84-03 C

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:

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

64

$$\frac{\cancel{187} \times 2.5}{\cancel{187} + 180} = \frac{467.50}{367}$$

1.27

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.