

83-17

# Scope and Methods of the Statistical Reporting Service

Prepared by the Statistical Reporting Service,  
U.S. Department of Agriculture

Miscellaneous Publication No. 1308

Washington, D.C. 20250

Revised September 1983

# Research

## Introduction

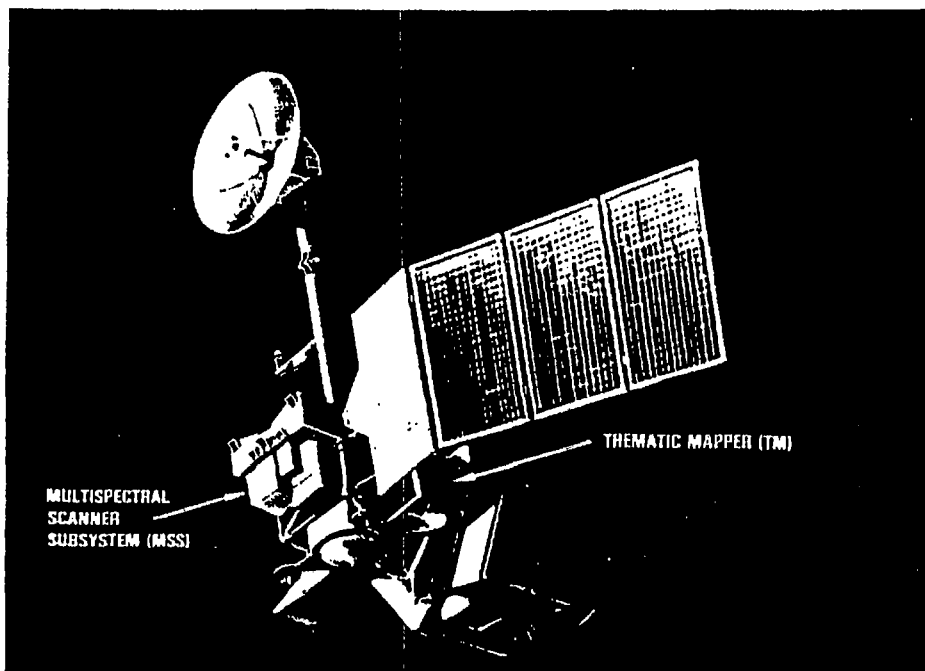
Research has been a formal SRS activity for more than 25 years. SRS continually conducts research aimed at improving the quality of crop and livestock estimates. This chapter describes the principal areas under study: remote sensing, crop-yield forecasting and estimation, sampling frames, and survey research.

## Remote Sensing Research

Remote sensing research is the study of imagery or digital data collected from aircraft or spaceborne sensors to detect valuable information about conditions and resources on the earth's surface. SRS has been a user of remote sensing products since the fifties when it began using aerial photography to construct area sampling frames. Thus, since the fifties, SRS has had an intense interest in being familiar with and using the most cost-effective photographic products available for its purposes. As described in Chapter 2, SRS presently uses aerial photography mosaics to stratify broad land use

and rectified aerial photographs to collect data from farm operators.

Numerous research efforts were conducted in the sixties and early seventies by SRS and other research institutions such as the National Aeronautics and Space Administration (NASA), Purdue University's Laboratory for the Application of Remote Sensing, the Environmental Research Institute of Michigan, South Dakota State University's Remote Sensing Institute and the University of California at Berkeley's Remote Sensing Research Program. These projects investigated the potential of aerial photography and aerial multispectral scanner data (digital energy readings) to meet SRS information needs. Topics addressed were diverse and included tree counts, fruit counts, livestock counts, and measuring the effects of corn blight. This previous research by SRS and other research institutions and the use of its area sampling frame procedures put the agency in a position to address the research needed for domestic crop estimation after the launch of NASA's first earth-resource monitoring satellite, Landsat 1, in 1972.



*Artist's rendition of Landsat 4 satellite in space, showing the multispectral scanner and thematic mapper. These devices sweep the earth's surface, measuring reflected energy. Scientists are learning to use satellite data to better estimate crop acreage.*

## Landsat Satellites

Four Landsat earth-resource monitoring satellites, launched since 1972, have collected digital information on the amount of energy being reflected, emitted, or absorbed by the earth's surface. The major measuring instrument on Landsats 1, 2, and 3 was a multispectral scanner (MSS) that measured energy in four bands of the electromagnetic spectrum every 18 days for approximately each acre on the earth's surface. Landsat 4's new measuring instrument, the thematic mapper (TM), takes measurements in 7 bands every 16 days for approximately each one-fourth of an acre. Landsat 4 also has an MSS on board that provides 16-day coverage.

Since the launch of Landsat 1 in 1972, SRS has investigated the potential contribution of earth-resource satellite data to its domestic crop acreage estimation program. SRS's operational use of area sampling frames already provides a statistically sound and cost-effective crop acreage 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 10 years, SRS has developed two major uses of Landsat data. The first is photo-interpretation of imagery from Landsat MSS and the return beam vidicon (a second sensor on board Landsats 1, 2, and 3) for broad land use stratification in constructing area sampling frames, as explained later in this chapter. The second major use of Landsat MSS data is to classify the digital data into crop types and regress SRS ground-collected data against the classified Landsat data for each crop type. The degree of success of these improved crop acreage estimates depends heavily on optimum timing of Landsat coverage, the extent of cloud cover, and the rapid delivery of Landsat data to SRS.

Methodology used by SRS in its domestic crop estimation program is described in Chapter 2. One major survey used for crop area estimation is the June enumerative survey (JES). The use of Landsat data for crop area estimation has been an extension of the JES. The JES provides data collected for a random sample of area segments. Landsat measures energy for each acre on the earth's surface. A statistical relationship, called a regression estimator, is developed between the ground-gathered data and the satellite energy readings.<sup>2</sup> The measure of success associated with this probability-based estimator is referred to as the relative efficiency (RE).

<sup>2</sup> See John Cochran, *Sampling Techniques*, third edition, John Wiley and Sons, 1977, pp. 189-199.

RE, or success, is defined as the multiplier that would be necessary to increase the sample size of the JES in order to obtain an estimate just as precise from the ground data alone as from the ground data plus Landsat. 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. Major research studies were conducted from 1972 to 1978 to determine the applicability of the regression estimator for large-area crop estimates. RE's generally ranged from 1.5 to 5.0. Based on this research, researchers concluded that it was indeed feasible to prepare large-area crop estimates through joint use of JES and Landsat data. Since 1978, in selected areas, the research results have been calculated on a timely basis and used as an indicator for the CRB's annual crop summary.

Specialized software and computer hardware were necessary to process the Landsat and JES data jointly. Starting in 1974, SRS and the University of Illinois' Center for Advanced Computation entered into a joint agreement to redesign a software system for this purpose. The system (EDITOR) was redesigned to analyze SRS's ground-gathered data and Landsat data and to calculate the regression estimates. Data processing was to be done at several locations using several computer systems and telecommunications networks. The Illiac IV supercomputer at NASA's Ames Research Center was used by SRS from 1974 to 1981 to classify the 10.5 million picture elements ("pixels") on a Landsat image into the selected crop and land cover types. Currently, SRS uses the CRAY-1S supercomputer for this task at the Ames facility. In 1983, SRS plans to process a volume of Landsat data equivalent to the amount of data collected for 75 national JES's.

## AgRISTARS Program

Starting in 1979, SRS research became a focal point of the Agriculture and Resource Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) research program. AgRISTARS is a 7-year interdepartmental research program designed to evaluate the potential of aerospace remote sensing data to meet many USDA information needs. The program was initiated by USDA and includes as partners NASA, the U.S. Department of Commerce, and the U.S. Department of Interior. Multi-agency teams of scientists were assigned to eight major project areas: early warning and crop condition assessment, foreign commodity production forecasting, yield model development, domestic crops

and land cover, renewable resources inventory, soil moisture, conservation and pollution, and supporting research.

SRS has overall management and coordination responsibility for the AgRISTARS Interagency Coordinating Committee and the Program Management Team. Within the structure of the interagency AgRISTARS research program, SRS assumes leadership for the Domestic Crops and Land Cover project (DCLC). SRS also has a supporting role in several of the seven other project areas.

The goals of the DCLC project are: to calculate timely and more precise crop area estimates at the State and substate level, using Landsat and SRS's JES data jointly in an expanding number of States; to conduct research on new methodology and on data from new satellite sensors and on ADP to process and analyze the Landsat and JES data; and to conduct research on land cover area estimation and mapping.

In early 1983, AgRISTARS was restructured and reduced from eight to three major projects: domestic commodity assessment research, foreign commodity assessment research, and land resources monitoring research. SRS maintained leadership of the domestic commodity assessment research and overall AgRISTARS management. The work done in the DCLC project continues under the domestic commodity assessment research project.

### Land Cover Research

In 1979, land cover was a completely new area of interest for SRS remote sensing research. The concentration before had always been on estimates of crop areas only. The DCLC project gave SRS a research charter to evaluate and develop techniques which might answer information needs of other USDA agencies or State and county-level agencies concerned with natural resource management. NASA's National Space Technology Laboratory (NSTL) had prior experience in examining land cover and geographic information needs of other USDA agencies as well as State and local agencies. Thus, SRS and NSTL personnel began joint remote sensing research efforts in 1980 to address land cover information needs.

The first technique that SRS chose to explore for statistically based land cover estimation was an extension of the JES, and JES-plus-Landsat, estimators. Since the JES sample was allocated mainly for crop and livestock

estimation, some additional nonagricultural strata segments were added to the JES sample in Kansas to gather information on the basic land cover definitions.

The SRS requirements for land cover research are that any estimates created must be statistically sound estimates and any map products created must have statistically sound measures of accuracy.

### Remote Assessment of Worldwide Crop Yields and Conditions

SRS is taking part, with the Agricultural Research Service (ARS) and Foreign Agricultural Service (FAS), USDA, and the National Oceanographic and Atmospheric Administration (NOAA) in the AgRISTARS program under the domestic commodity assessment research project in the development of remote assessment methods to measure crop yields and monitor crop conditions in the United States and other countries. The methods being researched rely on inputs from many sources including weather data as reported through existing domestic and foreign sources, published material on cropping practices, satellite remote sensing data, analysis of the effects of changes in government programs and evaluations of the extent of disease and pest problems. Mathematical models bring together the effects of combinations of these inputs to provide condition or yield values which can be interpreted alone or in a time series with previous data.

Research has been under way since the outset of the AgRISTARS project, and several methods developed are being used operationally by FAS. The primary areas of application, outside the United States, are for the major countries that produce or import crops produced by the United States. Two major crops of interest are wheat and soybeans or their substitute crops. Countries of interest are the U.S.S.R., People's Republic of China, Brazil, and Argentina.

In the United States, data are collected, analyzed, and published on a current basis in direct support of yield and condition assessment. This is not the case for many foreign areas. Remote assessment is thus complimentary to other domestic data collection sources and is a primary method for independent assessment over foreign areas. In the United States, such methods could prove useful to obtain yield information for States not included in the objective yield programs and for crops not having objective yield surveys.

Even though there are many similarities between condition assessment and yield estimation there are some differences. Condition assessment relates measures of the current year with similar measures in previous years, or previous points during a given growing season, whereas yield estimation would produce a point value that would "stand alone" without previous data.

Developing crop yield models relating factors that seem to affect production, such as weather events, is not a new endeavor. SRS and its predecessor agencies have been involved in the development and use of weather-yield regression models for some time. What is new about the current effort to develop and use these models is their application to forecasting crop yields in foreign production areas and their reliance on remote sensing. Large-scale use of such models is now possible using advanced computers and improved statistical techniques.

Researchers have developed a set of criteria to evaluate proposed crop yield models for use by USDA agencies. Evaluation of candidate models, in terms of their ability to "measure up" against these criteria or desirable characteristics, also helps model developers to improve the models and to identify those models which are the best candidates for further research. The eight criteria currently emphasized in model evaluation and comparison research are: yield indication reliability, timeliness of model output, minimum cost of model operation and development, the provision of accurate current measures of modeled yield reliability, objectivity, consistency with scientific knowledge, adequacy of geographic coverage and detail, and simplicity.

Crop yield model development must assess both the short- and long-term factors which influence yield levels. The plant process model, discussed elsewhere in this chapter, attempts to explicitly account for important factors of both types. However, to the extent this is not fully achieved, either because of model limitations or limited data availability in many foreign areas, the plant process model may need to be calibrated over a period of years or by location to the observed or reported yields. Somewhat simpler models, which are designed to reflect more general relationships of a crop's yield to the plant's environment, are being developed from a time series of yield and environmental data.

In condition assessment, SRS participates with scientists from ARS, NOAA, and NASA to develop methods with particular reference to early warnings of events that may substantially affect crop yield. This research is aimed at providing USDA agencies having responsibility

for foreign and domestic production forecasting, with a timely capability to monitor crop development and those conditions affecting production potential. Some of the products developed may also have application in the SRS input to the weekly weather and crop program. Among the techniques being studied are models to provide alarms for potentially damaging ambient conditions, including water and temperature stresses, the prediction of insect epidemic, and plant stress caused by flooding.

### Look to the Future

In the eighties, the researcher may be able to make use of aerospace data from a variety of sources. At the same time, costs may be greater and some basic limitations of satellite data remain to be hurdled. As new sources of aerial photography or satellite imagery become available, SRS personnel will evaluate their potential for cost-effective improvements in constructing area sampling frames. Potential sources are as follows: the Government's national high-altitude photography (NHAP) program, MSS and TM image products from Landsat 4 and D' (which will be identical to Landsat 4), the large-format camera on future space shuttle missions, and the French "SPOT" satellite image products. The major problems to be resolved are the cost and geometric rectification of these products.

The second major use of Landsat data by SRS is the use of digital data for crop and land cover area estimation and crop yield assessment. Potential future sources for this digital application are: both MSS and TM data from the Landsat 4 and D', French SPOT data, and any future private-sector Landsat MSS, TM, or new sensor missions. Research results from TM simulator data over Missouri in September 1979 indicated that the increased TM resolution will improve the efficiency by a multiplier of two to three over MSS for major crop area estimates.

Major concerns of SRS regarding future applications that use spaceborne data are as follows: frequency of satellite data coverage during short (1-month) "optimum windows," the cost of "one scene" of satellite data, rapid delivery of satellite data to SRS, and the total project cost. Using 1982 cost estimates the RE of the regression estimator should be 2.5 or larger for a cost-effective improvement in statistical precision. The SRS research results from 1975 to 1982 have been mixed concerning cost-effectiveness. The majority of results have met the criterion of the cost-effective improvements in precision. However, the increased cost of the new and

higher resolution data and the associated data processing costs will be a major concern. The frequency of cloud-free images (coverage) is a serious operational problem at this time. The need for cloud-free coverage for at least one pass during the relatively short (approximately 1-month) optimum temporal windows for crop-area estimation is a critical need of SRS. A 9-day interval is the preferred minimum satellite configuration to ensure this coverage for any operational SRS domestic crop area estimation program that uses satellite and JES data. At present, the Landsat program provides 16-day coverage. Also, the 21-day target average for data delivery needed by SRS for Landsat data and images has never been met for any SRS full-State project. Twenty-one day turnaround for products is essential for future SRS projects. However, NOAA plans to provide Landsat MSS data to users in 12 to 27 days in the near future.

Completing research and development required to extract agricultural statistics information from MSS data took several years. It is expected that the same will also be true for the new higher resolution sensors. However, the research and development time period is expected to be somewhat shorter due to the presence of experienced statistical, data processing, and remote sensing staff. As with the MSS, however, the potential information content is significant, and the eighties will be a decade of exciting research on these new sensors.

### **Crop Yield Forecasting and Estimation**

The Research Division develops and improves objective methods of estimating and forecasting yields for a wide variety of crops. The estimation of crop yields at harvest and forecasting of yields during the growing season are two distinct phases of this research. For most crops, the development of methods of estimating harvesting losses constitutes an additional phase.

### **Objective Yield Surveys**

SRS has a continuing program of research to improve operational objective yield forecasts and estimates and to reduce the cost of data collection. In addition, research is underway to develop new objective yield surveys for rice, grain sorghum, and sunflower. As technology and cropping practices change, SRS researchers conduct field studies on objective-yield crops to evaluate the size and number of plots needed for sample fields, the relationships in the forecasting models, and procedures used for obtaining counts and measurements.

They also observe the plant growth characteristics of new crop varieties.

Another form of research, the validation study, tests for overall levels of bias in objective yield procedures. Researchers compare the yield, calculated from a large number of plots harvested in a field, to the actual weighed production the farmer harvests.

Development of objective yield procedures for new crops begins with a trial adaptation of existing procedures for estimating yields. Literature on crop development characteristics suggests which plant characteristics would be most useful to count and observe. Procedures are modified based on pilot studies in one or more States, and historical data are accumulated to construct forecasting models. Developing reliable estimating and forecasting procedures generally takes 3 or more years.

### **Plant Process Models**

A plant process model (PPM) is a computer simulation of plant growth. Most models simulate the day-by-day growth of a plant by calculating the effects of management variables (such as plant density and cultivar), soil properties and initial soil water, and real or hypothetical daily weather conditions. The models ultimately estimate gross grain yield of a typical plant in a homogeneous plant community.

SRS is researching the feasibility of ultimately using PPM's to produce more timely and accurate yield forecasts based on growing conditions during the current crop year. This presupposes that a model could accurately mimic (or model) plant growth with minimal data requirements.

Many computer simulation models of plant growth developed by agricultural scientists have been used to manage water resources, diseases, pests, and other factors affecting crop yield. Yet no PPM has been extensively validated for a wide geographic range of growing conditions. Needed techniques are still being developed and refined, and data required for an extensive validation would be substantial.

SRS objective yield forecasts assume that historic regression relationships—usually for the most recent 3 years—will continue for the current year. However, if environmental conditions for a given crop year are unlike those during the previous years (or if they are but the regression fit is not particularly good), the forecast may be somewhat inaccurate. Since PPM forecasts are