

**COMPUTER-AIDED STRATIFICATION
AND SAMPLING**

BY
CATHY M. TOMCZAK

The National Agricultural Statistics Service (NASS) employs digital processing of satellite and geographic attribute data to create an area sampling frame for agriculture. This brochure will describe how computer-aided stratification and sampling (CASS) is used to facilitate the frame construction and sampling process. Ultimately, a specified number of sample blocks of land, called segments, are desired for survey enumeration. Sample segments are selected from various land use strata that encompass each State's total land area. Each segment must have identifiable physical boundaries for enumerators and have a known probability of selection for estimation.

Introduction

The **CASS workstation** includes several pieces of equipment (shown below). The monitor displays the color Landsat image and the text using a Graphical User Interface (GUI). The keyboard is used to enter commands or filenames, and the mouse interacts with the GUI and creates or edits polygons and segments. The function keys are also used to turn the overlay planes on and off, to change the color of the overlay planes, and to expand or contract the scale of the image.

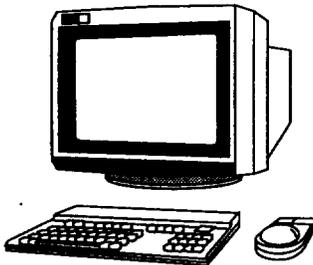


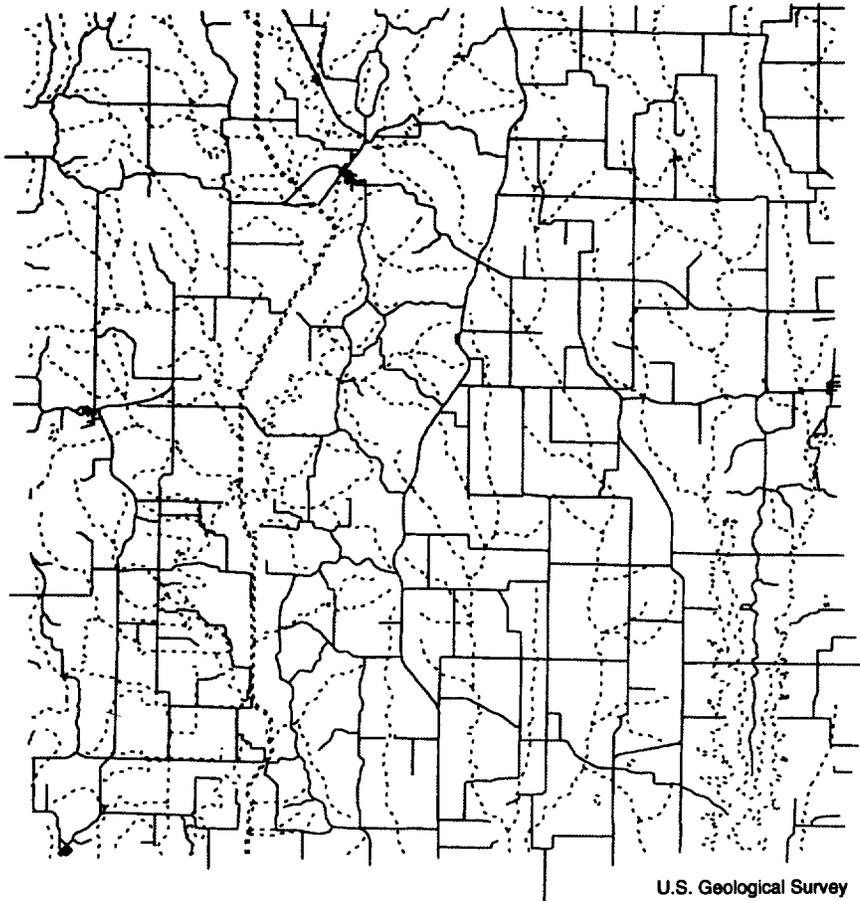
Figure 1: Image Data



EOSAT Corporation

The primary inputs to CASS include Thematic Mapper (TM) Landsat data (Fig. 1: Macon County, MO) and U. S. Geological Survey Digital Line Graph (DLG) data (Fig. 2). The Landsat data provide 30-meter resolution in 7 bands. Three of these bands are selected to be displayed as a red, green, and blue plane on a graphics workstation. The DLG data are displayed as an overlay plane to aid in identifying physical boundaries on the TM data. A DLG display for roads (solid red) and waterways (dotted blue) is shown in Fig. 2. Census digital data (acronym TIGER) is used to display county boundaries.

Figure 2: Road and Water DLG



U.S. Geological Survey

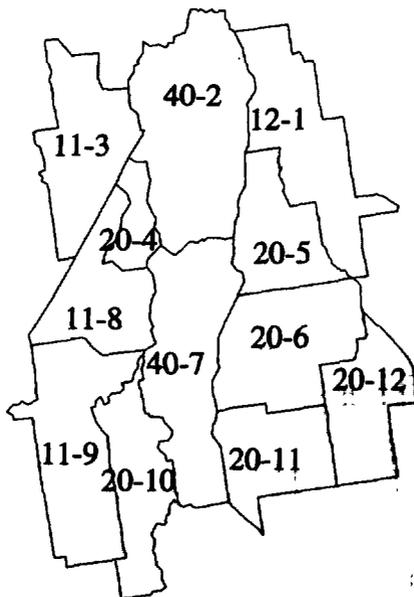
Two initial tasks are to create a **color map file** and a **control point file**. The color map file assigns each data value from the Landsat image to the appropriate red, green, and blue color intensity. Many color variations are possible. The optimal color map for stratification is the one which best distinguishes cultivation. The optimal color map for sampling is the one which best distinguishes physical features (boundaries). The control point file registers (or rubber-sheets) the DLG data to the TM data.

Stratification

To create a new area frame, the stratifier must divide the land into homogeneous land use strata. Strata are typically defined as low, medium, or high percent cultivated, nonagricultural lands, urban use, ag-urban, or water. The user must consider which crops are grown in the State, what maturity phase the crops were in at the time the satellite image was taken, and how to interpret the colors on the Landsat image.

The previous year's June survey data are used to aid in the proper interpretation. The stratifiers interpret the color Landsat image on the display screen, and divide each piece of land to a Primary Sampling Unit (PSU) within a specific size range. Physical boundaries are identified on the Landsat image itself, enhanced by DLG data. This is done until all land areas in a county have been assigned to a stratum and PSU, as in Fig. 3. Quality checks are done to assure accuracy and consistency throughout the State.

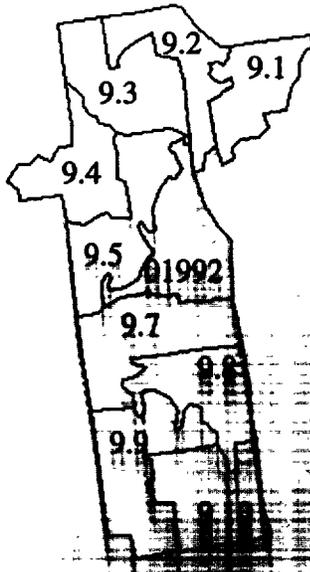
Figure 3: PSU's



Sampling

A sample of PSU's within each State is selected with probability proportional to size. The size is based on the number of individual segments contained in the PSU. Each one of the sample PSU's must then be subdivided into segments of equal size. Each segment has a specific target size depending on its land use stratum. Each individual segment resembles the full PSU as closely as possible, and is delineated with the best physical boundaries available (Fig. 4). Note that segment 9.6 was selected to be in the sample, and is given the segment number 01992. The person doing this work must also be able to interpret the colors on the Landsat image to make each segment "look like the full PSU." Quality checks are also done in this phase. The Sample Select Unit creates a color print of the selected segment and a segmap (a county map which identifies the locations of the sampled segments), which are then used in the Sample Preparation Unit to prepare the selected segments for the States. They also place an order for an enlarged black and white photo (scale = 5.28" or 8" per mile, depending on the photo source) of each individual sampled segment.

Figure 4: Segment Breakdown



Preparation

The color (segment) print is used as a reference to transfer the selected segment boundaries onto the photo enlargement. There are several quality checks done during the preparation of the materials to assure accurate boundary transfers, boundary quality, and segment size.

A copy of the segmap and the finished photo enlargement (Fig. 5) are sent to the State for use by the survey enumerator.

Figure 5: Photo Enlargement



Benefits

- The Landsat data can be dynamically colored, rather than having fixed colors as in the Landsat image print.
- The base display scale of the Landsat image (1:100,000 currently) can be changed. The base scale of the next Landsat satellite will approach that of the current National High Altitude Photos (1:40,000 scale).
- The digital data are available every 18 to 21 days. Aerial photographs are taken every 5 or 6 years.
- The area frame can be revised rather than recreated anew when updating the frame.
- Time and labor can be saved. Fewer people are required to complete the same number of States in a given period of time.
- CASS makes it possible to maintain a latitude/longitude database of all selected segments.
- An increase in quality should result as the analysis is done at 1:100,000 scale rather than the 1:250,000 scale, checks for missing or overlapping land are done, there is no need to transfer boundaries between materials, and the user is given immediate feedback on the required dimensions of size.

Under Development

- An enhancement currently being developed is a GUI which will provide more operational flexibility to the user. With the current system, the user must exit from one operation to access a feature in another. The GUI system will allow access to all features without exiting the current operation.

Future Possibilities

- Unsupervised land classification.
- An edge-enhancing filter.
- The exchange of information between CASS and the Agency's Geographic Information System.
- Use of imagery with resolution approaching the level of detail of aerial photography. Such resolution would allow the segment coverage to be produced directly from the image used for the breakdown rather than ordered from the photo lab.



**National
Agricultural
Statistics
Service**

U.S. Department of Agriculture