REMOTE SENSING PROGRAM OF

THE NATIONAL AGRICULTURAL STATISTICS SERVICE: FROM A MANAGEMENT PERSPECTIVE*

BY

GEORGE A. HANUSCHAK
CHIEF, SURVEY TECHNOLOGY BRANCH
NATIONAL AGRICULTURAL STATISTICS SERVICE
U.S. DEPARTMENT OF AGRICULTURE

1992

I. SUMMARY

The National Agricultural Statistics Service of the United States Department of Agriculture has been utilizing digital earth resource observation satellite data since the launch of Landsat 1 in 1972. There are three major applications in the U.S. agricultural statistics program. These are area sampling frame construction, crop area estimation and crop condition assessment. In a related application, the Agency plans to use a geographic information system for farm chemical data. These four applications are in various stages of development and operational implementation.

The area sampling frame construction process is the largest operational application. Area sampling frames remain as the statistical foundation of many U.S. probability based agricultural surveys since area frames have complete frame coverage and no duplication of land areas. NASS has been utilizing area frame sampling since 1954 and nationwide since 1965. The land area of the country (state by state in the U.S.) is divided into broad land use or land cover strata. From 1954 to 1978 aerial photography was the primary source of information for this stratification. From 1978 to 1992, Landsat image products have been the main source of stratification information. Starting in 1988, NASS started a joint research effort with the National Aeronautics and Space Administration (NASA) to convert the stratification process from a basically manual image interpretation process to an almost entirely digital process by combining Landsat thematic mapper digital data with digital line graph map data from the U.S. Geological Survey. Starting in 1992 this process is operational and the area sampling frame for the state of Oklahoma to be used for the 1993 major mid-year Agricultural Survey was developed using the new process. In addition the state of California is also being started in 1992. French SPOT data is also being investigated as a source for delineation of city and urban boundaries. Benefits of the new method include both improved cost efficiency and timeliness.

The second application is the use of Landsat thematic mapper data in combination with area sample frame based ground-gathered data to improve the precision of rice and cotton acreage estimates in the Mississippi Delta region of the U.S.. These estimates are calculated in an operational timeframe and provided to the Agency's Agricultural Statistics Board as input to the official estimates released by the Agency during the crop season. The well documented regression estimator approach is used. The Delta region was selected because of the excellent separation characteristics of rice and cotton from competing spectral land covers and because of the North-South orientation and relatively small growing region compared to the Midwest or

* Delivered at the European Community's Eurostat and Agriculture Conference on "The Impact of Remote Sensing on the European Statistical Information System," Bad Neuenahr, Germany, Sept. 1992.

Great Plains regions of the U.S.. Multi-temporal Landsat data is used and regional, state and county level estimates are calculated. In addition county level classification color coded theme map products are provided to the state offices. This project began in 1991 and will be done on an annual basis. The Agency has a long history of similar projects with Landsat Multi-Spectral Scanner Data from 1972-1990. Accurate cost estimates have been kept for the time series 1972-1992 for these projects for cost benefit analysis comparing the new method to the conventional area frame ground-gathered data approach. The statistical measure of performance used is the relative efficiency which is the ratio of the variance of the ground data only direct expansion estimator (numerator) and the variance of the regression estimator (denominator). Larger values of the relative efficiency reflect a larger gain due to adding Landsat data into the estimator process. For the 1991 Delta project, the relative efficiency for rice was 3.90 and for cotton it was 3.25. That is, the sample size on the ground would have to be increased by a factor of 3.25 to 3.9 to match the precision of the Landsat-based acreage estimate. These were cost effective improvements in the precision and accuracy of the State level crop acreage estimates with no additional respondent burden on farm operators. In addition, sub-state estimates and classification maps were provided.

The third area of utilization of remotely sensed data involves the use of vegetative indices calculated from National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer (AVHRR) sensor. NASS has been slow to get into this area because of its very extensive ground-gathered objective yield forecasting and estimation program already provided excellent information on crop conditions. However, due to the daily satellite passes and the spatial nature of the AVHRR data there is now interest in calculating and mapping vegetative indices similar to the operational program that Statistic's Canada has had since 1988. NASS is currently populating a historic data base of AVHRR data and testing the hardware system to support this type of activity. NASS is using the Land Analysis System (LAS) software from the Earth Resource Observation Satellite (EROS) data center in Sioux Falls, South Dakota and NASA's Goddard Space Flight Center. This program is still in the development stage but the goal is an operational program that would sell on a subscription basis special crop condition assessment data and color map products similar to the Statistics Canada program.

The fourth and newest area is the use of geographical information systems for providing management additional information about agricultural issues by taking advantage of the spatial aspects of the data and by overlaying several layers of data such as farm chemical applications, soil types, slope and water flow, crop and land use covers, etc. NASS is in the very early stages of the utilization of GIS based data and related analysis. NASS has procured the ARCINFO GIS software system and also has a Sybase relational data base software system that integrates with ARCINFO. NASS is in the process of populating a farm chemical data base and GIS layer at the moment. In addition, a small pilot project looking at Global Positioning System recorders is being done. After completion of these tasks, other layers will be considered as analysis goals and potential become better clarified.

Overall NASS is a fairly extensive user of space based remotely sensed data and related spinoff technologies such as GIS in its' U.S. Agricultural Statistics Program. However, in relation to the overall NASS mission of providing agricultural statistics on hundreds of items throughout a year, the portion of NASS's program that utilizes remotely sensed data is not large. Especially if one compares it to the sometimes seemingly "utopian and panacea type" claims of many in the Remote Sensing Community in the 1970's then it seems like a fairly modest (albeit an important one) addition to the U.S. Agricultural Statistics Program. The Agency has, however, been able to quite successfully integrate and supplement its' existing probability based (area, list and multiple frame sampling) estimation program by utilizing digital and image space based remotely sensed data for the four applications described above.

II. AREA SAMPLING FRAME CONSTRUCTION

The most recent thorough description of the conventional paper based area sampling frame construction process was given by Cotter and Nealon (Aug. 1987 - reprinted June 1989). Earlier thorough descriptions of the paper based process were given by Houseman (November 1975) and Huddleston (April 1976). The first description of the use of Landsat images in the conventional paper based system was given by Hanuschak and Morrisey (October 1977).

The most current description of the new system called the Computer Assisted Stratification and Sampling (CASS) system was given by Cotter and Mazur (Feb. 1992). An earlier description of the computer assisted CASS system was given by the National Aeronautics and Space Administration personal and contractors Cheng, Angelici, Slye and Ma (November 1989). These previous papers document both old and new procedures in substantial detail which won't be duplicated in this report.

In brief, the CASS procedure is primarily electronic and digital based (both maps and Landsat TM data) using color graphics displays and the conventional procedure was primarily paper based. Assuming the accuracy of the image interpreted stratification of both methods to be approximately equivalent, there is a substantial productivity gain with the new method and a slightly smaller total cost. The new method is now operational and is being used for the area sampling frame construction for the relatively large land area States of Oklahoma and California.

III. CROP AREA ESTIMATION IN THE MISSISSIPPI DELTA REGION OF THE UNITED STATES

NASS staff used Landsat Thematic Mapper Data to operationally calculate improved crop acreage estimates for rice and cotton in the Mississippi River Delta Region in 1991 and 1992. The Landsat Thematic Mapper used in conjunction with area frame based ground-gathered data in the form of a regression estimator. The ratio of the variances, also called the relative efficiency, of the regression estimator and the ground data only direct expansion estimator is the measure of statistical gain from using Landsat Thematic Mapper Data.

In 1991, for rice the relative efficiency averaged 3.90, for cotton it was 3.25 and for soybeans it was 2.45. Results from 1992 are not available yet but increased cloud cover may well decrease the statistical gains compared to 1991. The relative efficiency can also be interpreted as the factor by which the ground data area frame sample size would have to be increased by to match the results of the regression estimator. Due to cloud cover and scene availability factors, the Landsat coverage area was divided into both multitemporal and unitemporal analysis regions. In addition, county level estimate were also calculated. Coefficients of variation for the county level estimates for the major rice counties ranged from 3.9 to 10.0 percent. Also, color coded crop classification maps were provided to the State Statistical Offices. The full details of this project are in a recent paper by Bellow and Graham (Aug. 1992). All the estimates were calculated using the extensive PEDITOR in-house software system developed by NASS and the

National Aeronautics and Space Administration Ames Research Center Staff over the years. The PEDITOR system's current status has recently been summarized in a paper by Ozga, Mason and Craig (Aug. 1992).

Accurate cost data has been collected and preserved in a data base by project managers since the mid-late 1970's. Thus, NASS has been able to look at Landsat projects (both Multi-Spectral Scanner and Thematic Mapper) from a rudimentary cost/benefit perspective over the years (1975-1992). The cost side of the equation has been relatively easy to measure. However, as in any cost/benefit analysis the assumptions made about the benefits are a key ingredient to the validity of the analysis. Current total Landsat project costs per State are approximately $175,000. Of the total, 63% is for salaries and benefits, 14% is Landsat data purchases, 12% is all data processing costs including amortized equipment costs on an annual basis, and 11% is a second visit to ground data sites where fields were not already planted on the first visit (See Figure 1.) In addition, costs per State for the already operational ground survey are approximately $60,000 for the States involved in the project. Landsat project costs have been dropping due mainly to advances in computer technology and in concert with dropping prices for any given level of technology. Ground data collection costs on the other hand are increasing due to inflation in salary, hotel and mileage costs for survey interviewers.

When total project costs are compared over time and divided by billions of bytes of input Landsat data processed, the project cost drop is dramatic (see Figure 2.) This was due to two main factors. The first has already been cited as the dropping prices of an ever improving computer technology. The second is staff productivity as more States and land areas are done with a constant number of research staff. With the Landsat Thematic Mapper sensor, it is estimated that a relative efficiency in the 3.0-4.0 range is required to be cost effective. Thus, the 1991 results for rice and cotton are judged to be cost effective improvements. This is especially the case since the fairly dramatic improvements in the precision and accuracy of the crop acreage estimates do not add to total respondent burden which is a major concern in U.S. agricultural surveys. However, the success across years and seasons is still dependent on the degree of cloud cover during the critical crop discrimination windows which are usually only 30 - 40 day windows at best. The probability of success for these projects would be increased substantially by having eight day coverage (two Landsat TM systems) instead of the planned one at a time Landsat 6 and 7 systems.

IV. VEGETATIVE INDICES

NASS has recently (last 18 months) begun to explore the possibilities of crop condition assessment utilizing vegetative indices from NOAA's AVHRR data. The Agency has been slow to get into this area because of its very extensive conventional ground-gathered survey data program to forecast and estimate crop yields. The conventional program utilizes both objective crop counts and measurements such as corn ears, ear length and circumference, field and laboratory weights etc. for each crop plus farmer reported yields. Both types of data have long well established time series and provide a relatively high performing system for forecasting and estimating crop yields. The most comprehensive document of the U.S. system for forecasting and estimating yields was by Huddleston (August 1978). For a current update, the Agency survey manuals would be the best source.

However, due to the daily satellite passes and the spatial nature of AVHRR, NASS research staff

FIGURE 1: 1991 Delta Project Costs

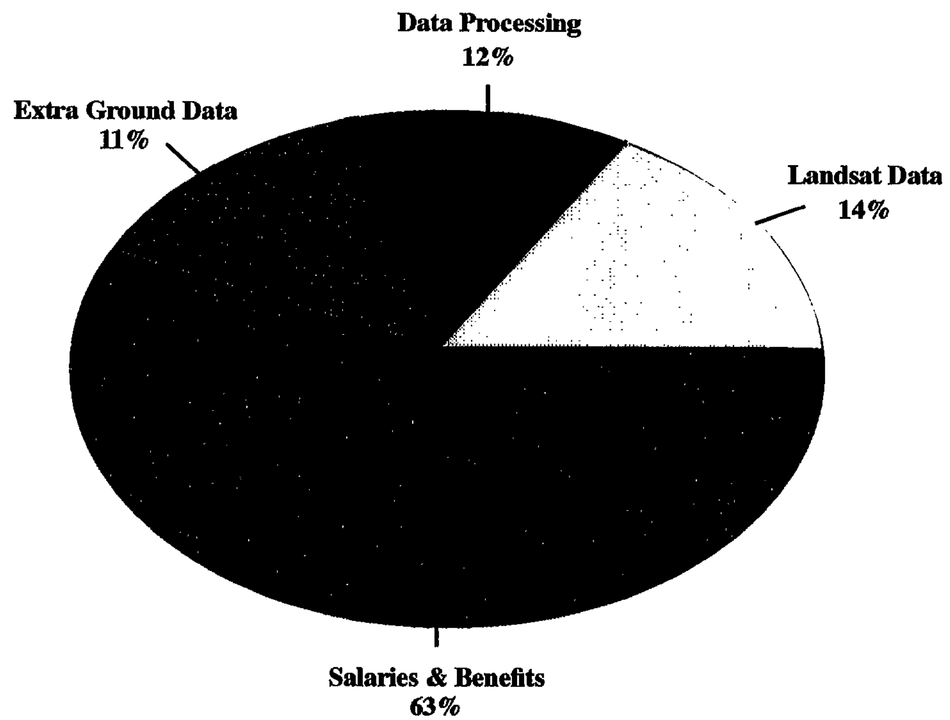
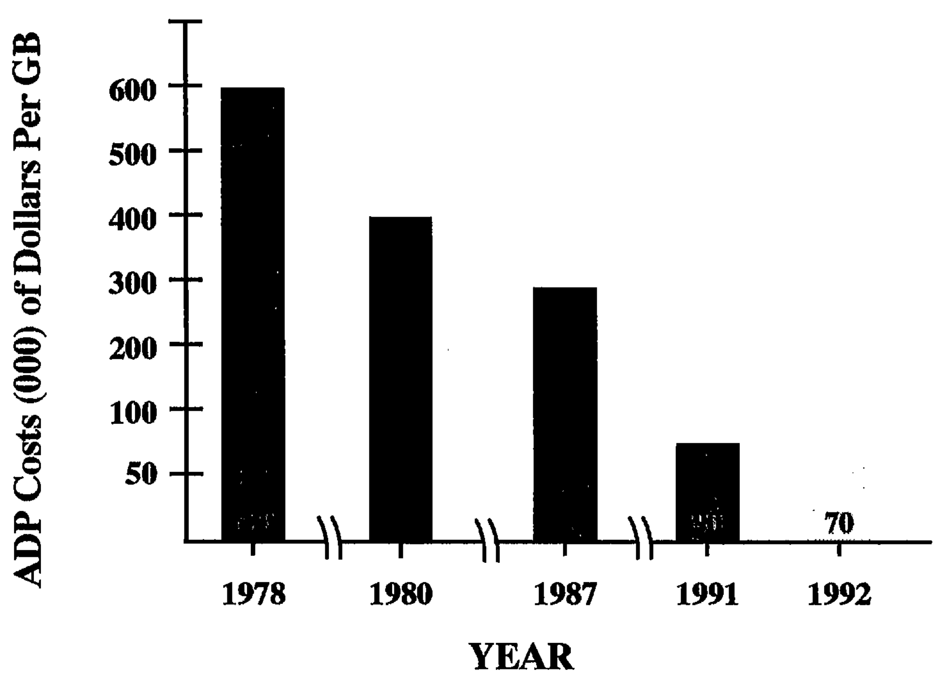


FIGURE 2: ADP Costs Per Gigabyte (GB) of Input Data
(1978 - 1992)



saw some new potential. In addition, close cooperation with Statistics Canada's Agriculture Division enabled NASS to observe their AVHRR vegetative index program which became operational in 1988. These facts combined inspired NASS research staff to initiate a program. NASS has begun a cooperative agreement with the Remote Sensing Laboratory of USDA's Agricultural Research Service to investigate vegetative indices as related to crop conditions. Condition assessment encompasses such topics as comparison of current year crop(s) growth to previous year(s), comparison of crop growth within a given year between States or counties, and drought monitoring. The AVHRR-based Normalized Difference Vegetative Index (NDVI) produced biweekly by the EROS data center will be specifically evaluated. Early research in crop condition assessment will center on the evaluation of NDVI color line printer plots, building a historic data base of NDVI and on the use of the NDVI for spring wheat yield models. A DEC VAXStation workstation has been purchased for this project; it will utilize a current version of the U.S. Geological Survey Land Analysis System LAS software.

A very recent event, Hurricane Andrew, that occurred in the U.S. in southern Florida and southern Louisiana will be looked at by comparing vegetative indices before and after. The southern Louisiana area will be of special interest since the same area is also already part of the Landsat TM Delta project previously mentioned.

V. ENVIRONMENTAL DATA AND GEOGRAPHIC INFORMATION SYSTEMS

The newest major addition to the Agency's survey program are farm chemical application data in various forms. Survey programs have been designed and implemented (1989 - current) to measure farm chemical applications at the farm level and at the individual field level on a sample survey basis. As part of the U.S. President's Water Quality and Food Safety Initiatives, NASS has become the surveyor of farm applied chemicals. As part of these initiatives, the tasks of putting these data in a data base and into a geographic information system were also assigned to NASS. NASS has utilized SYBASE (a UNIX based relational data base system) and ARCINFO (a GIS system) as the software to provide the necessary platforms for storing, retrieving and analyzing the sample survey farm chemical data in relation to water quality when other data layers are added such as soils, land use, vegetative indices, topography, etc. Data at the published level and micro data will soon be entered into these systems. Confidentiality of farmer reported data will be strictly protected as only use for statistical purposes will be allowed and individual data will not be revealed in any form of publication. In addition, a small pilot project was initiated to look at Global Positioning Systems (GPS) recorders for getting accurate coordinates of field locations. A recorder was used to label points within several sample segments in Ohio. This technology, as reported by many other applications scientists, seems to meet most accuracy needs. However, the up front capital investment in equipment, software, training, etc. was judged to be too high for current Agency applications. However, as costs continue to drop, the GPS technology holds substantial promise for several Agency applications such as GIS, area frames, etc.

VI. IN HOUSE COMPUTER SYSTEMS

To service these requirements, a wide range of microcomputer technologies are interfaced in-house. Large volume remote sensing analyses are performed on a VAXCLUSTER of a MicroVAX 3500 and a VAXStation 3100. Other technology research applications, such as GIS, are performed on a UNIX system which utilizes a SUN 4/380 server with SPARC and SUN IPC workstations (both stand alone and client server forms). Both servers have a 9-track tape and Exabyte tape cartridge capabilities in addition to several disk drives and other peripherals.

Smaller volume analyses utilize 386 and 486 personal computers as stand alone and/or client workstations to both the SUN and VAX servers. All servers, workstations and personal computers are connected together on an ETHERNET network using Network File Server, DECNET and TCP/IP protocols. Peripheral equipment includes high resolution color monitors, printers, scanners, video cameras, and digitization tablets. Other equipment includes laptop and notebook computers, such as GRID Pads and Zenith Supersports and Zeos.

VII. LOOK TO THE FUTURE

The future of all four of the efforts described in this report (area sampling frame construction, crop acreage estimation using a regression estimator, vegetative indices, and geographic information systems) is bright concerning the technology aspects. The pressure will be on economic factors and showing cost effective improvements or new products in the budget decision time schedules and framework.

As far as the technological aspects, the U.S. Government has recently increased its commitment to future Landsat's 6 and 7. The U.S. Government is firmly supportive of the NOAA/AVHRR program. NASS is firmly supportive of area frame sampling as its statistical foundation to complete universe coverage without duplication in the frame. NASS of course, complements this with list and multiple frame sampling as well. NASS staff are also investigating panel surveys calibrated to the universe as a potential path to reducing total respondent burden. Geographic Information Systems are proliferating throughout the public and private sectors on a worldwide basis.

The one down side on sensors is that for forecasting and estimating a dynamic event like crop production frequent satellite coverage is required. One Landsat TM or enhanced TM at a time, only gives 16 day coverage. For acreage estimation with the regression estimator, optimum classification windows are often only 30-45 days in length. Usually, that gives only 2 or 3 chances to get data during the optimum window. If those 2 or 3 chances are substantially cloud-covered, then the statistical gains of the regression estimator can drop dramatically. NOAA/AVHRR gives daily coverage but with much different resolution than Landsat TM or SPOT. Thus, it enables large scale looks at the vegetative indices across time but doesn't provide a vehicle for estimating acreage accurately compared to ground-gathered data systems. Perhaps some private sector systems could be developed to better meet agriculture's needs.

The challenge will be to speed up the R&D process as much as possible to evaluate if cost beneficial application of these various technologies is appropriate under most likely declining budgets. Substantial progress has been made but work remains. The U.S. and other government commitment to space borne sensors seems to be at a quite healthy stage. The next 5 - 10 years will be crucial to complete R&D, and to apply the technology where it makes sense in a cost effective manner.

In addition, new sensors such as several nation's radar based systems and NASA's Earth Observing System Data and Information System (EOSDIS) will be new systems of data to evaluate. It is difficult to envision that preciously few research resources in NASS can address new sensors as well as current sensors. NASS staff will observe other research such as European and Canadian research on radar systems for agriculture and land cover and NASA research on EOSDIS. Radar sensors overcome the cloud problem but also have different characteristics and require different processing methods. If substantial demonstration of potential cost effective improvements are completed, then NASS research staff would re-evaluate its resource allocation. However, given current resource availability and NASS applications, we will continue to focus on Landsat TM, NOAA/AVHRR, area frame sampling (general and specialty crops), and geographic information systems especially related to environmental data such as farm chemical data. In fact, it will be a serious challenge to even address these four applications appropriately under cost and staff constraints.

VIII. ACKNOWLEDGEMENTS

The author sincerely acknowledges the contributions of many persons associated with this effort. Especially those units and staff listed here:

Remote Sensing Section Staff
Technology Research Section Staff
NASA Ames Research Center Staff For This Project
Area Sampling Frame Section Staff
Various Functional Unit Staff Members Throughout the Agency
State Statistical Offices
Agricultural Research Service Staff For This Project
USDA Remote Sensing Coordinator
Environmental Statistics Section Staff
Agency's Program Planning Committee
Charles E. Caudill, Administrator of NASS

IX. REFERENCES

Allen, J.D. and Hanuschak, G.A., 1988, "The Remote Sensing Applications Program of the National Agricultural Statistics Service: 1980 - 1987," U.S. Department of Agriculture, NASS Staff Report No. SRB-88-08

Battese, G.E., Harter, R.M. and Fuller, W.A., 1988, "An Error-Components Model for Prediction of County Crop Areas using Survey and Satellite Data," Journal of the American Statistical Association, 83(401): 28 - 36.

Bellow, Michael and Graham, Mitchell, 1992. "Improved Crop Area Estimation in the Mississippi Delta Region Using Landsat TM Data," American Society of Photogrammetry and Remote Sensing Convention, Washington, D.C.

Bellow, M.E. and Ozga, M., 1991, "Evaluation of Clustering Techniques for Crop Area Estimation using Remotely Sensed Data," American Statistical Association 1991 Proceedings of the Section on Survey Research Methods, Atlanta, GA, pp. 446 - 471.

Caudill, Charles E. and Hanuschak, G.A., 1983. "Management Issues of Integrating Earth Resource Satellite Data in to the U.S. Department of Agriculture's Domestic Crop-Area Estimation Program," Semi-Annual Meeting of the Institute of Management Science/Operations Research Society of America, Orlando, FL.

Cheng, Thomas D., G.L. Angelici, R.E. Slye, and M. Ma, 1989. "Computer-Aided Boundary Delineation of Agricultural Lands." NASA Ames Research Center, Tech. Memorandum 102243, Moffet Field, California.

Cochran, William G., 1977. "Sampling Techniques," New York, NY: John Wiley and Sons, Inc.

Cook, P.W., 1982. "Landsat Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service 1972 - 1982," Washington, D.C.

Cotter, James and C. Mazur, 1992. "Automating the Development of Area Sampling Frame Using Digital Data Displayed on a Graphics Workstation," Eurostat Conference on New Technologies for Statistics, Bonn Germany.

Cotter, Jim and Jack Nealon, 1987. "Area Frame Design for Agricultural Surveys," U.S. Dept. of Agriculture, National Agricultural Statistics Service.

Craig, Michael E., 1992. "Applications of Advanced Technology for Agricultural Statistics." United Nations, Conference of European Statisticians, Bratislave, Czech and Slovak Federal Republic.

Hanuschak, George, 1977. "Landsat Estimation with Cloud Cover," Proceeding of the 1976 Symposium on Machine Processing Remotely Sensed Data, West Lafayette, Indiana.

Hanuschak, G.A., R.D. Allen and W.H. Wigton, 1982. "Integration of Landsat Data into the Crop Estimation Program of USDA's Statistical Reporting Service 1972 - 1982." Paper presented at the 1982 Machine Processing of Remote Sensed Data Symposium, West Lafayette, IN.

Hanuschak, G.A., and K.M. Morrissey, 1977. "Pilot Study of the Potential Contributions of Landsat Data in the Construction of Area Sampling Frames," U.S. Department of Agriculture, Statistical Reporting Service.

Holko, Martin L., and Richard S. Sigman, 1984. "The Role of Landsat Data in Improving U.S. Crop Statistics," Paper presented at the Eighteenth International Symposium on Remote Sensing of Environment, Paris, France.

Houseman, Earl E., 1975. "Area Frame Sampling in Agriculture," U.S. Department of Agriculture, Washington, D.C.

Huddleston, Harold F., 1976. "A Training Course in Sampling Concepts for Agricultural Surveys," Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

Huddleston, Harold F., 1978. "Sampling Techniques for Measuring and Forecasting Crop Yields," Economics, Statistics and Cooperative Service, U.S. Department of Agriculture, Washington, D.C.

Johnson, R.A. and Wichern, D.W., 1988, "Applied Multivariate Statistical Analysis," Prentice Hall, Englewood Cliffs, N.J.

Ozga, M., Mason, W.W. and Craig, M.E., 1992. "PEDITOR - Current Status and Improvements," in Proceedings of the ASPRS/ACSM Convention, Washington, D.C.

Ozga, M., W. Donovan and C. Gleason, 1977. "An Interactive System For Agricultural Acreage Estimates Using Landsat Data," Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, IN.

Sigman, Richard and Gail Walker, 1982. "Use of Landsat for County Estimates of Crop Areas: Evaluation of the Huddleston-Ray and Battese-Fuller Estimators," SRS Staff Report No. AGES 920909, U.S. Department of Agriculture, Statistical Reporting Service.

Winings, Sherman B., 1982. "Landsat Image Availability for Crop Area Estimation." Paper presented at the Eighth International Machine Processing of Remotely Sensed Data symposium for Applications of Remote Sensing, West Lafayette, IN.