PEER-REVIEWED ARTICLE

# An Image Analysis System to Develop Area Sampling Frames for Agricultural Surveys

**James J. Cotter and Catherine M. Tomczak**

## Abstract

*The National Agricultural Statistics Service (NASS) has been developing area sampling frames as a vehicle for conducting surveys to gather a variety of agricultural data nationwide. In 1987, NASS was awarded a National Aeronautics and Space Administration research grant to develop a digitally based system to automate this process which had been conducted using a labor-intensive, paper-based technique. This system, the Computer Aided Stratification and Sampling (CASS) system, was developed by NASS and the Ames Research Center Ecosystem Technology Branch, and has now been implemented into NASS' operational program. This paper discusses and compares the manual procedure and the new methodology and the results of this research effort.*

## Introduction

Since 1954, the National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) has been developing, using, and analyzing area sampling frames 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 (Cotter and Nealon, 1987). Statistical data developed by NASS on the Nation's agriculture are essential for the orderly development of production and marketing decisions by farmers, ranchers, and other agribusinesses. These agricultural data series are also used for monitoring the ever-changing agricultural sector and for making and carrying out agricultural policy relating to farm program legislation, commodity programs, agricultural research, agricultural chemical usage, rural development, and related activities.

An area frame for a land area, typically 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 delineated based on factors such as ownership, or based simply on easily identifiable boundaries as is done by NASS. Area frames are critical to producing quality estimates, as they provide complete coverage with all land areas being represented in a probability survey with a known (not necessarily equal) chance of selection.

States are selected to receive new area frames based on a statistical analysis of the survey results over time. The utilization of land is always changing. As a result, the frame "ages" and the survey statistics begin to decrease in reliability. Because only two new frames (on average) can be developed in a year, frames are prioritized according to need.

The manual procedure used to develop area frames was very labor intensive. The development of an area frame on paper-based materials for a typical or average state may require 11,000 hours and cost over $150,000. CASS has been responsible for cutting the time requirement by approximately half.

This paper will briefly describe the materials and procedures used in developing a paper-based area frame. This will then be followed by a description of the new automated procedures for developing area frames using digital inputs, which is now operational. For more information on area frame development, consult the authors.

## Paper-Based Area Sampling Frames

### Materials Used

Area frames are developed on a state by state basis. The materials used in the stratification process (see examples in Cotter and Nealon (1987)) include

*Satellite Imagery*: Historically, a paper-based image product from the Landsat satellite was used. Of the two types of scanners that are available, the Multispectral Scanner (MSS) and the Thematic Mapper (TM), the TM is the preferred product for stratification, though TM is more costly due to its better resolution. The paper TM product is scaled at 1:250,000.

*National Aerial Photography Program (NAPP)*: NAPP is the product of a consortium of federal agencies which use aerial photography. 1:40,000-scale, 9-inch contact prints are used. NAPP is a primary stratification tool. Nearly all (99 percent) of the U.S. has been photographed through the NAPP program.

*Topographic Quadrangle Map*: These maps are produced by the United States Geological Survey (USGS) and the preferred scale is 1:24,000 (7.5-minute series) which makes them beneficial for urban and ag-urban stratification and sampling.

*Bureau of Land Management Map*: These maps, scaled at 1:100,000, show the distribution of the federal and state land. They were useful in western states for delineating (public/private) range strata.

*USGS 1:100,000-Scale Topographic Map*: These high quality maps provide NASS with an accurate map base on which to work.

### Stratification

Satellite photo products and black-and-white aerial photography are used to identify land-use strata on the 1:100,000-

Photogrammetric Engineering & Remote Sensing,
Vol. 60, No. 3, March 1994, pp. 299–306.

0099-1112/94/6003–299$03.00/0

USDA-NASS, Room 4813 South Building, Washington, DC 20250.

scale map base. Table 1 displays the complete set of land-use categories which were used in the development of Missouri's area frame in 1987. The strata for general cropland can vary slightly depending on the amount of cultivation and agricultural activity in a state. This table also shows the target size range of first-stage sampling units called Primary Sampling Units (PSUs, discussed in the next section). The purpose of stratification is to reduce the sampling variability by creating homogeneous groups of sampling units. Although certain parts of the process are subjective, precision work is required of the personnel stratifying the land to ensure that overlaps and omissions of land area do not occur and that land is correctly assigned to land-use categories.

TABLE 1. LAND-USE STRATA CODES, DEFINITIONS, AND PRIMARY SAMPLING UNIT SIZES

| Stratum Code | Definition | PSU Size ($\text{miles}^2$) min. | target | max. |
|---|---|---|---|---|
| 11 | General Cropland, 75% or more cultivated. | 1 | 6-8 | 12 |
| 12 | General Cropland, 50-74% cultivated. | 1 | 6-8 | 12 |
| 20 | General Cropland, 15-49% cultivated. | 1 | 6-8 | 12 |
| 31 | Ag-Urban, less than 15% cultivated, more than 100 dwellings per square mile, residential mixed with agriculture. | 0.25 | 1-2 | 3 |
| 32 | Residential/Commercial, no cultivation, more than 100 dwellings per square mile. | 0.1 | 0.5-1 | 1 |
| 40 | Range and Pasture, less than 15% cultivated. | 2 | 12-16 | 24 |
| 50 | Non-agricultural, variable size. | 1 | n/a | n/a |
| 62 | Water | 1 | n/a | n/a |

Initial training of personnel stresses the need to use quality boundaries. A quality boundary is a permanent or, at least, long-lasting geographic feature easily located on the ground by a field interviewer. If an interviewer cannot accurately locate the sampled area, or 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 nonsampling errors will occur. This will result in estimates of population characteristics which are biased and do not truly represent the population of interest.

When the objective of using permanent boundaries conflicts in actual practice with the objective of obtaining homogeneous sampling units, permanent boundaries take precedence. For example, cultivated fields can be located at the base of a mountain but a good boundary does not exist to include these fields in the appropriate strata, and, therefore, they may be placed in a <15 percent cultivated strata with better boundaries. Roads and rivers make good strata boundaries, while intermittent streams and field edges do not and should rarely be used. The geographic features most frequently used for strata boundaries ranked from highest to lowest quality are

- paved highways,
- secondary all-weather roads,
- local farm to market roads,
- railroads, and
- permanent rivers and streams.

The stratification is performed on a county by county basis for administrative purposes. Each stratification analyst works a 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 TM satellite imagery. The imagery is used primarily to ascertain where the cultivated and non-cultivated areas are present in a county. TM imagery, which usually covers the most recent growing season, is the best medium for land-use stratification. Using TM data for locating crops and pasture and photography for locating boundaries, the analyst must make subjective decisions on placing areas in appropriate strata on the photography. These decisions on assigning land to strata are based on experience and training in photointerpretation.

After stratification on photography has been reviewed and approved, strata boundaries are transferred to a map base (also called the frame map). Once this transfer is completed, the next phase of stratification begins - construction of primary sampling units.

### Construction of Primary Sampling Units

Rather than dividing an entire frame into final sampling units, called segments, strata are divided into PSUs. A random sample of PSUs is further divided into segments, resulting in a tremendous savings in labor costs. Segments will eventually be visited by an interviewer to gather agricultural information.

The desired size of the PSU varies by strata, but averages six to eight segments. Because a PSU is a collection of segments, the minimum PSU size is, by definition, one segment (see Table 1). In delineating PSUs, the main focus is not homogeneity of land use; that has already been accomplished with land-use stratification. Rather, the main concern is to achieve a desired size with good boundaries while trying to maintain that each PSU is a smaller representation of the stratum as a whole.

Completed frame maps are then reviewed, and PSUs are examined for closure. The numbering system is checked for strata identification accuracy and sequential accuracy. Frame maps are further checked to ensure that omissions and overlaps do not exist. Once these checks have been accomplished, frame maps are ready for the next step in the process - measuring the size of the PSUs.

### Digitization

The 1:100,000-scale frame maps are digitized in order to

- measure PSUs accurately for subsequent sampling,
- ensure quality, and
- retain a digital backup copy of the frame map.

Using the map scale, the area of each PSU in a county is calculated in terms of square miles and is then stored in a file for that county.

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. County area is allowed more variance because of the smaller area involved and because PSUs are allowed to cross county boundaries. Because estimates of agricultural data items are made for the state, stratification is never allowed to cross state boundary lines. Therefore, only a small amount of error is allowed at the state level.

PSU areas are then accumulated for each stratum at the state level. The PSU area (e.g., 10.5 miles) divided by the target segment size for the stratum (e.g., 2.0 miles) is equal to the total number of segments in that PSU rounded to the nearest integer (e.g., 5). 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 segments to be sampled for the entire state. Using Missouri as an example, 387 segments are sampled from a population of 58,080.

### Sample Selection

After the total number of sample segments to be used in a state has been determined, a separate program is run to select PSUs which will be further broken down into sample segments. The PSUs are selected with probability proportional to their size. This is the first stage of sampling. Selected PSUs are located on the frame map and their boundaries are then transferred to photography. The selected PSU is then divided along identifiable boundaries into the required number of segments. Each segment has a specific target size (see the minimum PSU size in Table 1) depending on the stratum with which it is associated such that each individual segment closely resembles the full PSU (as much as possible) with the best physical boundaries available. Segments are manually numbered, and a random number is chosen to select the sample segment with equal probability. This completes the second stage of sampling. For information on the statistical formulas involved in the sample selection, contact the authors.

### Sample Preparation

After the segment has been randomly chosen within the PSU, sample preparation (the last step) takes place. The sampled segment is located and identified on a map for use by the field enumerator. In addition, the most recent photo coverage of the segment is ordered as an enlargement from the Agricultural Stabilization and Conservation Service (ASCS), U.S. Department of Agriculture. The enlargement is obtained to facilitate data collection activities such as delineating crop fields and locating farmsteads. Identification information such as county name and segment number are scribed onto the enlargement prior to being mailed to the State Statistical Office.

## Digital-Based Area Sampling Frames

### Research Background

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 Multispectral Scanner data and associated ground data, was written by NASS and ECOSAT personnel. Eighty percent of PEDITOR, a portable version of EDITOR, is currently the primary software tool for remote sensing operational work in NASS. PEDITOR was written at ECOSAT. ECOSAT assembled a prototype microprocessor-based workstation, called MIDAS, for NASS and assisted with an experiment to determine performance characteristics of the system when generating area estimates in the NASS operational environment. ECOSAT also created display software for NASS that is compatible with PEDITOR (not portable). This was the precursor for the system required for area frame development.

In an effort to automate and improve the process in which area frames are developed using digital inputs, NASS entered into a cooperative agreement with the National Aeronautics and Space Administration (NASA). The project with NASA began in 1988 with a NASA Research Grant from the Earth Observation Commercial/Applications Program. Although the initial research agreement with NASA expired in the fall of 1991, NASA continues to provide software support through a cooperative agreement with the Ecosystem Science and Technology Branch (ECOSAT), Ames Research Center, Moffett Field, California. The result of this NASS/NASA project is a new area frame system called the Computer Aided Stratification and Sampling (CASS) system.

### The CASS Workstation

Initially, several configurations of equipment were considered. The resulting system is shown in Figure 1.

A UNIX-based Hewlett-Packard (HP) workstation is being used to handle data processing and storage requirements. The HP workstations possess the minimum capabilities for area frame development, that is, three image planes, four overlay planes, and a 1024 by 1280 display coordinate system. The image planes display three bands of satellite data using 24 bits. The graphics overlay planes are used for various purposes, such as displaying digital road, water, and county boundary data; PSUs for a county and its neighbor; and the command menu (small window overlaying the image on the display terminal).

A graphical user interface (GUI) has been written under the X Window System and is currently being tested. This interface will make the software easier to use and allow the flexibility to handle changing hardware and software technology. An image is reduced to 1024 by 1024, allowing the remaining area on the graphics terminal to display the menus.

### The CASS System

The resulting CASS software developed by NASS/NASA is an image analysis system which incorporates digital imagery and digital line graph data. TM data (1:100,000 scale, 30-metre resolution) from the EOSAT Corporation (see Figure 2a)

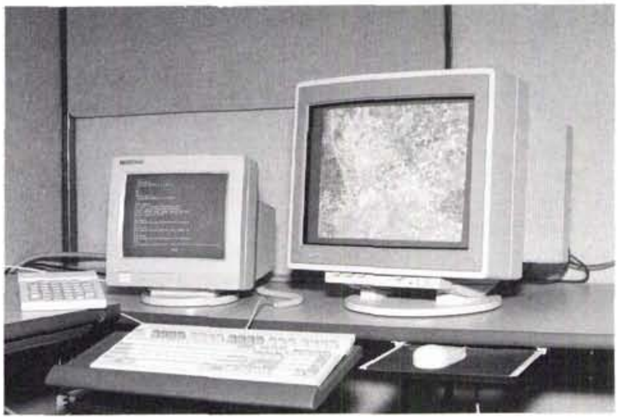
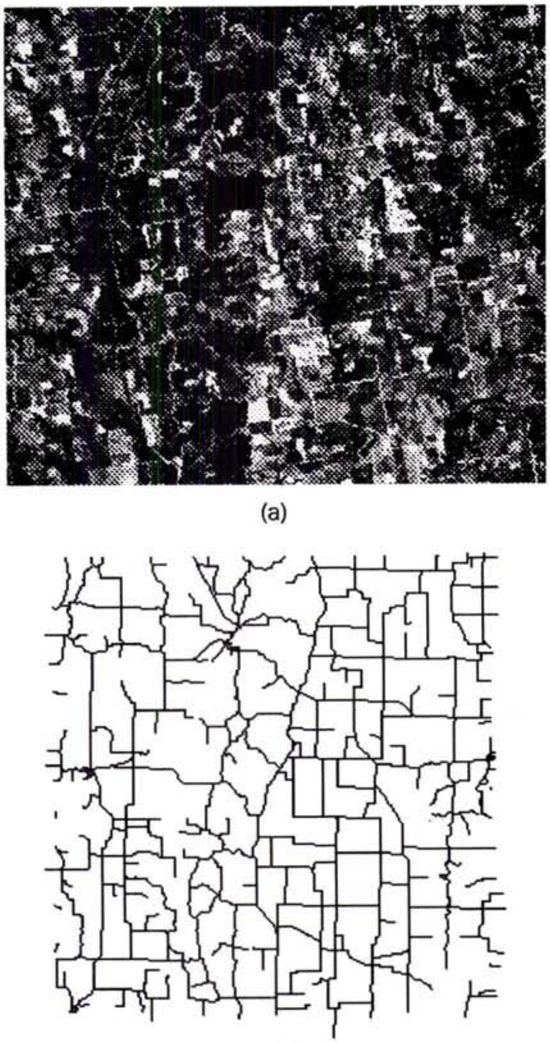


Figure 1. CASS workstation.


(a)

(b)

Figure 2. (a) TM image—Macon County Missouri. (b) DLG road data.

serves as a base to delineate land use according to the stratification scheme. U. S. Geological Survey's Digital Line Graph (DLG) data at a 1:100,000 scale is used for boundary identification, by overlaying onto the digital image using a graphics plane. The reason for using a visual approach to land-use classification, which is subjective, rather than a supervised classification approach, which is repeatable, is that good physical boundaries are required. Also, multitemporal and ground truth data are too expensive.

*Displaying and Coloring Satellite Data* — TM bands two, three, and four were used for optimal agricultural land-use classification. The digital nature of the data enables the analyst to create a color map which best distinguishes cultivation and boundaries.

*Display and Registration of DLG Data* — U. S. Geological Survey's transportation and hydrography DLG data (see roads in Figure 2b) are used, as well as political boundary data from the Census Bureau's digital map database (TIGER), the Topologically Integrated Geographic Encoding and Referencing System. The 1:100,000-scale DLG boundary data are not complete, and 1:2 million-scale DLG boundary data are too inaccurate for NASS's needs. Because the TM data are more recent than the DLG data, DLG data are used mainly as a reference. If the analyst can identify a physical boundary directly in the TM image, it is used. If a physical boundary can not be distinguished in the TM image, but a DLG boundary exists, it is used. The TM image (scene) is precisely overlayed with DLG, by registering the DLG data to the backdrop of satellite data. Several matching TM and DLG points are selected, and a least-squares regression is run to fit the remainder of the data. These points and the regression are saved in a file and used each time a DLG file for that scene is displayed. This registration file also enables the user to determine latitude and longitude coordinates of a given point.

*Other Inputs* — Other reference materials include 1:100,000-scale U.S. Geological Survey maps, some small scale aerial photography, topographic quadrangle maps for city areas, information from NASS State Offices, and information on planting and harvesting dates for the major crops in that state.

*PSU Delineation* — In each county, polygons are drawn and tagged with the appropriate PSU number, which consists of a stratum number and a sequence number (see Figure 3a). This is done by determining the particular stratum in which to place a unit of land, by interpreting the color TM image. At the same time, a PSU within some specific size range is delineated (refer to Table 1), using physical boundaries identified by DLG and/or TM data. In CASS, this is done by keying in a PSU number, and then utilizing the mouse to pick points along desired boundaries. When a PSU (or polygon) is closed, the area is immediately calculated and displayed. This allows the user to determine if the PSU is within the target size for that stratum. If polygons are too small or too large, they can be combined, split, or reshaped. When a county is completed, the polygons are saved to a file to be reviewed by another experienced analyst. The analyst has the ability to check for overlapping polygons and holes (or missing land areas). At any time, he/she can list PSUs that have been created, to check for proper PSU numbering and that PSU areas are within tolerance. Refer to Cheng *et al.* (1989) and Cheng *et al.* (1992) for further information.

*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 (outside of CASS) to draw a sample of PSUs which will be further divided into segments. Only those PSUs which were chosen by the sample select program are divided in CASS. The user displays the file (saved in the previous step) and enters in the PSU number to be sampled. Software then erases all but the sample PSU from the screen. Many of the same functions which were involved in delineating PSUs during the stratification phase are used to divide the PSU into equal size segments. For example, the mouse is used to pick points along an identifiable boundary. When the segment (polygon) is closed, the size is immediately displayed, and segments can be merged or split, or boundaries can be reshaped. Similar quality control checks for overlaps and omissions are done. Because segment areas are much smaller (typically one square mile), boundaries are harder to find. Occasionally, field edges, section lines, or point to point must be used. When the PSU has been completely divided into segments, one is selected randomly using the segment selection command (see segment number 01992 in Figure 3b). Its latitude and longitude are then deter-

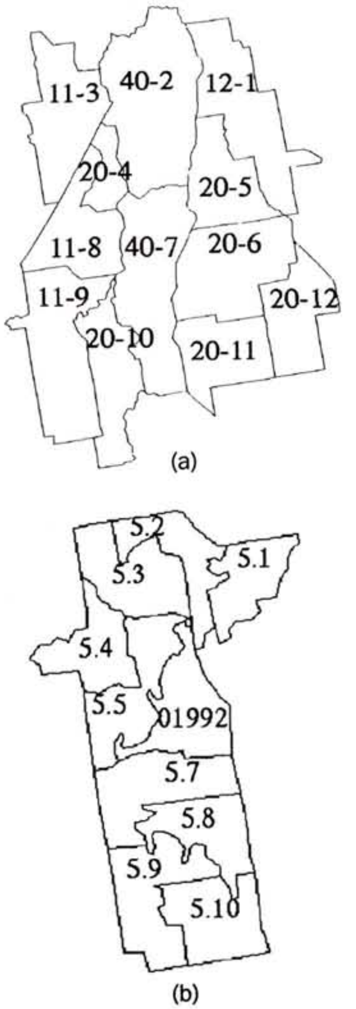


Figure 3. (a) PSU delinieation. (b) PSU divided into segments.

mined (for use in the NASS Geographic Information System), and a photo enlargement is ordered. Lastly, the boundaries for sample segments are transferred to enlarged photos by the Sample Preparation Unit (see Figure 4) by displaying the segment, or utilizing a print of the digital image and segment boundary.

## Research

Once the initial CASS system was developed (hardware and software), two pilot tests were conducted in portions of Missouri (to see if CASS was possible) and Michigan (to work out the "bugs"). Enhancements were made as a result of these tests.

### *Missouri*

The purpose of the initial test (conducted in 1988-1989) was to gain basic experience with the software, compare CASS to the paper-based method, and determine the speed of frame construction. 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 of NASS had developed a new area frame for Missouri in 1987 for use in 1988.

This test proved that stratification using CASS was possible and reasonable. CASS proved to be faster, as two weeks were required for paper stratification and digitization, and three days for CASS. However, this did not reflect setup time or problem resolution.

Because this was the first test, analysis was limited to comparing CASS to the paper frame. The results illustrate the subjective nature of the work, as the photointerpretation and boundary selection process (using both small scale paper and digital image products) is a difficult one. In this test, five different people stratified each of the three counties. Two people used the paper method, and three used CASS. They are represented in Figure 5 as paper1, paper2, cass1, cass2, and cass3. The percent of the total three-county area (excluding the urban, water, and non-cultivated strata) is given by strata, by person. That is, the percent land area in the >75 percent cultivated strata ranged from 23 percent (cass3) to 37 percent (cass1).

### *Michigan*

With the completion of the Missouri test, a more substantial test was done to mimic a pseudo-operational environment. A 21-county area in Michigan was selected for this purpose. At the time of the study (1989-1990), the state had just received a new frame in 1989 (implemented in 1990). Also, the Remote Sensing Section of NASS had recently completed work in the dry bean area of Michigan in regards to supervised classification (therefore, TM data were available). In this test, only one person worked each county, because this was typical of how the system was to be used operationally. Results of stratification in CASS were influenced by the recency of the paper stratification, because an analyst doing CASS strati-

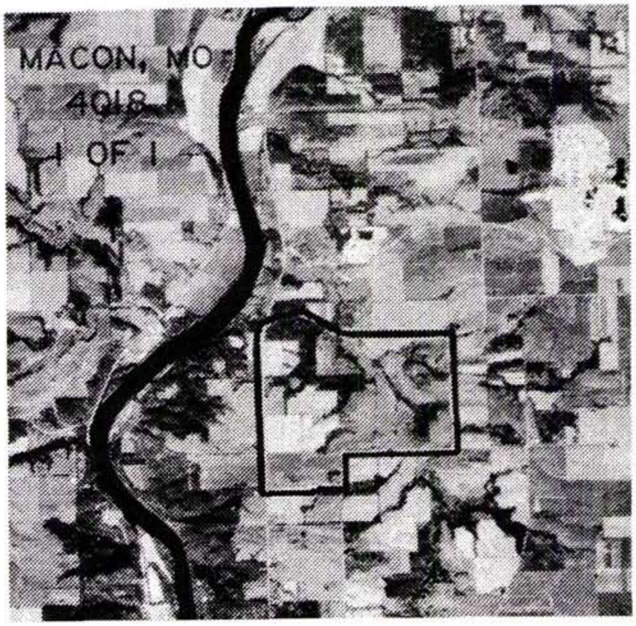


Figure 4. Sample segment on photo enlargement.

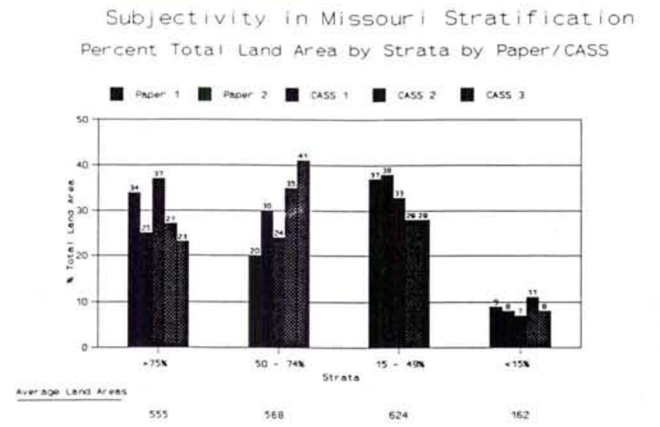


Figure 5.

fication might recall what occurred in the paper stratification.

As before, an analysis of this 21-county area (both quantitative and qualitative) was limited to comparing CASS to the paper frame. The total area in each land-use stratum was measured, and the percentage difference between paper and CASS (given as CASS-paper/paper) was calculated. This is shown in Figure 6.

- As to the 14.6 percent difference in the 50 to 74 percent cultivated stratum, the analysts generally favored the CASS results, because they could better identify pasture, which is not considered cultivated land.
- Concerning the 37.8 percent difference in the <15 percent cultivated stratum, the analysts also favored the CASS results, because they felt they were better able to identify and include the woodland areas.
- Finally, the 14.3 percent difference in Ag-Urban stratum was mostly due to the lower resolution of the TM data compared to aerial photography and quadrangle maps. The analysts could identify density of houses on the aerial photography (somewhat), and could easily locate road boundaries on the quadrangle maps. However, density of roads had to be used in some suburban areas on the TM, and the older age of the DLG compared to the TM, explained problems identifying road boundaries where new developments had been built. Also, the average PSU size in an Ag-Urban stratum was larger in CASS (1.6 square miles) than on paper (1.2 square miles).

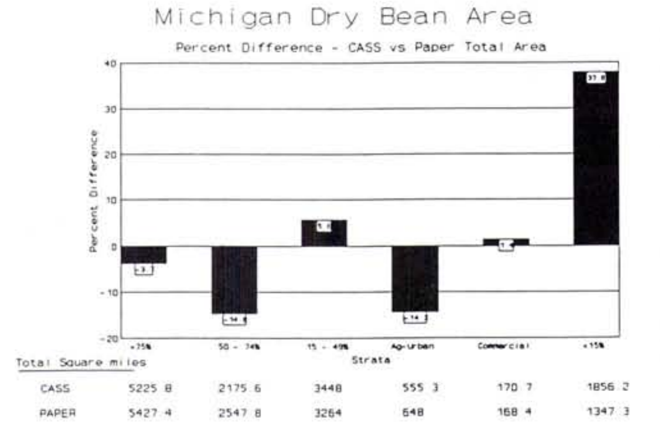


Figure 6.

In Michigan (as in Missouri), no ground truth verification was sought. With the experience of the analysts, and the high cost in obtaining sufficient data, emphasis was on developing CASS to make it operational.

Each of the five analysts visually compared the CASS and paper area frames for a given county, not to measure quantitative differences, but to see generally where differences occurred, and why. This qualitative aspect of the study is worth noting.

In general, because of the increased resolution of digital TM versus the paper TM product previously used, and the dynamic rather than fixed colormapping abilities (many more colors were visible), the CASS-produced stratifications were found to be more accurate than those developed from paper imagery. However, the stratification of urban areas was less accurate in CASS, because digital TM (approximately 30-metre resolution, 1:100,000 scale) proved more difficult to interpret than the aerial photography (approximately 2-metre resolution, 1:40,000 scale) and quadrangle maps (1:24,000 scale) previously used. For this same reason, boundary selection in CASS is more difficult. The topographic maps provided the user with detailed information on boundary type, which was lost when displayed in an overlay plane using limited line types.

In two cases, clouds were a problem. In one case, a cloud totally obscured a piece of land, resulting in its misclassification. In the second case, an island was obscured and missed by a wispy cloud.

It was during the Michigan test that the first counties which crossed TM scenes and Universal Trans Mercator (UTM) zones were encountered. Before, county polygon files were stored in row and column format (within a scene). In order to keep these files displayed in the correct part of neighboring scenes (e.g., southern tip of one scene, and northern tip of the neighboring scene), they were converted into UTM coordinates. For the same reason, special procedures were necessary to handle counties which overlapped two UTM zones. This complicated the process substantially.

### Operational Phase

Following the completion of the Michigan test, CASS became operational. Since then, two states' area frames have been developed in CASS: Oklahoma and California. Oklahoma's new area frame was used for enumeration purposes beginning in June 1993, and California's new area frame will be used for enumeration purposes starting in June 1994.

The most valuable lesson learned during this time was proper registration of the DLG to the TM scene. Initially, control points from the four TM scene corner points were used. These proved to be too inaccurate. The main problem occurred where TM scenes overlapped, that is, a given county is covered by two different scenes but not entirely by either one. The difficulty was then in getting the two TM scene registrations to "agree," each with different number and spatial placement of control points. Using trial and error on difficult situations, the best results were obtained when placing a substantial number of points, at equal spacing. By placing points as close to the scene edge as possible, the problem of a given county having a control point in one scene but not in the other scene is diminished. Problems with UTM zone changes within a TM scene were solved by locating control points throughout the entire scene (e.g., UTM zones 14 and 15), but reordering so that the first control point was in the zone of interest (e.g., zone 15). Thus, two registrations (one for each zone) would exist for that scene. At one point, Census TIGER data were tested, because it was more recent than DLG data. It proved to be less geopositionally accurate in some areas (critical to NASS work), so DLG was used.

Two problem areas had previously been identified: clouds and urban areas. When obtaining TM data, a compromise must sometimes occur between the most recent date (preferably in the last year) and the cloud cover. Also, spring or fall months are critical, when cultivation can best be identified (planting or harvesting has occurred, but trees are not at peak foliage). Therefore, cloud coverage under 10 percent is accepted, though it depends on the clouds' density and location. Aerial photography is ordered to supplement the TM in clouded and urban areas (large cities). With segment sizes of one-tenth and one-quarter square miles, sample selection for Commercial and Ag-Urban strata is done on quadrangle maps, rather than in CASS.

Also, the Sample Preparation Unit had a few problems transferring segment boundaries onto photo enlargements. For the first time, the TM imagery was more recent than the enlargements. This more accurately reflected what existed on the ground, but the segment boundaries had to be drawn on the older photo enlargements. In some cases, problem boundaries were selected due to the resolution of TM being lower than aerial photography, and the difficulty in distinguishing DLG line type (much easier on paper 1:100,000-scale maps). For example, intermittent streams in Oklahoma were used as boundaries, which did not show up in the photos, and may or may not show up on the ground. For the first time, latitude and longitude coordinates were determined for each sample segment.

## Resource Considerations

The original purpose in developing CASS was to improve NASS area frames (discussed above), and use resources more efficiently. In order to continue using CASS operationally, management had to be shown that resources were being spent wisely. The resource considerations for CASS can be broken down into several categories: materials, labor/staffing, hardware, and software.

### *Materials*

Material costs are significantly higher using CASS. The cost of digital TM data is approximately four times greater per scene than the paper product. In the past, materials averaged 20 percent of the total cost of creating a new state frame. In Oklahoma, materials were 58 percent (see Table 2). However, this could change in the near future, depending on the cost of satellite data.

### *Labor/Staffing*

- Stratification using all paper products was very labor intensive. Users would analyze TM imagery, stratify land area, transfer boundaries twice, and then digitize the boundaries. With CASS, digitization is built in and boundary transfers are eliminated, resulting in excellent labor savings.
- In the past, labor was about 80 percent of the total cost of producing a new state frame. In Oklahoma, labor was 42 percent (see Table 2).
- Not only are the total number of hours important, but also the personnel issues and changes are important. Staff has changed from more part-time, to fewer full-time employees. This results in a higher per-person cost, but total hours are greatly reduced (approximately half). Also, the full-time staff will result in less turnover, thereby retaining experience and reducing training costs for new employees.

### *Hardware*

There are one-time and maintenance charges. Hardware startup costs were partially offset by the NASA grant, and can be amortized over the life of the system. The systems will need to be upgraded or replaced over time, but the present trend in workstation prices is down while the amount of computing power per dollar is rising. At this time, a suitable HP workstation runs about $20,000 for the general public.

### *Software*

The software was developed, corrected, and enhanced with funds from the NASA grant, and from NASS contributions. These resources will decrease in the very near future to an amount needed only to maintain the software. Because the software was developed with government funding, it is public domain software. However, no software support mechanism is currently available from NASS or ECOSAT.

Changes in technology continue to have an effect on the CASS system (hardware and software). As machines phase in to X Windows, CASS must be flexible enough to incorporate this new technology.

## Summary

Overall, the advantages and disadvantages of CASS must be considered in order to determine its success.

### *Advantages of CASS*

The analysts generally agree that stratification in CASS is better for several reasons. First, satellite data provide more recent data (potentially available every 16 days) than aerial photography (may be five years old). Because an area frame is used for about 15 years, the most recent imagery at the

TABLE 2. COMPARISON OF COST AND HOURS FOR RECENTLY DEVELOPED AREA SAMPLING FRAMES

| | # counties | sq. miles | year | % labor | % mater. | total cost | total hours | |
|---|---|---|---|---|---|---|---|---|
| Oklahoma | 77 | 69,067 | 93 | 42 | 58 | 153,913 | 4,459 | CASS |
| Arkansas | 75 | 52,482 | 92 | 79 | 21 | 137,937 | 10,193 | paper |
| Georgia | 159 | 58,334 | 91 | 81 | 19 | 183,726 | 14,927 | paper |
| Alabama | 67 | 51,014 | 91 | 85 | 15 | 140,646 | 11,460 | paper |
| Michigan | 83 | 57,448 | 90 | 76 | 24 | 134,359 | 10,459 | paper |
| Louisiana | 64 | 44,384 | 90 | 81 | 19 | 122,128 | 10,050 | paper |

time of stratification is desired. Second, the land-use determination is more accurate, as the scale of TM data has gone from 1:250,000 on paper, to 1:100,000-scale digital data. Finally, a dynamic color map is available to enhance the image and bring out the cultivation.

The automation of this process has eliminated the tedious, error prone process of transferring from a satellite image print, to aerial photography, to a 1:100,000-scale USGS map, to a digital file. Also, CASS uses computer-generated random numbers for segment selection. In the paper process, random numbers were printed on a sheet of paper, and left out on a desk. Some felt that the analysts could find out which random number was next, which might influence their segment boundary selection.

The digital nature of the data is a benefit. First, PSUs (and segments) can be more easily revised in CASS by moving digital boundaries, and because the size of the PSU (and segment) is known immediately, it can be resized if it does not fall within the suggested limits. Next, it will allow a frame to be updated rather than having to start from scratch (which is necessary with paper frames). Finally, sample segment locations are being identified. Because they are now georeferenced, they can be used as data layers in other Geographic Information Systems.

### *Concerns*

The identification of good physical boundaries (which must be located on the ground by enumerators) is more difficult on TM. The scale has gone from 1:40,000 on aerial photography to 1:100,000 on TM, and the resolution has decreased. Also, the DLG is not as clearly distinguished in an overlay plane as it is on paper.

Other concerns are urban stratification and cloud cover. Houses cannot be seen as well in TM data, so the analyst must resort to using such characteristics as density of roads. At this time, aerial photography is being used to supplement TM data in urban and clouded areas. Also, sample selection in urban areas is being done on quadrangle maps.

## Future Considerations

Much is possible in the future. The resource considerations of CASS (higher/lower costs) could change, depending on the future Landsat program. It is believed that costs would drop if the government were to take over control of the Landsat program.

Several of the concerns discussed above could be solved in the near future. The problem with urban stratification is being handled by using SPOT data in some of the larger city areas (e.g., New York City). It will enable the analyst to identify houses and distinguish among physical boundaries.

Several options exist in identifying potential segment boundaries. First, the use of edge enhancing filters will be tried to "bring out" boundaries in the TM data. Second, digital orthophotography will be used (when available) to aid in identifying potential segment boundaries. Third, future satellites may offer better resolution data at an affordable cost. Finally, updated DLG would be a tremendous help. In one area in California the TM data were from 1991, the aerial photography was from 1987, and the DLG data were from 1983.

At this time, the photo enlargement is the Area Frame Section's primary output. The data needed to purchase photos are located on microfiche. Each segment must be located on the microfiche to obtain this information. In the near future, photos may be ordered using latitude and longitude coordinates, because segment latitude and longitude is calculated in CASS. Also, when the resolution of the imagery data improves, clear prints of the segment may be generated directly from digital data (e.g., this is now possible with digital orthophotography).

As various state and federal agencies get involved with geographic information systems, agricultural data (such as on property classification) might be used to make a state's stratification more accurate. Also, the Remote Sensing Section of NASS could provide crop-classified satellite imagery to assist in the development of area frames.

Ground truthing for each CASS state will be limited to those sample segments visited by enumerators each June. Any state getting a new area frame will have an analysis done to see how that new frame affects agricultural statistics. Also, the percent cultivation found by the enumerator in each segment is compared to the stratum assigned during frame development.

## Conclusion

The analysts in the Area Frame Section who work daily on building area frames and selecting samples across the U.S. are proud of CASS, and feel it is a big success. The ability to stratify based on the amount of cultivation is much improved. The difficulties in dealing with urban stratification and boundaries (especially in sample selection) will be solved through the use of SPOT, digital orthophotography, and/or panchromatic satellite data.

Progress continues to be made and prospects for the future are encouraging. Reduction in costs and increases in efficiency will continue as more states are done in CASS, when NASS can order photography based on latitude and longitude, and when data resolution is improved enough to print photo enlargements. Also, TM data costs could drop, and DLG costs have already decreased.

The dramatic increase in the use of Geographic Information Systems is causing a demand for digital data. Two different environmental groups have already input digital area frames into ArcInfo using DLG format files. NASS has also started using digital area frames in their own GIS. Much will be determined by future technology and application trends.

## References

Cheng, T., G. Angelici, R. Slye, and M. Ma, 1989. *Computer-Aided Boundary Delineation of Agricultural Lands*, NASA TM-102243, NASA-Ames Research Center, Moffett Field, California, 23 p.

———, 1992. Interactive Boundary Delineation of Agricultural Lands Using Graphics Workstations, *Photogrammetric Engineering & Remote Sensing*, 58(10):1439-1443.

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