Conference
Land use, Calif.

ECONOMIC FACTORS IN IMPLEMENTATION OF REMOTE SENSING INTO AGRICULTURAL INFORMATION SYSTEMS

Harold F. Huddleston and Robert M. Ray III

Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C. 20250

Center for Advanced Computation
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

JAN. 1978

Abstract

Methodological issues are discussed relevant to the cost effectiveness of remote sensing information systems implemented for monitoring agricultural crop production for geographic regions corresponding to U.S. states. Attention is given to methods by which computer-processed LANDSAT data can be integrated into existing cooperative Federal-state ground surveys to improve the accuracy, geographic detail and timeliness of production estimates for certain crops. A statistical methodology is illustrated by which the information gain afforded by LANDSAT for such estimates can be quantified. Where appropriate spatial sampling theory is practiced and adequate ground survey data is collected, these same techniques can be used for statistical estimation of other land cover characteristics discernible with LANDSAT. Specific land information needs vary widely across U.S. regions. Thus, remote sensing information systems utilizing LANDSAT in the prescribed manner should prove most cost-effective over time to local, state and Federal users where implementation occurs at the state level and is coordinated closely with existing Federal-state cooperative agricultural reporting services. Such systems should also prove exportable to less-developed countries where more rational management of agricultural and other land uses is of critical importance.

Introduction

Within the Department of Agriculture there are four service agencies that use remote sensing information in some form in providing the services they render. These are the Forest Service (FS), the Soil Conservation Service (SCS), the Agricultural Stabilization and

Conservation Service (ASCS), and the Statistical Reporting Service (SRS). Each of these four service agencies of USDA has its own separate operating costs specific to the basic services that it provides. In rendering these services, however, each agency functions in part through Federal-state cooperative programs where individual states are able to modify the basic Federal program according to its specific needs and fiscal commitments. Thus, due to the nature of the operation of these programs, there are necessarily two sets of economic costs and benefits, one set related to basic Federal objectives and another set related to state and local objectives. In addition, personnel of USDA service agencies are asked frequently to lend technical assistance to government agencies in other countries.

The development of orbital remote sensing platforms such as LANDSAT, in conjunction with rapid advancements in computational and digital telecommunication technologies, on the surface would seem to increase dramatically the quantity and quality of information products that USDA service agencies should be capable of supplying to public and private clients. The world-wide coverage of LANDSAT's digital multispectral imagery suggests comprehensive inventories of global agricultural production and surficial natural resources. Simultaneously, the nominal one-acre resolution of LANDSAT's sensors suggests detailed maps of land covers and land uses for localized regions.

The real difficulty in integrating the large volumes of remote sensing data available from LANDSAT into agricultural information systems lies in the interpretation of the data itself into useful information. Given the vast quantities of data available, it seems that computer-aided interpretation offers the most promising approach. In addition, the methodology appropriate for large geographic areas, such as a LANDSAT frame, simultaneously provides estimates for smaller geographic areas. While numerous demonstrations have been conducted illustrating the potential of computer interpretation of LANDSAT data, cost-effective remote sensing information systems have yet to be integrated within agency practice. In part this may be explained by the lack of any clear assignment of responsibilities for such systems among Federal, state and regional agencies. In the authors' opinion, such systems must be implemented at the state level within the context of specific Federal-state cooperative programs. It is at this level that Federally-supported ground surveys can most conveniently be coordinated with specific state needs. Equally important, state governments have both the political power and fiscal means to identify and address the specific land management problems for which they have primary responsibility.

AGRICULTURAL INFORMATION SYSTEMS

Primary Goal of an Agricultural Information System

The four service agencies referred to earlier provide information in the form of tables, graphs and maps corresponding to various geographic units at state and national levels. Each data source may be stored in the system based on the original data collection units as elements or layers in a complex data base by different geographic formats, such as discrete points, two-dimensional iso-lines, topographic contour lines, grids, or by small irregular or regular areas related to a base map. The different information elements can represent historical data series, current-year or current-crop data, or real-time data which enters the data base at the time the analysis is needed or to be performed. The analysis module may derive the designated information from the data base and output the desired characteristics in one of the geographic formats corresponding to those used in imputation of the data. An improved geographic-based information system would serve many common interests of the SCS, FS, and SRS since these agencies are all involved in statistical surveys. The frequency of the surveys will differ by agency and each will have its own kind of sampling units that are peculiar to their surveys. Likewise, the sample designs will have many features in common such as stratification, multi-stage sampling, and estimation techniques. Consequently, the costs and software can be shared by the several Federal agencies as well as state and local agencies with similar problems.

We will concentrate on the specific goals of one agency, the Statistical Reporting Service (SRS), which has elements common to the other agencies. The primary goal of an agricultural information system is the accurate estimation of agricultural production for political units or geographic units both large and small within a state and for the nation. Objectives supporting this goal are: (1) increased accuracy of the estimates, (2) increased timeliness of estimation, and (3) increased information concerning geographic distribution of various agricultural crop productivities. The cost of developing systems to the agency must be balanced against increased information gains. Experience to date suggests that estimation of productivities for several crops must be output from the system, since the principal cost is for personnel required to conduct ground surveys for a wide variety of agricultural products in delivering the traditional services of the agency. Consequently, the goal to improve accuracy, timeliness, and geographic resolution of reporting formats at reasonable costs poses the following question: For how many crops can the utilization of LANDSAT significantly improve the quality of statistical estimations?

Three potential benefits to the estimation of crop production from LANDSAT remote sensing technology are: (1) more frequent updating of state-wide landscape stratification to achieve more efficient spatial sampling, (2) more accurate crop acreage estimation, and (3) improved yield estimates. It should be noted that crop production information is a result of these three operations and better accuracy in any or all operations will result in improved outputs from the system. In some cases only a small improvement in one operation will result in a fairly large improvement in the production due to the multiplicative nature of the estimation improvements across operations.

Methodology to Achieve these Goals

At present information systems exist for providing crop production information. Current systems involve straightforward use of statistical sampling techniques and direct expansion of survey results in relationship to an area sampling frame constructed for an entire state. More comprehensive agricultural information systems will involve more sources of data that are correlated in some manner, with production or characteristics being estimated and related to the same area sampling frame. With regard to the particular crop characteristic being estimated, e.g., the acreage of an individual crop, the remote sensing and other information must exist in the system as an auxiliary or supplementary variable bearing a known relationship to the particular characteristic of the crop produced. The key to using such auxiliary variables efficiently with ground surveyed crop acreage data is through a procedure known as double sampling within the state sampling frame. For this type of problem, LANDSAT data is appropriate as an auxiliary variable, because it simultaneously gives potentially a wall-to-wall sampling of land covers with a relatively small resolution element of one acre. In addition, through traditional surveys of crop reporting services, "ground truth" data is obtained that may be used for calibrating computer classification procedures by which LANDSAT data may be interpreted.

To be effective the double sampling estimation technique depends upon good correlation between ground observations and computer LANDSAT classifications over samples of ground truth fields. Earlier papers[1,2], as illustrated below, indicate applications of double sampling with survey data and LANDSAT to individual crops result in very large or very small information gains depending on the crop being estimated. Preliminary studies indicate that yield estimates may also be improved through double sampling, but here information gains, though positive, are found to be substantially less. It should be noted that although information gains are typically computed for areas approximately the size of one LANDSAT frame because of the desire for statistical accuracy, the technique also yields a large amount of detailed land cover information for smaller areas such as counties, townships, and even individual fields. Where appropriate ground truth information is collected, nonagricultural as well as agricultural land covers may be estimated with the same procedures.

Example of Crop Production

The estimation or forecasting of domestic crop production is a major concern to USDA. For Federal purposes, current information is generally considered adequate for states and major production areas, but for state purposes more information is desired for small geographic areas within states. In this section, a method of combining SRS ground survey data and LANDSAT data to improve crop production estimates is illustrated in such a way that the gain in information, as measured by estimation accuracy, can be stated quantitatively. The basic approach has been to seek an integration of software systems developed jointly by SRS and CAC for more cost-effective machine interpretation of LANDSAT data with existing SRS ground surveys. Software systems have been implemented explicitly for interactive digitizing, storage and retrieval of large quantities of crop-acreage information collected routinely by SRS in the course of the extensive field surveys associated with its ongoing agricultural production estimation methodology.[1]

The technique illustrated is potentially more cost-effective for crop production estimation because LANDSAT sampling covers large geographic areas and the cost of obtaining field data for training computer classification procedures would not be significantly increased over that already collected for existing crop acreage and yield estimates. However, where LANDSAT data is implemented within estimation systems, slight modifications in existing field data collection procedures may well lead to substantially more efficient procedures for estimating crop productions. Similarly, small area or county estimates of acreage and yield may be obtained and used to derive county production estimates not now provided by SRS.

Acreage Estimation Procedure. Following ILLIAC IV classifications of all LANDSAT pixels for a specific analysis district, e.g., a particular set of counties wholly contained within a LANDSAT frame, a classified pixel total for each crop type is determined for the entire analysis district itself as well as each county. Classification results for each crop type are also aggregated to obtain individual totals for all segments sampled (ground units observed or enumerated) within the district.

An estimator[2] of the total acreage for a particular crop in a particular analysis district i and its sampling error is then computed as follows. The total acreage is estimated using the double sampling methodology for each stratum or analysis district as

$$\hat{A}_i = N_i(\bar{a}_i - \hat{B}_i(\bar{x}_i - \bar{X}_i))$$

and, assuming a sufficiently large sample of segments, the variance is

$$V(\hat{A}_i) = N_i^2\, V(\bar{a}_i)(1 - r_i^2)\left(\frac{n_i - 1}{n_i - 2}\right) .$$

For individual analysis districts, the normal approximation for small samples is used, that is $V(\hat{A}_i)$ for large samples multiplied by $(1 + \frac{1}{n_i - 3})$

where $\hat{A}_i$ = total acres of the crop within all area segments contained within the i^{th} analysis district

N_i = total number of all segments contained within the i^{th} analysis district (known from sampling frame)

n_i = the number of ground area segments sampled in the i^{th} district

$\bar{a}_i$ = average number of acres of the crop reported per area segment for all n_i area segments sampled in the i^{th} district

$\bar{x}_i$ = average number of pixels classified into the crop per area segment for all n_i area segments sampled in the i^{th} district

$\bar{X}_i$ = average number of pixels classified into the crop per segment over all possible segments for the i^{th} district n_i area segments sampled in the i^{th} district

y_{ij} = number of acres of the crop enumerated for j^{th} segment sampled in the i^{th} district

x_{ij} = number of pixels classified into the crop for the j^{th} segment sampled in the i^{th} district

$\hat{B}_i$ = the regression coefficient between y_{ij} and x_{ij} based on the

r_i^2 = correlation coefficient squared between y_{ij} and x_{ij} for the i^{th} district

$$V(\bar{a}_i) = \{\sum_{j=1}^{n_i} y_{ij}^2 - [(\sum_{j=1}^{n_i} y_{ij})^2 / n_i]\} / n_i(n_i - 1) .$$

Based on 33 segments falling in 10 western Illinois counties within LANDSAT image ID#2194-16042 of August 4, 1975, estimates of crop acreages and estimate errors were computed for several crops or land cover types. The estimates are shown in Table 1 and their squared sampling errors in Table 2. The rectangular LANDSAT data window containing the 10 counties included 4,887,960 pixels and required less than 80 seconds for classification of the ILLIAC IV. Column 3 of Table 2 indicates the gain or loss of information by using the remote sensing information in conjunction with the conventional area sample ground data. Column 3 in Table 1 shows the raw total of pixels classified and converted to acres by the factor 1.114, the acreage of one pixel. This type of direct estimate can lead to serious biases in the estimates for individual crops, with the extent of bias varying among crops. For an individual county within the analysis district the same type of estimation employing the regression parameters for the large geographic area can be used to the extent that the regional landscape may be considered homogeneous to the complete set of counties.

Table 1. Estimates of Agricultural Cover Types

Crop or Cover Type	Reported Acres July 27	Regression Estimate	Pixel Count x 1.114
	--- (000 acres) ---		
Corn	1286	1390	2105
Soybeans	631	701	610
Perm. pasture	533	434	678
Hay	179	154	104
Alfalfa	69	71	14

Table 2. Variances of Estimates of Agricultural Cover Types for 10-County Analysis District

Crop or Cover Type	Variance Reported (10^6 acres^2) (1)	Variance Regression Estimate (10^6 acres^2) (2)	Information Gain or Loss (1) ÷ (2) (3)
Corn	17202	2459	7.00
Soybeans	5880	847	6.94
Perm. pasture	4489	1096	4.09
Hay	630	376	1.67
Alfalfa	155	135	1.14

Yield Estimation. In order to derive production, the yield for the same area or group of counties is needed. This can be accomplished in any one of several ways. Here we describe one procedure based on SRS data during the growing season for the same set of segments as used for acreage. The correlation of satellite spectral reflectance information with appropriate ground survey plant yield data has been investigated to obtain statistical estimators with measurable standard errors for small areas contained in a single LANDSAT frame. The method is illustrated in terms of corn yield in Illinois during the 1975 crop year.

All sample fields of corn selected for SRS objective yield ground surveys were located on the LANDSAT images, and the digital values for the four spectral channels for all pixels within each field were used to compute the mean channel values for all fields. Mean vectors for the spectral data were obtained from the LANDSAT imagery for August 4, 1975, while the field data relate to a 10-day period centered on August 28, 1975 (for the September 1 forecast) and at harvest. Specifically, the mean vectors for 27 corn fields with 16 or more pixels were computed.

In addition to the individual field data, to predict actual yield for the analysis district the mean vector for the four LANDSAT channels would be needed for all pixels classified as corn; that is, the entire population of classified pixels for the area must be examined to identify the corn pixels to employ double sampling for yields. The information from the tape containing the classification results for the frame must be matched to corresponding spectral channel values for the same pixels on a second tape to derive information needed for the yield estimation procedure. The double sampling model(s) is the same as given earlier except the independent variables are now a vector of four channel values. However, the estimation of the yield is achieved through a double sampling regression estimator using the classified LANDSAT data collated with the ground data for the larger geographic area. In addition, it is possible to use the same double sampling regression estimator to obtain estimates for any smaller area for which the regression relationship is appropriate. One regression could be used to forecast final yield based on ear counts and size of ears on September 1. A second regression relates the actual grain harvested and weighed to the same four channel spectral values from LANDSAT from the August 4 image.

While a number of different variables or combinations of variables based on the field mean vectors and variance vectors were investigated using the August 1975 imagery in western Illinois, only two sets of variables gave statistical significance consistently: (1) means of channel 2 and channel 4, and (2) means of channel 2 and 4 plus variances of channel 2 and 4. The regressions based on data set (1) for September 1 and final harvest for the 10-county area within the LANDSAT frame of August 4 are as follows:

September 1: $\hat{Y}_s = \bar{y}_s + B_2(\bar{x}_2 - \bar{X}_2) + B_4(\bar{x}_4 - \bar{X}_4)$

$R = .56$

Harvest: $\hat{Y}_h = \bar{y}_h + B_2'(\bar{x}_2 - \bar{X}_2) + B_4'(\bar{x}_4 - \bar{X}_4)$

$R = .49$

where $\hat{Y}_s$ = yield per acre for the geographic area

$\hat{Y}_h$ = yield per acre for the geographic area

$\bar{y}_s$ = yield per acre for a sample of fields on September 1

$\bar{y}_h$ = yield per acre for the sample fields at harvest

$\bar{x}_2$ = mean spectral value for channel 2 on August 4 for sample fields

$\bar{x}_4$ = mean spectral value for channel 4 on August 4 for sample fields

$\bar{X}_2$ = mean spectral value for channel 2 on August 4 for the entire geographic area

$\bar{X}_4$ = mean spectral value for channel 4 on August 4 for the entire geographic area

B_2 and B_4 = regression coefficients (also B_2' and B_4')

R = multiple correlation coefficients

The gain in information by use of spectral data for yield estimation may be computed in a manner similar to column 3 of Table 2 based on the ratio of variances. For corn these information gains are in the range of 1.27 to 1.42. Based on these data sets for western Illinois in 1975, the potential information gain is obviously much less than that for acreage estimation. However, the use of the LANDSAT spectral data for both acreage and yield would result in an information gain of approximately 7.0 x 1.3 = 9.1 for estimation of corn production.

This gain of information in estimating production of a factor of almost 10 for the high quality image used would probably translate at the state level into a factor of 3 or 4. In general, correlations and information gains are reduced over larger areas due to the improbability of high-quality, late summer LANDSAT data for an entire state. However, for a small area, such as an individual county, with a high-quality frame, the information gain would probably average greater than 10. A gain in information by a factor of 4 would be roughly equivalent to $65,000 of additional resources which would be needed based on the current area sampling system to reduce the sampling errors of corn and soybean production in Illinois by one-half. The benefits of the reduced errors in Illinois has not been estimated but the marginal increase in costs of the LANDSAT analysis should be much less than the $65,000 providing real-time imagery is available and assuming that all initial fixed costs of establishing an operational information system have already been provided.

Summary and Conclusions

New technologies such as LANDSAT, more versatile computers, and digital communication systems now suggest more favorable information benefits per dollar spent for more comprehensive agricultural remote sensing information systems. Experiments to date with strategy of double sampling suggest that improvements in agricultural production estimation accuracy, timeliness, and geographic resolution will be worth the costs of more comprehensive systems development where a number of crops or land covers will benefit or small area estimates benefit state or local agencies. These substantial benefits to individual states for specific problem areas seem most probable where local data systems and information needs are integrated within improved agricultural information systems which would be implemented in conjunction with current USDA Federal-state programs. The set of economic factors which impact the development of such remote sensing information systems for the type of services discussed can be grouped as follows:

1. Personnel and service factors associated with the Federal agency for traditional programs.

2. Personnel and service factors associated with the state and local agencies for the traditional services.

3. Factors related to the operation of these new technologies.

4. Factors related to the use of more complicated procedures for combining data sources to improve further the accuracy of land cover estimates.

5. Costs and benefits associated with providing new services which may be expected to result from the extensive coverage of LANDSAT.

6. Factors associated with losses incurred because LANDSAT fails to provide coverage due to cloud-cover problems.

Given that such advanced remote sensing information systems are integrated into specific USDA Federal-state programs, a variety of spinoff benefits to other agricultural and nonagricultural government agencies may be anticipated. For example, current crop cover in conjunction with digitized topography available from the Defense Mapping Agency should indicate areas where soil erosion management policies should be concentrated. Such systems should allow forestry resource monitoring with only minor upgrading. They should also allow an inventory of irrigated croplands and hence assist planning of agricultural water supply demands. Through geographically detailed estimation of crop acreages such systems should assist in determining the patterns of use of pesticides, herbicides, and fertilizers. Integrated with land cover and topographic data, they should yield geographic information vital to water quality management of streams and rivers. Additionally, such systems should enable dynamic monitoring of urban development patterns, the encroachment of development onto prime agricultural lands, and shifting demographic distributions.

These advanced domestic agricultural production monitoring systems should be very close to those systems that are needed throughout the world in less-developed countries for more rational planning of land uses and agricultural activities. Thus their domestic implementation may well suggest the development of lateral programs at the international level that would function in a manner similar to the flexible U.S. Federal-state programs.

References

1 Huddleston, Harold F., and Ray, Robert M. III, "A New Approach to Small Area Crop-Acreage Estimation," Annual Meeting of the American Agricultural Economics Association, Pennsylvania State University, State College, Pennsylvania, August 1976.

2 Ray, Robert M. III, and Huddleston, Harold F., "Illinois Crop-Acreage Estimation Experiment," Proceedings of the 1976 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.