

SEMINAIRE

СЕМИНАР

SEMINAR

STATISTICAL COMMISSION AND
ECONOMIC COMMISSION FOR EUROPE



CES/SEM.31/R.31
26 February 1992

CONFERENCE OF EUROPEAN
STATISTICIANS

ENGLISH ONLY

Seminar on integrated statistical information
systems and related matters (ISIS '92)

(Bratislava, Czech and Slovak Federal Republic, 18-22 May 1992)

APPLICATIONS OF ADVANCED TECHNOLOGY FOR AGRICULTURAL STATISTICS

Report submitted by the U.S. Department of Agriculture,
National Agricultural Statistics Service, Research and
Applications Division, Survey Research Branch¹

I. INTRODUCTION

1. The National Agricultural Statistics Service (NASS), through a network of State Statistical Offices (SSO), surveys farms, ranches, and agri-businesses to make estimates on every aspect of U.S. agriculture. NASS must constantly revise its data collection procedures and estimation methodology to keep pace with advances in technology. The Survey Research Branch of NASS is responsible for research and development of statistical methodology and advanced technology with the goal of improving the Agency's current and future operating programs.

2. New computer, data collection, and data processing technologies are constantly being evaluated. Two of the Branch's Sections deal directly with technology related issues. The Remote Sensing Section administers the agency's activities related to the use of remotely sensed data for crop forecasting and estimation. The Technology Research Section administers and is responsible for research and development related to the implementation of new technologies. Some current research topics for both Sections are mentioned below.

¹ Prepared by Michael E. Craig

II. OVERVIEW OF CURRENT TECHNOLOGY RESEARCH

3. The Technology Research Section (TRS) is currently investigating a very wide range of applications. These include: interactive data editing, geographic information systems, computer assisted personal interviewing, mobile computerized data entry and edit, visual data analysis, heterogeneous network research, digital research survey data library, touchtone data collection, and electronic archiving of historic research reports.
4. Interactive data editing research will evaluate and implement the use of interactive software on PC-LAN's to detect suspicious data values for review. Blaise interactive editing software has been purchased from the Netherlands, Central Bureau of Statistics for all SSO's and HQ.
5. Recent advances in portable computer technology have allowed new approaches to data collection. Research projects to consider these approaches have been labeled as Mobile Computerized Data Entry and Edit (McDEE) studies. McDEE studies include such topics as computer assisted personal interviews (CAPI) and on-site data collection where no interview is required. This research considers alternatives to the paper and pencil way of collecting survey data. McDEE has the potential of collecting data faster, cheaper and with fewer errors by incorporating several technological factors into the data collection arena. Since the data are being collected in a digital form, no re-keying of the data is necessary. The data can then be sent into the office using a modem. Using a software solution to collect data can keep the flow-control of moving through the instrument much more accurate. The McDEE project described below involved the recording of livestock sales receipts by enumerators using a laptop computer to directly enter the data.
6. Another project will use Geographic Information Systems technology to analyze spatially referenced data. NASS has been given the responsibility to build a GIS database as part of USDA's Water Quality Program Plan to support the President's Water Quality Initiative. Environmental Research Systems Institute's ARCINFO has been chosen as the GIS software package. The use of Geographic Positioning System (GPS) receivers is being investigated as a way to tag the data points with latitude/longitude geographic referencing.
7. A library of SAS datasets from the Agency's Agricultural Survey Program is being established. The datasets are being transported from mainframe computers and archived in-house on tapes accessible through a Sun Workstation. Users can retrieve datasets and convert them for use with either Sun SAS or PC SAS. Efficient access to farm record level survey data is critical to the research efforts of the Agency.
8. Other TRS projects in initial stages include: the evaluation of a software package called PC-Wave as a tool for the display of advanced statistical graphics, the joint utilization of DOS and Unix operating systems running on the same network for database applications, and an evaluation of the technology of electronic scanning and archival storage of historic reports.
9. Remote Sensing Section topics include the improvement of crop acreage estimates in the Mississippi River Delta Region using Landsat Thematic Mapper Data and an evaluation of weather satellite data, such as the Advanced Very High Resolution Radiometer (AVHRR), for assessment of crop condition.
10. The 1991 Delta Remote Sensing Project implemented a crop acreage regression estimator based on digital classification of Landsat satellite images. This project produced state and county estimates in an operational timeframe, and is the first NASS large volume test of the estimation system for the Landsat Thematic Mapper (TM) scanner data. Spring and late summer TM images over the same areas were combined for the digital analysis. Earlier applications of the remote sensing based regression approach utilized rented mainframe computers and large amounts of paper generated products which had to be manually overlain on

light tables. The current in-house microcomputer system utilizes a MicroVAX 3500 linked via local area network to desktop personal computers. Recent enhancements to the PEDITOR software system allow users to analyze satellite data in a paper free environment. In addition to numeric estimates, color land use maps based on classified satellite images will be generated on a Tektronix 4693DX thermal printer for each county included in the 1991 Delta Project.

11. NASS has begun a cooperative agreement with the Remote Sensing Laboratory of the Agricultural Research Service to investigate the use of weather satellite data, such as the Advanced Very High Resolution Radiometer (AVHRR), for assessment of crop condition. To date, no satellite data technique has been able to provide direct indications of crop yield with acceptable precision or timeliness. However, significant progress has been made in the area of crop condition assessment. 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.

III. IN-HOUSE COMPUTER SYSTEMS

12. 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 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 9-track tape and Exabyte tape cartridge capabilities in addition to several external disk drives and other peripherals.

13. Smaller volume analyses utilize 386 or higher 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, scanners, video cameras, and digitization tablets. Other equipment includes laptop and notebook computers, such as GRiD Pads and Zenith Supersports.

IV. THE CASS PROJECT

14. A major research effort over the last three years that has now become operational is the Computer Aided Stratification and Sampling (CASS) project. Many NASS estimates are based on enumeration of uniquely described areas of land called segments. These segments are selected from area sampling frames constructed for each state.

15. The CASS System incorporates digital inputs in the form of thematic mapper (TM) LANDSAT data and USGS Digital Line Graph (DLG) data. The TM data (1:100000 scale) serves as a base to delineate land use according to our stratification scheme. For boundary identification, DLG data at a 1:100000 scale is overlaid onto the TM image using a graphics plane. The following steps define the CASS approach.

a. **Displaying satellite data** - Three bands of the seven band Landsat image are loaded into the system for display at 30 meter resolution. The image can be zoomed on the screen to various levels. This satellite image provides the most recent look at the area to be stratified. This digital product parallels the use of the paper color print of the satellite data and the black and white high altitude photography.

b. **Color mapping** - The display window uses 8 bits for each of the three bands yielding a sharp and colorful image. A color map is one which assigns brightness levels to each of the bands. A dynamic mapping function allows the user to control the brightness and contrast of each of the 3 planes. These settings can then be preserved in a file for later use. Several other functions allow the user to look at a single band, look at a histogram of each band, and do a linear, piecewise, or equiprobable mapping. The optimum mapping is the one which best distinguishes the cultivation and boundaries. This is checked by locating previous survey fields on the screen, and noting the colors.

c. **Display and registration of DLG data** - The DLG data are used as a reference. The stratifier must be sure that they are using good boundaries for the PSU and segment delineation steps. Current digital data includes transportation and hydrography. Political and administrative boundaries will be available soon. The original DLG files exist in a unit called a "panel." A CASS function reads these files from tape and creates disk files. These disk files can then be added together and placed in one file to cover a larger area, such as a county. If a county crosses UTM (Universal Transverse Mercator) zones, the data can be converted so that a single file can be saved. (Panels from different zones cannot be added together.) Another CASS function displays the data file to the screen. In order to precisely overlay the TM image with the DLG, another program is used to "rubber-sheet" (or register) the DLG data to the backdrop of satellite data by selecting several matching points in each data set and running a least squares regression to fit the rest of the data. These points and the regression are also saved in a file and used each time this particular DLG file is displayed. Lastly, this registration file allows the user to determine latitude and longitude coordinates of a given point.

d. **PSU delineation** - In this step, polyons will be drawn off and tagged with the appropriate stratum along with a unique PSU number. This is done by first determining what land use stratum a piece of land belongs to by interpreting the color TM display, and assigning it to a PSU (within a specific size range), using the physical boundaries identified by the DLG data and/or the LANDSAT image itself. This is done by keying in the PSU number (a stratum number followed by a sequence number), and using the mouse to pick points, and a button box to choose the appropriate commands. When the PSU is closed, the area is immediately displayed. This allows the stratifier to determine if the PSU is within the target size for that stratum. The polygons can be combined, split and reshaped. They are saved to a file to be later reviewed by an experienced stratifier. One check allows the user to check for overlapping polygons, while another checks for holes (or missing land areas). At any time, the user may list PSUs that have been created so far, to check numbering and areas.

e. **PSU breakdown into segments** - After the entire state has been stratified and the total area for each stratum has been calculated, a separate program is run to draw the sample of PSUs (first stage of sampling) which will be further broken down into final sampling units or segments. Not all PSUs are broken down - only those which were chosen using the sample select program. The user displays the stratification file (saved in the previous step) and types in the PSU number which is to be sampled. The software then erases all but the sample PSU from the screen. Many of the same functions which were involved in delineating the PSUs during the stratification phase are used to delineate the PSU into equal size segments. Similar quality control checks are done. When the PSU has been broken down into segments, a segment can be selected randomly using the segment selection command. This is then the second stage of sampling.

16. The CASS Workstation includes several pieces of equipment. The display terminal displays the color Landsat image, whereas the menu screen displays the software. The keyboard is used to enter commands onto

the menu screen, where the mouse is used to interact with the display terminal. The button box is also used in connection with the display terminal to handle the overlay planes (change the color of the display or turn the display on and off) and to zoom the image.

17. A UNIX-based Hewlett-Packard (HP) workstation is being used to handle the processing and storage requirements of the massive amount of data. The HP workstations possess the minimum capabilities for area frame development, i.e., three image planes, four overlay planes and a 1024 x 1280 display coordinate system. The configuration allows for three bands of satellite data in the image planes while utilizing the graphics planes for displaying DLG, PSUs, and the command menu. A recent modification allows the DLG to be displayed in more than one overlay plane. One reason for this is to separately color the transportation and hydrography data.

18. Some benefits of using CASS are listed below:

- a) The satellite data provides more recent data than the high altitude photography (which is important when a frame is used for 15 years), and provides better resolution than the current satellite image print.
- b) The tedious, labor intensive, error prone process of transferring from a satellite image print, to a high altitude photography, to a 1:100000 USGS map, to a digital file can be eliminated.
- c) The current process of digitizing PSUs to determine area (after a state has been completely stratified on paper) and as a backup, can be eliminated. The ability to immediately determine PSU or segment size will help the stratification and sample selection process.
- d) The digital aspect of the frame will allow a frame to be updated rather than having to start from scratch (which is currently done on paper due to the age of the paper, and the impossibility of erasing color pencil).
- e) Sample segments are chosen by the software using a random number generator instead of looking through a set of random number sheets.
- f) Allows the stratifier to more easily revise a PSU or PSUs.
- g) Allowing easier exploration of specialized area frames.
- h) Provide better control over the homogeneity of PSUs within a stratum.

19. Future considerations for CASS are:

- a) We have looked at the unsupervised and supervised classification of satellite data and hope to use it as an aid to the stratification process. However, more work must be done.
- b) As of yet, only bands 2, 3 and 4 have been used for display purposes. We are in the process of reviewing the use of the other bands (especially 5) to see more detail, and better see through clouds.
- c) When Landsat 6 becomes available, we will need to utilize the panchromatic band (which gives 15 meter resolution) with the spectral bands (which will still have 30 meter resolution).

- d) We are investigating the use of filters to improve the ability to detect "edges" or boundaries in CASS.
- e) A graphical user interface is currently being investigated.
- f) We are looking into using Census Tiger data and the corresponding Public Law File to locate the boundaries of a Census block on CASS and look up the number of households, to better determine the urban land classification. Topographic Quad maps may also be used to determine urban classification and the appropriate boundaries.
- g) The polygon data can be converted into a format that a Geographic Information System (GIS) can read. The possibility for reading GIS data into CASS is worth looking in to.

V. REFERENCES

Cotter, J. and Mazur, C.(1991): Automating the Development of Area Sampling Frames Using Digital Data Displayed on a Graphics Workstation. U.S.Department of Agriculture, National Agricultural Statistics Service Staff Report. Washington, D.C.

Hanschak, G.(1991): FY 1992 Survey Research Branch Missions and Projects Executive Summary. U.S.Department of Agriculture, National Agricultural Statistics Service Staff Report. Washington, D.C.
