efforts at ARC in 1980 were focused on a review of EDITOR procedures and the beginning of a study of alternatives for transferring some EDITOR procedures to an offline mode for better control of processing costs.

5. Several internal research efforts were conducted by SRS personnel in 1980. Topics addressed included sampling the LANDSAT pixels, multispectral classification, and estimating potato acreage in the Red River Valley of North Dakota. Highlights of these research efforts were the following:
   a. The study on sampling LANDSAT pixels showed promise for less than full scene classification.
   b. The study on multispectral classification yielded the expected improved results in precision for the regression estimator. However, it is difficult to overcome the cloud cover problems in obtaining both spring and late summer imagery without making the substate analysis districts too small.
   c. In the Red River Valley study an additional sample of smaller JES type segments was used in addition to the JES segments to improve precision of potato acreage estimates. However, the only summer imagery available was dated September 6, 1980 and was outside the optimum window for potatoes.

H. 1981 AGRISTARS DOMESTIC CROPS AND LAND COVER

In 1981, the major crop acreage element of DCLC expanded to four states. The new states were Oklahoma and Missouri. A detailed paper on this 1981 AGRISTARS DCLC Four State Project is also found in the proceedings for this symposium. Highlights of the 1981 crop area work are as follows:

1. In expansion of the major crop acreage work to Oklahoma and Missouri in 1981, state office procedures were similar to those used in 1980. The 1981 project was marked by a number of equipment failures. It was necessary to use several backup procedures and locations in order to complete digitization in time for analysis. The adjustments made did indicate the feasibility of digitization in several locations by digitizing from field boundaries placed on acetate overlays rather than having to ship actual original aerial photographs.

2. A greater percentage of Kansas imagery was obtained in 1981 than in the 1976 and 1980 projects. Coverage of nearly 85 percent of the total wheat acreage was obtained. However, nearly 30 percent of the wheat acreage was covered by early March (March 5 and 6) imagery which was earlier than the desired temporal window. The RE at the state level was 2.3. Results by analysis district for the early dates were lower than normally expected in Kansas (1.9 and 2.2). The rest of the analysis districts had RE's ranging from 2.6 to 5.5. Similarly, in Oklahoma it was necessary to use some early March imagery. The RE at the state level was 1.35. No LANDSAT data was used in the eastern third of Oklahoma due to the small percentage of wheat there and late LANDSAT data delivery to SRS. RE's for the analysis districts with March images were 1.25 and 1.72. RE's for the other analysis districts ranged from 1.94 to 4.01.

The percentage of imagery available in Iowa was the lowest experienced of the three efforts to study Iowa (1978, 1980, and 1981). Only about 65 percent of the corn acreage and 60 percent of the soybean acreage in Iowa were covered by 1981 summer imagery. Results for soybeans in Iowa were encouraging with relative efficiencies ranging from 2.91 to 15.81 by analysis district. The state level RE was only 1.63 due to the amount of land area that was cloud covered. The state level RE for corn was 1.56 and analysis district RE's ranged from 1.25 to 5.06. The amount of corn and soybean acreage covered by LANDSAT in Missouri was slightly higher than in Iowa. Relative efficiencies varied considerably by analysis district for corn and soybeans respectively. State level RE's were 2.16 and 2.08 for corn and soybeans respectively. An interesting spinoff in Missouri was the calculation of improved acreage estimates for the rice crop. The entire rice crop in Missouri is contained on one LANDSAT scene and a good image was available for that scene. The relative efficiency for that analysis district was 6.0. Sorghum estimates using LANDSAT imagery were also calculated and relative efficiencies were 1.23 and 1.31 for scenes available.

3. CLASSY was used for all 1981 analyses. The procedure resulted in a standardization of procedures and in less analyst time than the previous approach. However, CLASSY used more computer resources than expected and each run took longer than desired to fully execute. As a result analysts had to further subsample pixels and the effects of this are being studied.

1982 Machine Processing of Remotely Sensed Data Symposium
4. In addition to CLASSY, other new procedures were implemented in 1981. Geometric quality of 1981 LANDSAT data was much better than experienced in early 1980. Both of the necessary registration stages, global scene registration and final segment location, were turned over to the SRS Remote Sensing Branch Support Staff. The Support Staff performed all global registrations. Point selection, editing, and checking required only one to three hours per scene. LANDSAT 1250,000 image products were used for each global registration. Several members of the Support Staff were trained to perform the final segment location on LANDSAT.

5. The delay in LANDSAT data delivery was the only major analysis problem. Many of the scenes that SRS needed for Iowa and Missouri were not in the data base until late September, October, or November and SRS was affected by the backlog of orders. In order to meet the Crop Reporting Board deadlines for consideration of the improved estimates in the Annual Crop Reports it was necessary for staff members to work considerably overtime to complete the analysis steps. It is hoped that data delivery will speed up in subsequent years since this excessive use of overtime is not a feasible operational procedure.

6. All full scene classifications were done at NASA-Ames on a CDC 7600 as the TILLIAC IV had been retired and the CRAY-1S wasn't available yet.

7. There were several promising research results during 1981. One that was not promising, however, was the evaluation of the quality of geometric information from LANDSAT data as received. If LANDSAT data which had been "registered" to ground control points was accurately enough registered geometrically (to within 2 or 3 pixels of true ground location) SRS would be able to skip the present global registration operation. However, an evaluation of scenes by SRS and NSTL personnel showed that quality was not uniformly high enough to use directly.

8. One procedure which SRS personnel conducted the research for and it showed great promise was "automatic digitization". Automatic digitization does not replace all manual operations but it uses a video camera, an image processing system, and a minicomputer to replace the present point mode of digitizing segment and field boundaries. Since the automatic digitization equipment is driven by a small minicomputer the operation also has the advantage of creating batched files off line from the main processing network. This additional cost savings feature means that automatic digitization is projected to pay for itself in two years or after eight states have been digitized. Automatic digitization equipment is no order for 1982 and plans are that it will be used for two states in 1982.

9. NSTL worked in 1981 to implement an algorithm for automatic shifting of segment and field boundaries to their true locations within LANDSAT data files.53 Testing was performed in 1981 for 20 segments. Results were promising for 17 of the 20 but there was no provision in the algorithm to identify cases for which the approach was "lost". Additional work has been done in the algorithm and additional tests appear even more promising. It is hoped to have this algorithm in EDITOR for testing in one state in 1982.

N. 1982 AGRESTARS DOMESTIC CROPS AND LAND COVER

In 1982, if LANDSAT data is available, the DCLC program will calculate regression estimates in six states (Kansas, Oklahoma, Colorado, Iowa, Missouri and Illinois) and continue to pursue its research efforts (both internal and jointly with NASA centers). Full scene classifications will be done on the CRAY-1S at NASA-Ames. Missouri will be the test site for a multitemporal crops/land cover area estimation research project. If LANDSAT data is not available SRS will step up its research efforts using data from new sensors, methodology and more cost effective data processing.

V. LOOK TO THE FUTURE

As previously outlined, SRS to date has found two major uses for LANDSAT imagery and data. The first use is to photo interpret image products for broad land use stratification in area sampling frame construction. As new sources of photo interpretation data become available, SRS personnel will evaluate their potential for cost effective improvements in area sampling frame construction. Currently planned new programs and sensors that show potential are as follows: (1) U.S. government's High Altitude Photography (HAP) program, (2) LANDSAT D and D1 - both MSS and TM image products, (3) Large Format Camera on a space shuttle mission and (4) French SPOT image products. The major source of concern by SRS is the cost of the production and geometric rectification of the products. The improved resolution of these products is exciting.

The second major use of LANDSAT data by SRS is the use of digital data, as previously described, in the form of an input to the regression area estimator. Products needed by SRS for this application are digital data tapes (preferably the raw data with some system geometric and radiometric corrections), photo like image products corresponding to the data on tape, and a 10 to 14 day turnaround. Currently planned new programs and sensors that show potential for this digital application are as follows: (1) LANDSAT D AND D2 - both MSS and TM data and (2) French SPOT data and (3) any future private sector LANDSAT MSS or TM missions. Research results from TM simulator data over Missouri in September 1979 indicated that the increased resolution would improve relative efficiencies by a multiplier of 2-3 for maize crops. That multiplier is times the current relative efficiencies being experienced with the MSS. That is the current MSS with MSS usually range from 1.5 to 3.5 at the state level. With the TM these would range from 3.0 to 10.5. The improved resolution of these products is very encouraging concerning results. However, major concerns of SRS regarding the future of the application of the crop area regression estimator remain as follows and are: (1) frequency of coverage during short one month optimum windows (2) the size and cost of "one scene" of data (3) rapid data delivery to SRS and (4) total project cost. Using the 1981 AGRESTARS DCLC Four State Project costs, the relative efficiency (RE) should be 2.5 or larger for a cost effective improvement in statistical precision.55 The SRS research results from 1975-1981 have been mixed concerning cost effectiveness. The majority of results have met the criterion of cost effective improvements.

1982 Machine Processing of Remotely Sensed Data Symposium
in precision using 1981 project costs. However, the cost of the new and higher resolution data and the data processing costs may be a major concern. Also, the smaller scene size and the frequency of coverage are also serious concerns at this time. The smaller scene size may cause problems with a shortage of statistical degrees of freedom for analysis areas. Frequent coverage during the relatively short (approximately one month) optimum temporal windows for crop area estimation is a critical need of SRS. Nine day coverage is a minimum satellite configuration for domestic crop area estimation. Also, the 10-14 day target average for data delivery to SRS of LANDSAT data and images has never been met for any SRS full state projects. Fourteen day turnaround is essential for future SRS projects.

Concerning full scene data processing, SRS plans to examine the capability and cost effectiveness of the CRAY-1S, CDC Cyber 205, and NASA Goddard Massively Parallel Processor (MPP).

The research and development with the MSS and extracting the agricultural statistics information took several years. It is expected that the same will be true for the new higher resolution sensors. However, the research and development time period is expected to be somewhat shorter due to experienced staff and data processing experience. As with the MSS though, the potential information content is exciting and the 1980s will be a decade of interesting and exciting research on these new sensors.

VI. ACKNOWLEDGMENTS

The authors wish to acknowledge the numerous contributions during the period 1972-1982 from various members of the following organizations:

1. Statistical Reporting Service (SRS) USDA
2. Laboratory for Applications of Remote Sensing (LARS) - Purdue University
3. National Aeronautics and Space Administration
4. Center for Advanced Computation - University of Illinois
5. EROS Data Center - U.S. Department of Interior
6. Iowa State University
7. University of California at Berkeley
8. Remote Sensing Institute (RSI) - South Dakota State University
9. Environmental Research Institute of Michigan (ERIM)
10. Agricultural Stabilization and Conservation Service (ASCS) - USDA

REFERENCES


34. FONT IN LARGER TYPE


1982 Machine Processing of Remotely Sensed Data Symposium


1982 Machine Processing of Remotely Sensed Data Symposium
AUTHOR BIOGRAPHICAL SKETCHES

1. George A. Hanuschak, Head, Applications Section, Remote Sensing Branch, SRS

EDUCATION

1971 B.A. Mathematics, Youngstown State University (Ohio)
1973 M.S. Mathematical Statistics, Ohio State University
1981 M.P.M. Management, Agriculture, Cornell University

EXPERIENCE

Statistics Instructor
1974 Des Moines (Iowa) Area Community College
1977 U.S. Department of Agriculture Graduate School

Mathematical Statistician - USDA/SRS

2. Richard D. Allen, Acting Director, Survey Division, SRS

EDUCATION

1963 B.S. Agricultural Economics, Iowa State University
1967-1968 Graduate Study Statistics, Iowa State University

EXPERIENCE

Agricultural Statistician

Mathematical Statistician - USDA/SRS
1968-1982 Research Division - Corn Blight Watch Experiment
Head, Yield Research Section
Assistant Statistician-in-Charge, Illinois SSO
Head, Crops, Prices and Labor, Methods Staff, Estimates Division
Leader, List Frame Project, Research Division
Chief, Remote Sensing Branch, Research Division

1982 Machine Processing of Remotely Sensed Data Symposium

3. William H. Wigton, Head, Research Section, Remote Sensing Branch, SRS

EDUCATION

1964 B.S. Mathematics, Marietta College, Marietta, Ohio
1971 M.S. Master of Statistics, North Carolina State University

EXPERIENCE

Mathematical Statistician
1967-1970 USDA/SRS, Research Division, Washington, DC