

2/92

AUTOMATING THE DEVELOPMENT OF AREA SAMPLING FRAMES USING DIGITAL DATA DISPLAYED ON A GRAPHICS WORKSTATION

James J. Cotter *

Cathy Mazur

National Agricultural Statistics Service

1992

ABSTRACT

The National Agricultural Statistics Service uses area sampling frames to conduct surveys to collect agricultural data in the United States. These area frames are created by assigning land area to land use categories, and then delineating parcels of land for sampling purposes. Currently, paper-based products are used such as satellite imagery, high altitude photography and United States Geological Survey map products. This fall, a sampling frame will be created using the Computer Assisted Stratification and Sampling (CASS) system which will enable us to automate this process by displaying digital satellite data and digital map data on a graphics workstation. The new system will enable the cartographer to create a more accurate sampling frame by accessing more recent data and will enable us to complete the process much more quickly.

Keywords

stratification, area frame sampling, digital satellite imagery, digital line graph.

* THE AUTHORS ARE RESPECTIVELY, HEAD OF TECHNOLOGY RESEARCH SECTION WITHIN THE SURVEY RESEARCH BRANCH AND GROUP LEADER IN CHARGE OF AREA FRAME DEVELOPMENT IN THE AREA FRAME SECTION WITHIN THE SURVEY SAMPLING BRANCH OF THE NATIONAL AGRICULTURAL STATISTICS SERVICE

1. INTRODUCTION

The National Agricultural Statistics Service (NASS) has been developing, using and analyzing area sampling frames since 1954 as a vehicle for conducting surveys to gather information regarding crop acreage, cost of production, farm expenditures, grain yield and production, livestock inventories and other agricultural items. An area frame for a land area such as a state or county consists of a collection or listing of all parcels of land for the area of interest. These land parcels can be defined based on factors such as ownership or based simply on easily identifiable boundaries as is done by NASS. Area frames provide complete coverage with all land areas being represented in a probability survey with a known chance of selection. The existing procedure used to develop area frames is slow, labor intensive and expensive. The development of an area frame for a single state may require 11,000 hours and cost over \$150,000. This paper will briefly describe the currently used materials and procedures in developing an area frame. This will be followed by a description of a research project which is now going operational to develop area frames using digital inputs. For more information on area frame development, consult the bibliography.

2. PAPER-BASED AREA SAMPLING FRAMES

2.1 Materials Used in the Current Procedure

Area frames are currently developed on a state by state basis. The materials used in the stratification process include:

Satellite Imagery: Satellite imagery is derived from digital data collected by scanners aboard satellites. Presently, the paper-based imagery product from the LANDSAT satellite is used. A scanner mounted on the satellite collects the reflected and emitted energy from the ground. Two types of scanners are used: a multispectral scanner (MSS) and a thematic mapper (TM). TM is the preferred product for stratification. TM is more costly due to its better resolution. The paper TM product is scaled at 1:250000.

National Aerial Photography Program (NAPP): NAPP is the product of a consortium of federal agencies, each of whom need and use aerial photography. Contact prints are used which are nine inches square and are scaled at 1:40000. NAPP is a primary stratification tool. Nearly all of the U.S. has been photographed through the NAPP program.

Topographic Quadrangle Map (Quad): These maps are produced by the U.S. Geological Survey (USGS) and the preferred scale is 1:24000 (7.5 minute series - 2.6 inches to a mile) which makes them useful for urban and ag-urban stratification and sampling.

Bureau of Land Management (BLM) Map: These maps, scaled at 1:100000, show the distribution of the federal and state land. They are useful in western states for delineating the range strata and for locating the boundaries of Indian reservations.

U.S. Geological Survey 1:100000 Map: These maps are of high quality and provide NASS with an accurate map base on which to stratify and digitize (to be defined later).

These materials, as they are currently being used, are all paper-based products.

Table 1. Land-Use Strata Codes and Definitions

STRATUM CODE	DEFINITION	TARGET SIZE	
		miles ²	kilometers ²
11	General Cropland, 75% or more cultivated.	6-8	15.5-20.7
12	General Cropland, 50-74% cultivated.	6-8	15.5-20.7
20	General Cropland, 15-49% cultivated.	6-8	15.5-20.7
31	Ag-Urban, less than 15% cultivated, more than 100 dwellings per square mile, residential mixed with agriculture.	1-2	2.6-5.2
32	Residential/Commercial, no cultivation, more than 100 dwellings per square mile.	.5-1	1.3-2.6
40	Range and Pasture, less than 15% cultivated.	12-16	31.1-41.4
50	Non-agricultural, variable size segments.		
62	Water		

2.2 Stratification

The process of land-use stratification is the delineation of land areas into land-use categories on photography and a corresponding map base utilizing satellite coverage as an aid. Table 1 displays the set of land-use categories which were used in the development of Missouri's area frame in 1987. The purpose of stratification is to reduce the sampling variability by creating homogeneous groups of sampling units. Although certain parts of the process are highly subjective in nature, precision work is required of the personnel stratifying the land (called stratifiers) to ensure that overlaps and omissions of land area do not occur and land is correctly stratified.

Perhaps the most important concept conveyed during the initial training of personnel is the idea of using quality boundaries. A quality boundary is a permanent or, at least, long-lasting geographic feature which is easily found and identifiable by an interviewer. If an interviewer cannot accurately locate a segment in a timely manner, there is the potential for nonsampling errors to be introduced into the survey data. If the field interviewer, unknowingly, does not collect data associated with all of the land inside the sampled area or collects data for an area outside of that selected, then survey results will be biased.

The objective of using permanent boundaries and the objective of obtaining homogeneous sampling units within a stratum often conflict in the actual practice of area frame stratification. Concessions may have to be made in marginal situations. Given that the area frame is to be used over a period of 15-20 years and represents a major investment, the best and most permanent boundaries must be used. Roads and rivers make good strata boundaries, while intermittent streams and field edges do not and should rarely be used. The following is a list of the most frequently-used geographic features which represent strata boundaries. The list is ranked by quality from highest to lowest:

- Paved highways.
- Secondary all-weather roads.
- Local farm to market roads.
- Railroads.
- Permanent rivers and streams.

The stratification is performed a county at a time for administrative purposes. Each stratifier is assigned a county and will work on that county until its completion. Stratification generally begins with determining the urban and ag-urban strata for the county. The agricultural areas are then stratified using the TM satellite imagery. The imagery is used primarily to ascertain where the cultivated areas and the non-cultivated areas are present in a county. The imagery is very useful in stratification because it is so timely. Although aerial photography may be one to five years old, the Landsat imagery usually covers the most recent growing season, providing a very recent look at the area. Using the Landsat imagery for locating crops and pasture and the photography for boundaries, the

Quality assurance is a major concern during the stratification phase. Throughout the process, checking and rechecking is performed to ensure a high quality product and to obtain the benefits of a second opinion by someone with a more experienced „eye“.

After the stratification on the photography has been reviewed and approved, the strata boundaries will then be transferred to the map base (also called the frame sheets or frame maps). The map will later be digitized (electronically measured) to determine the areas of the primary sampling units (to be defined later). Once this transfer is completed, the transfer is examined by the checker and the next phase of stratification is begun...construction of primary sampling units.

2.3 Construction of Primary Sampling Units

The next step in the development of the area frame is to further subdivide the strata into primary sampling units (PSUs). The desired size of the PSU varies by strata, but contains, on the average, six to eight final sampling units or segments. The minimum PSU size is generally one segment. The use of primary sampling units introduces economic saving into area frame sampling. An entire frame need not be divided into segments in order to select a sample. The strata blocks are broken down into PSUs and a sample of PSUs is randomly selected. Only the randomly selected PSUs will be further subdivided into segments - saving a tremendous amount in labor costs. In delineating PSUs, the main focus is not homogeneity of land-use - that will have already been accomplished with the land-use stratification. The main concern is to achieve the desired size with good boundaries while trying to maintain that each PSU is a smaller representation of the strata as a whole.

The frame maps are reviewed by a statistician for completeness as a final check. The polygons created by drawing the PSUs are examined to make sure they form a closed polygon. The numbering system is checked for strata identification accuracy and sequential accuracy. The frame maps are further checked to ensure that omissions and overlaps do not exist. Once these checks have been accomplished, the frame maps are ready for the next step in the process...measuring the PSUs.

2.4 Digitization

The conversion of map points into two-dimensional X-Y coordinates is called digitization. Digitization involves electronically measuring the area of the PSUs on the frame maps. The PSUs need to be measured to determine the number of segments per PSU for sampling purposes. Electronically recording the PSU areas allows:

- measuring the PSU accurately,
- quality assurance,
- retaining a digital backup copy of the frame map in the unlikely event that a frame map is lost.

(0,0), is established for X-Y coordinates on the map. X-Y coordinates, tagged with the appropriate identification, uniquely describe the borders of a PSU and therefore create a polygon for each PSU. The digitizing software records the X-Y coordinates in a file. Using the map scale, the area of each polygon (PSU) in the county is calculated in terms of square miles and stored in a file for that county.

The PSU areas for each county are summed and compared against the official county size. The same procedure is done for the state area. County areas are allowed to vary 3.0 percent from the published area. The accumulated state area is only allowed to vary 0.5 percent from the published area. The county area is allowed more variance because of the smaller area involved and because primary sampling units are allowed to cross county boundaries. Since the stratification is never allowed to cross state boundary lines, only a small amount of error is allowed.

The PSU areas are then accumulated for each stratum at the state level. The area of the PSU divided by the target segment size for the stratum is equal to the total number of ultimate sampling units (segments) in that PSU (rounded to the nearest integer). Summing the number of segments will yield the total number of segments in the stratum. This information will be used in determining the number of samples to be chosen for the entire state.

2.5 Sample Selection

After the total number of sample segments to be used in a state has been determined, a program is run to select the PSUs which will be further broken down into sample segments. Each segment has a specific target size depending on the stratum it is associated with such that each individual segment closely resembles the full PSU (as much as possible) with the best physical boundaries available. The selected PSUs are located on the frame map and the PSU boundaries are transferred to photography. The selected PSU is then broken down along identifiable boundaries into the required number of segments. The segments are manually numbered and a random number is chosen between one and the number of segments in the PSU. The segment corresponding to the random number is the selected segment. A photo enlargement of the selected segment will eventually be sent to the interviewer for enumeration

3. DIGITAL-BASED AREA SAMPLING FRAMES

3.1 Research Background

NASS has been involved in a cooperative agreement with the National Aeronautics and Space Administration (NASA), the U.S. Space Agency, to develop area frames using digital inputs. The project with NASA began three years ago. Although the initial research agreement with NASA will expire in the fall on 1991, NASA will continue to

ment with the Ecosystem Science and Technology Branch (ECOSAT), a group located at Ames Research Center, Moffett Field, California.

NASS and ECOSAT have a history of cooperation in remote sensing research. The two agencies have worked together on a number of projects since the late 1970s. EDITOR, a software package for large area crop acreage estimation utilizing LANDSAT MSS data and associated ground data, was written by NASS and ECOSAT personnel. Eighty percent of PEDITOR, a portable version of EDITOR, that is, a system suitable for implementation on a variety of hardware devices, was written at ECOSAT. PEDITOR is the primary software tool for remote sensing operational work in NASS. ECOSAT assembled a prototype microprocessor-based workstation, called MIDAS, for NASS and assisted with an experiment to determine the performance characteristics of the system when generating area estimates in the NASS operational environment. ECOSAT created display software for NASS that is compatible with PEDITOR and a precursor for the system required for area frame development.

The development of the area frame software has been an iterative process. The software supplied by ECOSAT to NASS has been constantly evaluated for completeness and ease of use by the Area Frame Section of NASS. This evaluation by the students and supervisors who work on the system, has led to a close working relationship with the people at NASA. Suggestions for improvement and modification are addressed on a very timely basis.

A new area frame system called the Computer Assisted Stratification and Sampling (CASS) system, based on PEDITOR concepts and integrated into the PEDITOR software will strengthen both the research and operational remote sensing programs at NASS and NASA and the development of area frames. One particular advantage to this approach is the ability to use digital information relating to land use from previous years' surveys as an aid to the development or updating of area frames. Other benefits will be discussed later.

3.2 The CASS System

CASS 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.

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-based satellite imagery and the black and white high altitude photography.

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.

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“ (7.5 or 15 minutes in size). 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 Universal Transverse Mercator (UTM) zones, the data can be converted so that a single county file can be utilized. (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.

PSU delineation - In each county, polygons will be drawn off and tagged with the appropriate PSU number, which consists of a stratum number and a sequence number. This is done by determining what land use stratum a piece of land belongs to by interpreting the color TM display. At the same time, a PSU within some specific size range is delineated, using physical boundaries identified by the DLG and/or TM data. In CASS, this is done by keying in the PSU number, and then utilizing the mouse to pick points along the desired boundaries. When the PSU is closed, the area is immediately calculated and displayed. This allows the stratifier to determine if the PSU is within the target size for that stratum. If the polygon is too small or too large, the polygons can be combined, split or reshaped. When a county is completed, the polygons are saved to a stratification file to be later reviewed by an experienced stratifier. The user has the ability to check for overlapping polygons and 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.

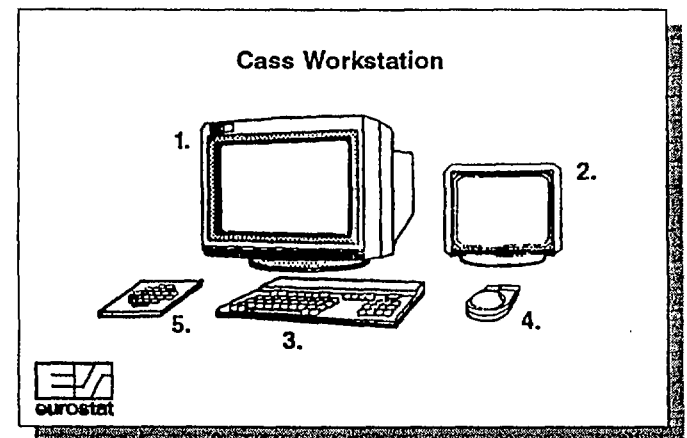
PSU breakdown into segments - After the entire state has been stratified and the total area for each stratum has been calculated, a separate program is run to draw the

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.

3.3 The CASS Workstation

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.

Figure 1



At this time, 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, that is, 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.

3.4 The Initial Test

The purpose of the initial test was to gain basic experien-

Digital data covering three north-central counties in Missouri (Linn, Livingston and Macon) were used. These counties were chosen partly because the Area Frame Section had developed a new area frame for Missouri in 1987 for use in 1988 and because of the availability of digital data. A few results are listed.

- 1) This test proved that stratification using CASS was possible.
- 2) CASS proved faster, as 2.5 - 3 weeks were required for the current stratification, and 2 - 3 days for CASS.
- 3) Several software enhancements were made. It was about this time that we decided to add sample select modules.
- 4) One important note is the subjective nature of the work. In this test, 3 people did the same counties in CASS and 2 people did the paper stratification.
- 5) It was also about this time that we switched from a SUN display to a Hewlett-Packard (HP) display. The screen size increased from roughly 512 to 1024 pixels. The number of overlays increased from 2 to 4, and the HP enabled the full TM scene to be loaded into memory rather than just one screen's worth.

3.5 The Expanded Test

The next phase of research was to represent a pseudo-operational environment. A larger area in Michigan was selected (21 counties) and system requirements were to

be considered. Reasons for selecting this area was first, that the state had just received a new frame in 1989 (implemented in 1990), and second, that the Remote Sensing Section of NASS had recently done some work in the dry bean area of Michigan in regards to supervised classification (therefore data was available). In this test, only one person did each county. Because of the recency of the paper stratification, a potential bias exists in the CASS stratification.

3.6 Analysis

This analysis is based on results from the expanded test using the 21 counties in Michigan. Evaluations are both quantitative and qualitative in nature.

Table 2 compares the total area stratified in the Michigan study using the CASS system versus the current manual mode. The percentage difference (given as CASS-operational/operational) is also shown. Some difference is shown for stratum 12 and stratum 31 with a larger difference shown for stratum 40. Comments from stratifiers generally favored the CASS system for the stratum 40 difference as they felt they were better able to pick out the woodland areas.

Figure 2

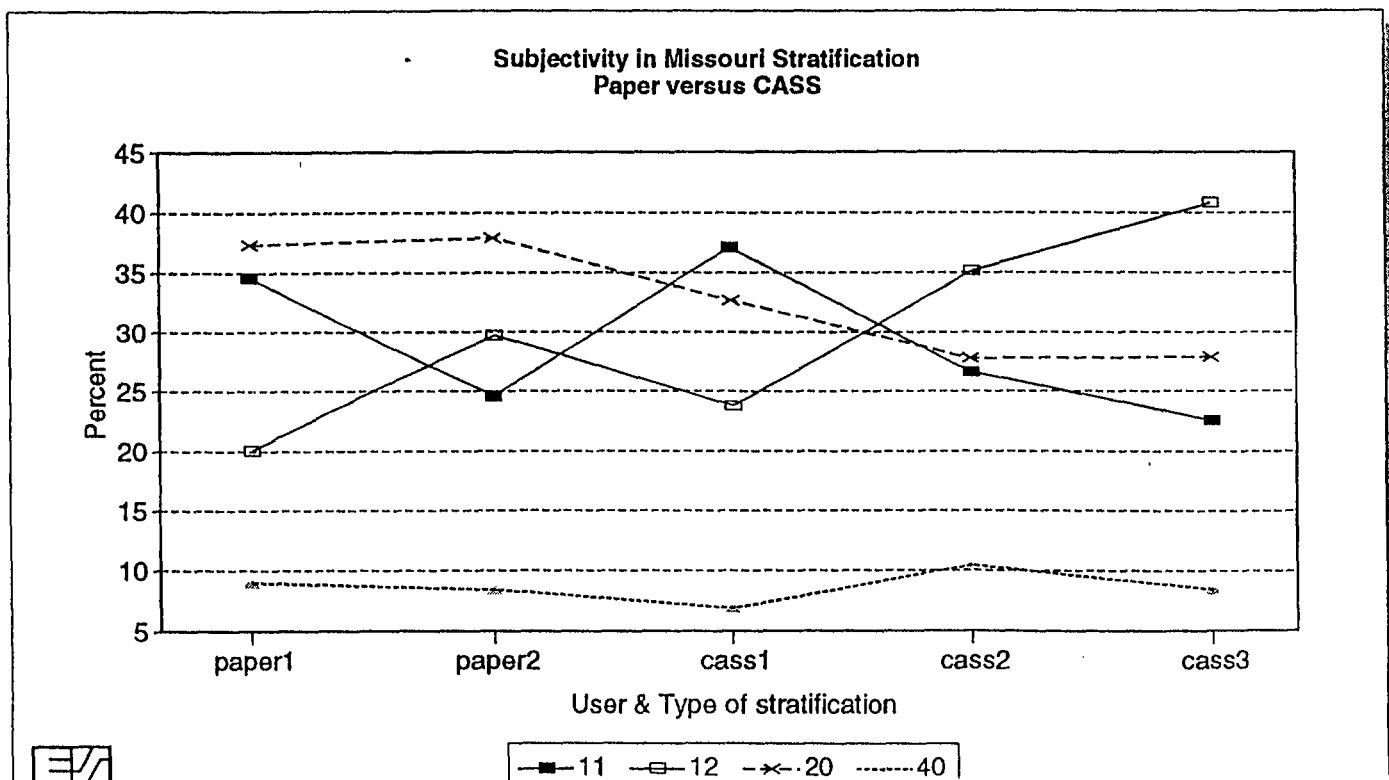


Figure 3

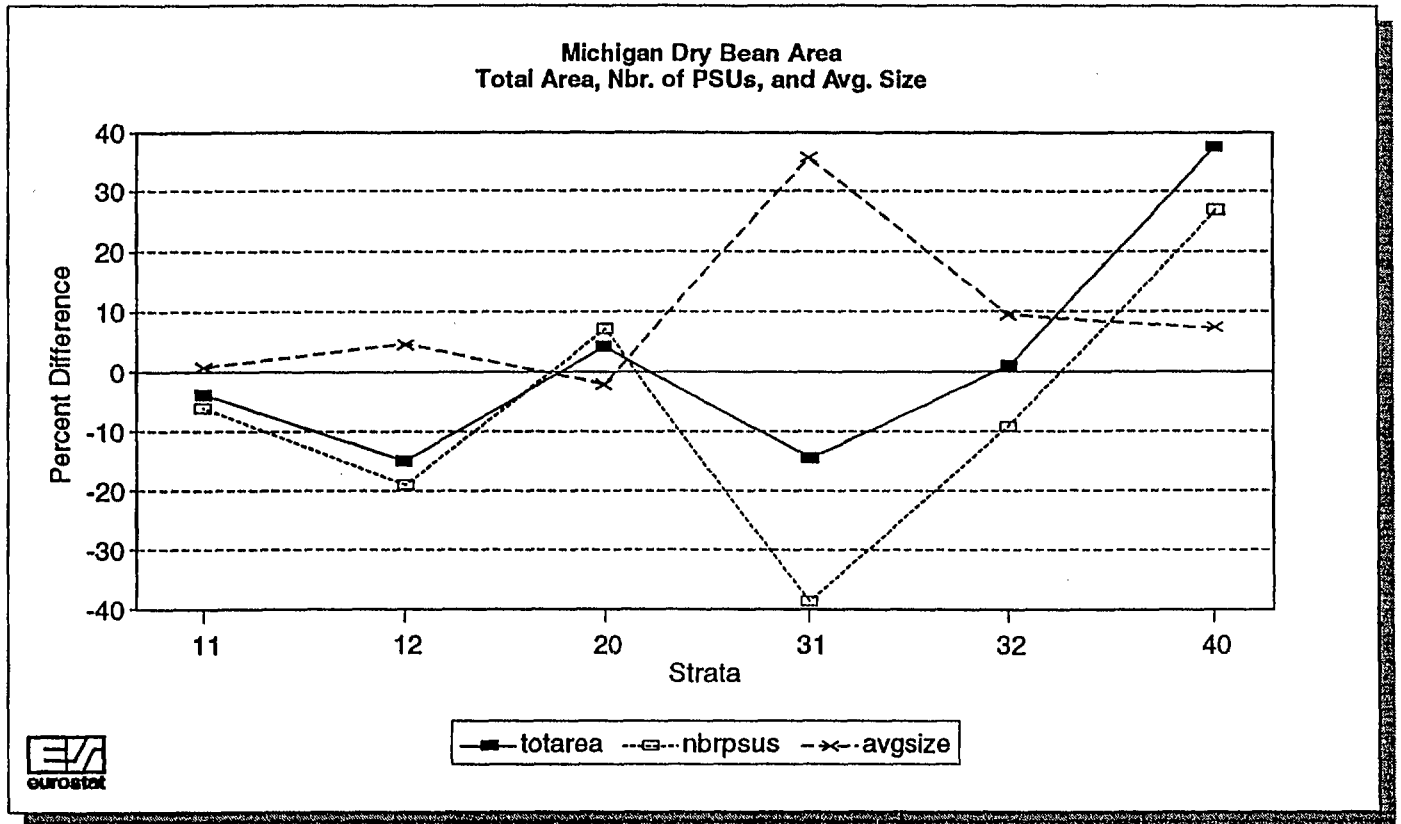


Table 2.

Comparison of Total Area by Stratum for the 21 Counties in the Michigan Study. Area is in square miles.

Stratum	Operational	CASS	%Difference
11	5427.4	5225.8	- 3.7
12	2547.8	2175.6	-14.6
20	3264.0	3448.0	5.6
31	648.0	555.3	-14.3
32	168.4	170.7	1.4
40	1347.3	1856.2	37.8

Table 3.

**Average PSU Size by Stratum for the 21 Counties
in the Michigan Study. Area is in square miles.**

Stratum	Operational	CASS	% Difference
11	6.4	6.5	1.7
12	5.3	5.6	4.9
20	6.1	6.0	-1.5
31	1.2	1.6	35.8
32	0.6	0.7	10.0
40	9.7	10.5	8.1

Table 3 describes the average PSU size by stratum as a comparison between the operational mode and the CASS system.

The quantitative data are interesting to review but the qualitative aspect of the study focuses on more informative issues. When the Michigan test was completed, the users compared the paper and CASS frames of 5 counties in Michigan. The issues raised, relate to the digital display, stratification, and sample selection.

A few improvements to the digital display will help the user do a better job the next time.

a) TM

- It was determined that the display of the satellite image could be improved. Initially, the zoom levels were set at 1, 2, 4, 8, and 16. At the end of the Michigan test, zoom levels were changed to 1, 2, 3, 4 and 5.
- Initially, each stratifier created a color map for their county. This created a problem during the review, as each person's color map was different. It was then decided to try one color map per LANDSAT scene in the future. Also, the stratifiers were not aware of some of the mapping techniques, which might have helped in the LANDSAT interpretation.

b) DLG

- The digital line graph data was entered into one overlay plane. Only at the end of the test, was a modification made to display DLG in different planes. In this way, roads can be displayed in one plane (in one color), and water in another plane (in another color).
- Also, the line display was improved. Because of an error, lines did not retain the dots and dashes when the image was redisplayed or zoomed. This has been fixed.

- Also, the points marked by the user will appear larger at the zoom levels two or greater.
- Finally, the lines all „look“ the same within a given overlay plane. Therefore, the stratifiers initially used illegal boundaries (such as swamp boundaries) which „looked“ the same as legal boundaries. A program was then written to delete all illegal boundaries from the DLG files. However, work is currently underway to better display the DLG so the user can „recognize“ the specific type of boundary.

A few comments about the stratification are mentioned below.

a) EXPERIENCE

- The color interpretation was a new experience, and it will take time to become proficient at this.
- Also, on paper, entire strata are delineated and then split into PSUs. On CASS, strata are determined but delineated at the PSU level.
- Looking back, several people commented that they might have done better with the experience they have now.

b) STRATIFICATION

- In general, the stratifiers felt that the stratification of cultivated areas was better in CASS. As to urban areas, the paper did better, due to increased resolution of the high altitude photography. On paper, houses can be seen somewhat, but on CASS the only guide to cities were the closeness of roads. A few ideas to handle this are covered in the „Future Considerations“ section. Lastly, the water was more detailed on the paper maps, but by displaying one band, the user can clearly see the water, and the DLG displays streams and rivers.

c) BOUNDARIES/CLOUDS

- As to the boundaries, the users felt the current method was somewhat better. This is due in part to the resolution of the data.
- In two cases, clouds were a problem. In one case, a cloud totally obscured a piece of land which resulted in its misclassification (should have been a stratum 11). In the second case, a fair sized island was missed due to a wispy cloud which totally obscured it. If the cloud is wispy, a different band combination may allow us to „see“ through it.

An analysis of the sample selection portion of CASS is currently being done. A comparison will be done using the segment size, the homogeneity of the PSUs, and the type of boundaries used. The important outcome was the distinction (as to the software) between the stratification and sample selection processes. The first attempts at sample selection proved to be fairly time consuming due to display difficulties. This can be improved by working with smaller TM data files (rather than using a full LANDSAT scene), by automating the identification of the PSUs location, and by working at different zoom levels. The speed can also be improved in the long run through the graphical user interface (see the „Future Considerations“ section), and in the short term by modifying the commands available with the mouse, and those which must be keyed in on the menu terminal. Lastly, the resolution of the TM image can be improved by the bands which are selected, and through increased resolution from future satellites. One important note is that boundaries are more vital to this process since the polygons are much smaller.

3.7 Benefits of Using CASS

In any project, such as CASS, several components of quality can be addressed. These are accuracy, resources, timeliness, and relevance. (Statistical Policy Working Paper 20, 1991)

As to accuracy, several benefits come to mind.

- 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 (1:100000 rather than 1:250000) than the paper-based satellite imagery. This relates to better agricultural stratification.
- b) The tedious, error prone process of transferring from a satellite image print, to high altitude photography, to a 1:100000 USGS map, to a digital file can be eliminated.
- c) PSUs can more easily be revised in CASS, than on paper.

- d) Sample segments are chosen by the software using a random number generator instead of looking through a set of random number sheets.
- e) Several concerns are boundaries, urban stratification, and cloud cover. The DLG is only as good as the date it was created or revised. Boundaries on LANDSAT are harder to identify than boundaries on the photos. In cases where the black and white photography is more recent than the 1:100000 maps, changes in boundaries could be determined. Also, houses cannot be seen as well using the current LANDSAT data. Ways to improve on these are discussed in the „Future Considerations“ section.

Resources and Timeliness go together. With CASS, we could do more states with the same number of people, or do the same number of states with fewer people. NASS prefers the latter, as personnel can be reallocated to other areas where there is need.

- a) A reduction of staff can occur, as the labor intensive stratification process is automated on CASS.
- b) A reduction of staff can occur, as the digitization process (to calculate areas and serve as a back-up) is no longer required.
- c) In general, CASS takes less time than the current process. A county takes about 2 weeks to do on paper, but only 2-3 days to do the same process on CASS.
- d) The ability to immediately determine the PSU size should help reduce extra labor involved in the sample selection process.
- e) Cost is an important resource. The hardware costs will be large initially, and then decrease substantially. The material costs will increase due to the purchase of digital TM data rather than TM imagery. Salary costs should be much less due to the reduced staff and the shorter time involved. Details on this are not yet available.

Relevance refers to the use of frames and sample segments by the end-user.

- a) 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).
- b) The locations of sample segments can be identified (latitude and longitude) and read into a Geographic Information Systems (GIS) database.
- c) Allows easier exploration of specialized area frames. The Remote Sensing Section in NASS can provide crop-classified satellite imagery to assist in the development of specialized area frames.

3.8 Future Considerations

There are several ways in which we hope to improve CASS. This involves the stratification process, the boundary determination, and other possibilities.

The stratification process may be improved in several ways.

- a) As of yet, only bands 2, 3 and 4 have been used for display purposes. By accessing other bands, we may see additional color detail, and even be able to see through clouds.
- b) We are looking into the use of unsupervised and supervised classification of satellite data and hope to use it as an aid to stratification.
- c) We are looking into using Census Bureau 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.
- d) A gentleman from the U. S. Department of Agriculture (USDA) is looking into a coordinated purchase of Landsat data by all agencies within the USDA. By sharing the data, all agencies should benefit, and CASS might benefit from possibly having more than one LANDSAT date per scene.

The boundary selection process is vital and may be improved in several ways.

- a) The use of other bands (especially band 5) should provide more detail.
- b) We are investigating the use of filters to improve the ability to detect „edges“ or boundaries.
- 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) The availability of 1:100000 DLG for administrative and political boundary data should be helpful. Also, „area“ records on the DLG file are not yet utilized, and could provide useful information.

A few other improvements are given.

- a) The NASA programmers are working on a graphical user interface for CASS, so that we can get rid of the „menu“ terminal, by displaying menus within „windows“.
- b) The polygon files created within CASS, have successfully been read into a GIS using ARC/INFO. An exchange of information the other way, may be possible.
- c) Latitude and longitude coordinates can be determined within CASS. This will help NASS to order photography more easily and accurately, and fill other potential needs.

Bibliography of Area Frame Development

- Ciancio, N.J., Rockwell, D.A., and Tortora, R.D. (1977): An Empirical Study of Area Frame Stratification. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.
- Cotter, Jim and Nealon, Jack (1987): Area Frame Design for Agricultural Surveys. U.S. Department of Agriculture, National Agricultural Statistics Service.
- Fecso, R. and Johnson, V. (1981): The New California Area Frame: A Statistical Study. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 22. Washington, D.C.
- Fecso, R., Tortora, R.D., and Vogel, F.A. (1986): Sampling Frames for Agriculture in the United States. Journal of Official Statistics. Vol. 2, No. 3. Statistics Sweden.
- Geuder, J. (1983): An Evaluation of SRS Area Sampling Frame Designs: Ordering Count Units and Creating Paper Strata. U.S. Department of Agriculture, Statistical Reporting Service Staff Report No. AGES830715. Washington, D.C.
- Geuder, J. (1984): Paper Stratification in SRS Area Sampling Frames. U.S. Department of Agriculture. Statistical Reporting Service, SF&SRB Staff Report No.79. Washington, D.C.
- Hanuschak, G. and Morrissey, K. (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 Staff Report. Washington, D.C.
- Houseman, E. (1975): Area Frame Sampling in Agriculture. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 20. Washington, D.C.
- Huddleston, H. (1976): A Training Course in Sampling Concepts for Agricultural Surveys. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 21. Washington, D.C.
- Jessen, R.J. (1942): Statistical Investigation of a Sample Survey for Obtaining Farm Facts, Iowa Agricultural Experiment Station Research Bulletin 304.
- Pratt, W. (1974): The Use of Interpenetrating Sampling in Area Frames. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.
- Statistical Policy Working Paper 20 (1991): Seminar on Quality of Federal Data. Federal Committee on Statistical Methodology Report. Washington D.C., pages 32-3.
- U.S. Department of Agriculture (1983): Scope and Methods of the Statistical Reporting Service. Publication No. 1308. Washington, D.C.
- U.S. Department of Agriculture (1984): Area Frame Analysis Package. Statistical Reporting Service Staff Report. Washington, D.C.
- U.S. Department of Agriculture (1987): International Training: Area Frame Development and Sampling. National Agricultural Statistics Service. Washington, D.C.

U.S. Department of Agriculture (1987): Supervising and Editing Manual, June Enumerative and Agricultural Surveys. National Agricultural Statistics Service. Washington, D.C.

Wigton, W. and Bormann, P. (1978): A Guide to Area Sampling Frame Construction Utilizing Satellite Imagery. U.S. Department of Agriculture. Statistical Reporting Service Staff Report. Washington, D.C.

~~A may allow us to „see“ through it. An analysis of the sample selection portion of CASS is currently being done. A comparison will be Area Sampling Frames Paper Technology Research Section Technology Research Section~~