1993
93-11

THE AREA FRAME: A SAMPLING BASE FOR ESTABLISHMENT SURVEYS

Jeffrey Bush, National Agricultural Statistics Service

Carol House, National Agricultural Statistics Service

Area frame methodology has formed one of the cornerstones of probability sampling for several decades. While area frames are frequently used in urban settings for household surveys and population censuses, those with a rural focus have proved valuable for targeting farm establishments to provide basic statistics on agriculture and ecological resources. Cotter and Nealon outline the advantages and disadvantages of area frame methodology. They state that area frames are highly versatile sampling frames providing statistically sound estimates based on complete coverage of land area. Although costly to build they are generally slow to become outdated. However, area frame sampling is generally less efficient than list sampling for targeting any individual item and is inadequate for estimating rare populations.

This paper describes four different area frame methodologies currently in use as a base for sampling in rural areas. These are: a) the area frame used by the United States Department of Agriculture's (USDA) National Agricultural Statistics Service; b) the area frame used by Statistics Canada for agricultural statistics; c) the hexagonal area frame used by the U.S. Environmental Protection Agency; and d) the area frame used by the USDA's Soil Conservation Service for the National Inventory Survey. The paper's greatest emphasis is on the NASS area frame. For this frame, the authors provide additional detail on area frame construction, sampling, data collection and estimation. The paper provides a profile of costs associated with these activities, as well as procedures to assess quality deterioration in an "aging" frame.

NASS AREA FRAME

The National Agricultural Statistics Service (NASS) is the major data collector for the U.S. Department of Agriculture. As such it has responsibility to provide timely and accurate estimates of crop acreages, livestock inventories, farm expenditures, farm labor and similar agricultural items. NASS also provides statistical and data collection services to other Federal and State agencies. They have used an area sampling frame extensively for over 30 years in the pursuit of these objectives. Area frame samples are used alone and in combination with list samples (multiple frame). NASS contacts approximately 50,000 farm establishments each year through their area frame sampling procedures.

This section updates the work of Cotter and Nealon as it describes the procedures used by NASS to construct area frames and sample from them. It discusses data collection procedures, estimators, and costs associated with these different activities. Finally it discusses methods to objectively assess the "aging" of an area frame.

Area Frame Construction and Sampling

NASS constructs area frames separately by state and maintains one for every state except Alaska. Generally, two new frames are constructed each year to replace out-dated ones. The most recent frame construction was for Oklahoma. It became operational in June 1993. This frame will be used as the "example" throughout this paper.

Frame construction produces a complete listing of parcels of land, averaging six to eight square miles in size, throughout a state. These parcels serve as the primary sampling units (PSUs) in a two stage design, and each contain a varying number of population units or segments. The sampling process selects PSUs, and only those selected PSUs are broken down into segments. This two stage process saves considerable time and money over that required to break the entire land area into segments.

Construction of and sampling from an area frame involves five basic steps: 1) determining specifications for the frame; 2) stratifying the land area and delineating PSUs within each stratum, 3) allocating stratum level optimal sample sizes; 4) creating sub-strata and selecting PSUs; and 5) selecting segments within PSUs. Each step is discussed in detail below.

Frame Specifications

The specifications for building an area frame consist of strata definitions and target sizes for both PSUs and segments within each stratum. Statisticians define these by examining previous survey data, and assessing urbanization and other trends in the state's agriculture. Table 2-1 lists the frame specifications for the Oklahoma frame.

Table 2-1: Oklahoma Frame Specifications

Stratum	Definition	Primary Sampling Unit Size			Segment Size
		Minimum	Desired	Maximum	
		(sq. miles)	(sq. miles)	(sq. miles)	(sq. miles)
11	>75% CULTIVATED	1.00	6.0 - 8.0	12.0	1.00
12	51-75% CULTIVATED	1.00	6.0 - 8.0	12.0	1.00
20	15-50% CULTIVATED	1.00	6.0 - 8.0	12.0	1.00
31	AGRI-URBAN:>100 HOME/SQMI	0.25	1.0 - 2.0	3.0	0.25
32	COMMERCIAL:>100 HOME/SQMI	0.10	0.5 - 1.0	1.0	0.10
40	<15% CULTIVATED	3.00	18.0 - 24.0	36.0	3.00
50	NON-AGRICULTURAL	1.00	none	50.0	pps[1]
62	WATER	1.00	none	none	not sampled

1. Segments are selected with probability proportional to size; i.e. PSU's are treated as segments.

Strata are based on general land usage. A typical NASS area frame employs one or more strata for land in intensive agricultural (50 percent or more cultivated), extensive agricultural (15 to 50 percent cultivated), and range land (less than 15 percent cultivated). Less frequently an area frame contains "crop specific" strata. This occurs when a high percentage of the land in a state is dedicated to the production of a specific type of crop, such as citrus in Florida. In addition, each area frame uses an agri-urban and commercial stratum (more than 100 homes per square mile) plus a non-agricultural stratum including such entities as military bases, airports, and wildlife reserves. Finally, large bodies of water are separated into a water stratum.

Boundary points for agricultural strata are generally restricted to a set of standardized breaks: 15, 25, 50 and 75 percent cultivated. To determine the exact breaks for a given state, the percent cultivated for each segment sampled under the old frame is calculated from survey data. The resulting distribution is examined using the cumulative square root of frequency rule proposed by Dalenius and Hodges. The standardized breaks may be collapsed or expanded based on the structure of the distribution.

Two criteria are the most important for determining target sizes for PSUs and segments within strata: availability of good natural boundaries and the expected number of farm establishments. Generally, a lack of good boundaries will prompt the use of larger target sizes, while a large expected number of farm establishments will prompt smaller target sizes.

Stratification and Delineation of PSUs

Once strata definitions are set, the stratification process divides the land area of the state into PSUs and assigns each to the appropriate land use strata. Each PSU must conform to the definition and target size outlined for its particular stratum. PSU boundaries become a permanent part of the area frame and must be identifiable for the life of the frame. Thus the stratifier uses only the most permanent boundaries available when drawing off PSUs. Acceptable boundaries include permanent roads, rivers, and railroads. The final product of the stratification process is a "frame" file which contains a record for every PSU in a state. Specifically, each record includes the PSU number, stratum assignment, county, and size. This frame file is maintained over the life of a frame as the sampling base.

Prior to 1990 the process of stratification used paper maps, aerial photography, satellite imagery, and a considerable quantity of skilled labor. The end product was a frame delineated on paper U.S. Geological Survey 1:100,000 scale maps. In 1990, NASS implemented its Computer Aided Stratification and Sampling System (CASS). CASS automates the stratification steps on a graphical workstation using digital satellite imagery and line graph (road and waterway) data from the U.S. Geological Survey.

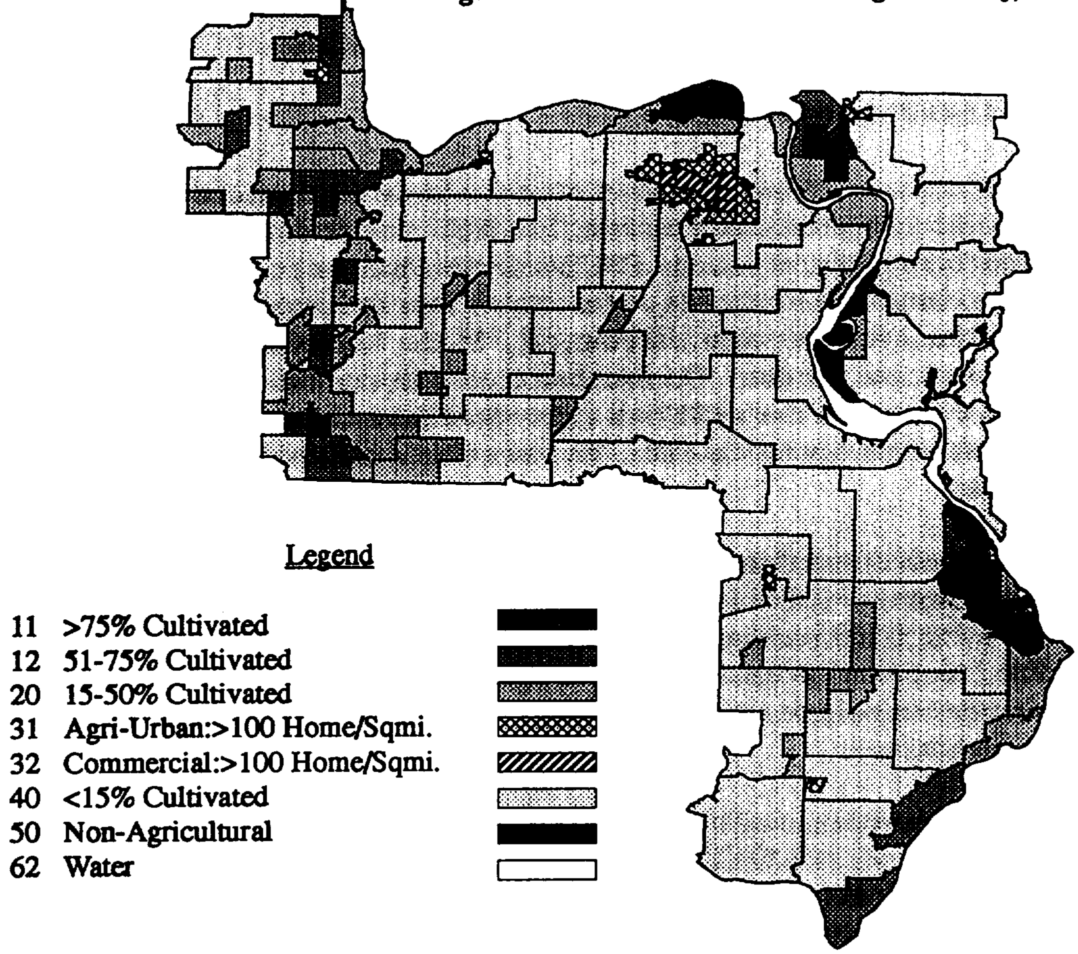
The digital satellite imagery employed by the CASS system is currently obtained from the thematic mapper (TM) sensor on the LANDSAT-5 satellite. The TM has a spatial resolution of 30 meters and is made up of 7 spectral bands. TM bands 1-5 and 7 reside in the reflective region of the spectrum while band 6 is located in the thermal infrared region. NASS experience with the imagery shows that bands 2, 3, and 4 highlight cultivated areas of land most accurately.

While TM data is very useful in providing information with respect to land usage, its large scale (30 meter resolution) renders it practically useless for identifying good PSU boundaries. Therefore the CASS system also uses digital files of U.S. Geological Survey 1:100,000 scale maps, in which feature class codes are assigned to all roads, water, railroads, power lines, and pipelines. The CASS system incorporates the road and waterway data from these files and overlays it on the TM imagery.

Personnel use a mouse and a "drawing" program to delineate boundaries of the PSUs and label them with their appropriate stratum number and sequence. As each PSU is completed, its size is immediately displayed. If the PSU does not fall within the particular target size, the stratifier immediately makes a correction. In addition, once a county has been completely divided into PSUs, the system can check for overlaps or omissions of land. Though the software provides many quality checks which save much time, reviews are still necessary to check the quality of stratification.

Figure 2-1 displays the stratification of Muskogee, OK which was performed on the CASS system. Notice the differing PSU sizes created with respect to each stratum.

Figure 2-1: Stratification of Muskogee County, OK



Sample Allocation, Sub-stratification and Sample Selection

The national sample size for the NASS frame is approximately 15,000. The two stage design selects 15,000 PSUs and then one segment per selected PSU. The sampling process is described in more detail below.

Following stratification a multivariate optimal allocation analysis is performed to allocate the first stage sample of PSUs between land use strata. This is a multivariate procedure because the area frame must target a variety of agricultural items. The analysis requires the following inputs: a) population counts of segments per stratum; b) estimated totals of important commodities from previous year's survey; c) standard deviations from previous survey derived by locating old segments in the new frame and strata; and d) target CV's for major commodities. The analysis produces stratum level sample sizes with expected coefficients of variation less than or equal to the target CV's.

At this point in the process, PSUs have been delineated and stratified according to their land use and the optimal number of PSUs to sample in each stratum has been determined. Next, each land use strata is further divided into sub-strata based, in part, on a criteria of agricultural similarity. This process improves the precision of estimates of individual commodities and facilitates sampling by replication. PSUs are grouped by county, and ordered within counties in a serpentine pattern starting in the North East corner. Counties are then ordered based on results of a clustering algorithm that groups counties with similar crop production. Together these steps produce an ordering of all PSUs throughout the state. Figure 2-2 displays the county ordering for the state of Oklahoma. Substrata then equally divide the ordered PSUs within each stratum, where one PSU is selected per replicate per sub-stratum.

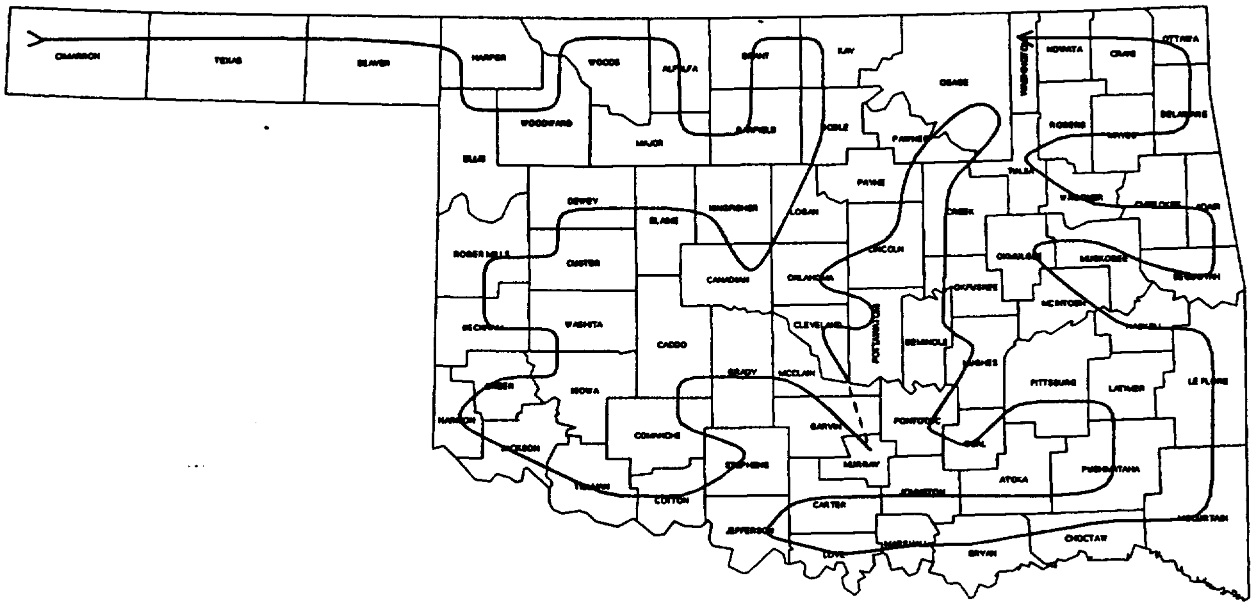


Figure 2-2: Oklahoma County Ordering

The sampled PSUs in each sub-stratum are randomly selected with probability proportional to the expected number of segments they contain and are assigned to a replicate. In non-agricultural and some range strata, where the lack of suitable boundaries is a problem, the PSUs themselves also serve as segments.

Replicated sampling has several advantages. First it facilitates sample rotation. Twenty percent of the sample in the NASS frame is rotated each year. Second, it allows estimation of year to year change from the 80 percent of the sample that did not change. Third, it simplifies the process of adjusting sample sizes to improve sampling efficiency.
The PSUs selected in the sample selection program are then located and further broken down into segments. The CASS system is used for this procedure as well. Just as the PSUs were originally delineated, so are the segments within each chosen PSU. Segments must be constructed using permanent boundaries, contain similar amounts of cultivation, and be equally sized. The CASS system randomly chooses one of the segments to sample from each PSU.

For data collection, segment boundaries are transferred to large scale NAPP photography. This process is completed by hand.

Data Collection and Estimation

Data Collection

NASS conducts a major area frame survey once each year in June. It is called the June Agricultural Survey (JAS). Approximately 15,000 segments are enumerated and yield approximately 50,000 farm establishments. These segments account for roughly 0.8 percent of the total land area of the 48 conterminous states. The survey produces estimates of crop acreages, grain stocks, number of and land in farms, livestock inventories, farm labor, and cash receipts at state, regional, and national levels. Importantly, the information collected during this survey provides a database of information about the farm establishments sampled through the area frame. This information is used as a sampling base for follow-on surveys for the remainder of the year. In particular, farm establishments are checked against the NASS list sampling frame to measure the incompleteness of that frame. The follow-on surveys generally use multiple-frame methodology, incorporating list samples with an area sample which account for this incompleteness.

Prior to the JAS data collection period, newly rotated segments along with residential, commercial, and non-agricultural segments are screened for the presence of farm establishments. States implementing a new area frame must screen all segments. Screening usually takes place in late April to early May. A questionnaire is filled out for segments which contain no agriculture.

The data collection period for the June Agricultural Survey begins June 1 and continues for two weeks. Enumerators conduct face-to-face interviews with operators of all farm establishments with land inside a segment, and account for all land within the segment. Enumerators are assigned anywhere from 8 to 15 segments to survey depending on distance between segments and the enumerator's experience level.

Estimation

In the NASS area frame, recall that segments are the population units and the second stage sampling units. The reporting units are the individual farm establishments within the segments. However, depending on the estimator that is used, these reporting units are defined somewhat differently. Sometimes the establishment reports only for its land *contained within the segment.* That part of a segment operated by a single establishment is referred to as a "tract." For other estimators, the establishment reports information for its entire operation. Other times only farm establishments whose operator lives inside the segment report information.

Three different estimators for summarizing area frame data are described below. Each has different advantages and disadvantages. Each may be used alone to estimate agricultural items, or in conjunction with a list frame to estimate for the undercoverage of that list.

Closed Estimator

The closed estimator simply sums data associated with all land *within the segment boundaries*, and expands these "segment totals" to represent the population. A state level sample estimate using the closed estimator may be expressed mathematically as follows:

$$\hat{Y}_c = \sum_{i=1}^{L} \sum_{j=1}^{s_i} \sum_{k=1}^{r_{ij}} y_{ijk}$$

where

$$y_{ijk} = \begin{cases} e_{ijk} \sum_{l=1}^{f_{ijk}} t_{ijkl}, & \text{if } f_{ijk} > 0, \\ 0 & \text{if } f_{ijk} = 0, \end{cases}$$

t_{ijkl} = the value of the survey item on the total *tract* acres operated for the l^{th} tract operation in the k^{th} segment, j^{th} sub-stratum, and i^{th} land-use stratum,

f_{ijk} = the number of tracts in the k^{th} segment, j^{th} sub-stratum, and i^{th} land-use stratum,

e_{ijk} = the expansion factor for the k^{th} segment in the j^{th} sub-stratum and i^{th} land-use stratum,

r_{ij} = the number of sample replicates or segments in the j^{th} sub-stratum, and i^{th} land-use stratum,

s_i = the number of sub-strata in the i^{th} land-use stratum,

L = the number of land-use strata in the state.

The closed estimator is simple and easy to use. Farm establishments report only for data within the segment boundaries. Reported data is easily verified and thus relatively free of reporting errors. The closed estimator can be very precise for estimating agricultural items such as planted acreages. However, other agricultural items, such as farm labor and cash receipts, can only be reported accurately for the entire farm establishment. A closed estimator is not reasonable for estimating such items. This approach usually requires a face-to-face interview to show the segment boundaries to the farm operator. Thus data collection costs are high.

Weighted Estimator

The weighted estimator uses entire farm data, and prorates (or weights) some portion of that data to each population unit (segment) in which the farm has land. A variety of weighting schemes are possible, the only restriction is that the sum of the weights for a farm across all population units will equal "one." NASS currently uses a ratio of "tract acres minus farmstead" to "entire farm acres minus farmstead" as its operational weight. Reported data for the entire farm is multiplied by this weight and summed to the segment level and then expanded for the entire population.

The state level sample estimate using the weighted estimator may be expressed mathematically as follows:

$$\hat{Y}_w = \sum_{i=1}^{L} \sum_{j=1}^{s_i} \sum_{k=1}^{r_{ij}} y_{ijk}$$

where

$$y_{ijk} = \begin{cases} e_{ijk} \sum_{l=1}^{f_{ijk}} a_{ijkl} z_{ijkl} = e_{ijk} \sum_{l=1}^{f_{ijk}} w_{ijkl} & \text{if } f_{ijk} > 0, \\ 0 & \text{if } f_{ijk} = 0, \end{cases}$$

w_{ijkl} = the weighted value of the survey item for the lth operation with land in the kth segment, jth sub-stratum, and ith land-use stratum,

a_{ijkl} = the weight for the lth agricultural operation with land in the kth segment, jth sub-stratum, and ith land-use stratum,

z_{ijkl} = the value of the survey item on the total acres operated for the lth operation with land in the kth segment, jth sub-stratum, and ith land-use stratum,

e_{ijk}, r_{ij}, s_i, L are previously defined.

The weighted estimator incorporates entire farm level data and thus can be used for any agricultural item. Once the "tract acres minus farmstead" value is established for each operation, less expensive collection procedures are possible as face-to-face interviews are not required. NASS has found, however, that weighted estimates are often biased upward when the weight depends on whole farm acreage. Farm operators under-report farm acreage (which included cultivated plus non-cultivated land), which in turn causes the weight to be biased upward. The NASS operational weight suffers from this problem. By eliminating the farmstead in the weight calculation, NASS simplifies screening in agri-urban strata, where a farm operator may reside apart from his/her operation.

Open Estimator

The open estimator is a special case of the weighted estimator, which gives a weight of "one" to farm establishments whose operator resides within the segment, and a weight of zero otherwise. Data need only be collected from resident farm operators, thus reducing data collection costs and respondent burden. However, many disadvantages are associated with the use of the open estimator, and NASS has discontinued its use. First, the estimates are less precise than other weighted estimators. Second, farm operator residences are sometimes missed when screening segments in agri-urban areas. This causes open estimators to be biased downward. Intensive, and expensive, screening procedures are needed to make this estimator work satisfactorily.

Cost

The construction and maintenance of a national level area frame is a costly undertaking with respect to both labor and materials. When constructed and maintained on paper, the cost of labor far outweighed the cost of the materials. Many hours were required for the delineation of strata and PSUs on several different media. Additional hours were required for reviews. The use of the CASS system has shifted the relative cost of labor and materials. Many activities are now automated. Using this system the stratification of an average county takes approximately 44 staff hours. Using paper materials the same county would take approximately 105 staff hours.

The Arkansas frame was the last one constructed using paper materials. The process used approximately 10,000 staff hours ($86,000). Materials (including paper satellite imagery, photography, and maps covering the whole state, photo enlargements of selected segments) cost approximately $30,000. Thus the cost of building the frame was approximately $116,000, with 75 percent of the total for labor. The Oklahoma frame was larger and cost approximately $124,000 to complete. With CASS, however, only 35 percent of the total was for labor. The major recurring cost with CASS is the purchase of digital satellite imagery. CASS also had significant up front costs for equipment and software development. Over time, we expect labor costs to increase and the cost of digital satellite imagery to decrease, making the CASS system a truly cost effective medium for the construction of area frames.

Data collection costs are also of interest. NASS enumerates approximately 15,000 segments during the June Agriculture Survey each year. Data is collected during a two week time frame by approximately 1600 enumerators. Costs average $180 per segment. This includes enumerator training, travel, screening, and data collection.

Quality Control and Assessment

Quality control and assessment is an ongoing process within the frame construction process and throughout the useful life of that frame. The following sections discuss the process of discovering and correcting problems with individual segments, and procedures for assessing the deterioration of an older frame.

Problem Segments

Occasionally a segment is selected that can not be efficiently enumerated. These segments are termed "problem segments" and require immediate, careful attention. Problem segments are generally caused by one of two situations: 1) segment boundaries are not well defined, or 2) the segment is too large or contains too many farm establishments to enumerate accurately in a reasonable amount of time.

The first assessment of the quality of segment boundaries occurs when the boundaries are copied onto aerial photography. Because the line mapping data overlaid on the satellite imagery in the CASS system is usually older than the aerial photography, some boundaries chosen with CASS may not appear on the photography. In those cases, cartographers make small adjustments to the segment boundaries to accommodate the boundaries on the photography. On rare occasions, PSU and stratum boundaries are also adjusted. Care is used to avoid changing the number of sampling units. The second assessment occurs during data collection. If a boundary error is found at this point, the segment is adjusted prior to next year's survey.

Problems associated with the size of the segment and with the number of interviews required are usually discovered during the initial screening. These are resolved by dividing the segment into a number of smaller parcels of land and randomly selecting one. The expansion factor for the new segment is appropriately modified.

Assessing the Deterioration of an Older Frame

Land utilization within each state is constantly changing. As a result, over time a state's area frame will contain an increasing number of segments that do not conform to their stratum's definition. This occurrence, in turn, damages the frame's ability to produce useful and accurate estimates. Frames exhibiting this characteristic are said to be "aging".

Bush describes a systematic approach to prioritize states for new frame construction. The approach consists of: 1) deciding upon objective criteria, or standards, by which to judge each frame, 2) ranking the states for each individual criteria, 3) assigning weights, or relative importance, to each criteria, and 4) using the weighted ranks to arrive at an overall ordering based on all criteria. Bush uses the following criteria in his assessment.

1) Percentage of segments meeting strata specifications. Assuming that almost all segments met their stratum definitions when the frame was new, this serves as a basic measure of stratification aging.

2) Relative importance of state to the national estimating program. A national level optimal sample allocation analysis is performed for commodities whose estimates rely heavily upon the area frame (as opposed to being estimated from the list frame of farm operators). The objective is to highlight states needing an increased sample size in order to reach national level precision goals.

3) Availability of current aerial photography. Though frame construction is now automated with the use of the CASS system, sampled segments are still delineated on large scale aerial photography and sent to the state offices for each survey. Ensuring the availability of current photography, therefore, decreases the possibility of adding non-sampling errors to the estimates.

This type of analysis is performed approximately every five years to insure that resources are used efficiently.

OTHER AREA FRAMES WITH A RURAL FOCUS

The remaining section of this paper reviews three other area sampling frames which are designed, in part, to collect information from farm establishments. These are 1) the area frame used by Statistics Canada for agricultural surveys; 2) the Environmental Monitoring and Assessment Program's hexagonal area frame; and 3) the area frame constructed for the National Resource Inventory Survey. The reviews are less detailed than the preceding one. They describe the purpose of each frame, and provide an overview of their design. The paper then compares and contrasts the four area frames from the prospective of collecting information from farm establishments.

<u>Statistics Canada's Area Frame (As Used for Farm Establishments)</u>:

The Agriculture Division of Statistics Canada has been conducting a survey of farm establishments using various forms of area frame methodology since the early seventies. The major purpose in using the area frame is to account for the incomplete coverage of farm establishments on the list frame. During this time frame the quality of the list frame has greatly improved, requiring less dependence on the area frame. The agricultural area frame in Canada relies heavily on use of Enumeration Areas (EAs) and data from the quinquennial census.

The design and construction of area samples is being fundamentally revised in Canada. The previous approach used Census of Agricultural Enumeration Areas as the primary sampling units (PSUs) for the area frame. Enumeration Areas classified as "ag" in the Census (i.e., contain at least one farm headquarters) were subsampled in a two stage design similar to that used for the NASS frame. Using natural boundaries, selected PSUs were broken into 10 to 30 segments of about 6 to 10 square kilometers. A second stage sample of segments was then selected, usually one per PSU. Julien and Maranda (1990) and Ingram and Davidson (1983) discuss the earlier design.

Trepanier and Theberge present a detailed look at the redesign in a paper presented at this conference. It is a single stage design which uses the Universal Transverse Mercator projection to divide the country into 3 x 2 kilometer rectangles or cells. (In the west, 1 x 3 mile segments are used instead of the cells, and a completely different methodology is planned for Prince Edward Island.) The boundaries of these cells and of the Census Enumeration Areas are digitized and overlaid. A computer proportionately distributes census data from an Enumeration Area into all cells that overlap that Area. Cells that straddle Enumeration Area boundaries are assigned data from both Areas. This process assigns measures of agricultural activity to the frame's sampling units. Cells that do not overlap agricultural Enumeration Areas are removed from the population. Likewise cells corresponding to urban and remote regions, forest and water are manually identified and removed. The remaining cells form the population of segments from which the single stage sample is drawn.

This population is stratified first on geographic location and then on a composite measure of agricultural activity. Sample allocation to major geographic regions is proportional to size. Allocation within geographic strata is proportional to the square root of size. The resulting sample consists of approximately 2000 segments. These are plotted on maps for data collection, where enumerators account for all land within the segments. Because of the lack of natural boundaries, the enumerator uses a grid to measure the area of each farm inside the segment rather than relying on the farmer's estimate. In the western part of the country the interview is even conducted over the telephone.

The Environmental Monitoring and Assessment Program's Hexagonal Area Frame

The United States Environmental Protection Agency established the Environmental Monitoring and Assessment Program (EMAP) in the late 1980s. While still in transition, this program is developing an integrated network for environmental monitoring with the following objectives: 1) to estimate, on a regional basis, the current status of and trends in the condition of the nation's ecological resources; 2) to monitor pollutant exposure and to understand the links between existing conditions and human-induced stresses; and 3) provide periodic statistical summaries to policy makers and the public. Inherent in these objectives is the need to statistically sample any land or water based ecological resource, including agricultural land. The information needed is clearly "area" based, and hence EMAP developed an area frame approach to their sample design.

A full description of the design of this area frame is contained in Overton, et al (1990). The process samples the land/water area of the conterminous United States via a grid composed of approximately 12,600 point locations, with 27 km. between points in each direction. The grid was constructed by centering a regular hexagon on the conterminous United States. The hexagon covered the targeted land area and parts of the adjacent continental shelf, southern Canada, and northern Mexico. Each side of the hexagon measured approximately 2,600 km. in length. Six equilateral triangles were constructed within the hexagon by connecting radial lines from the center to each vertex. Next, each side of the equilateral triangles were divided into 96 equal parts. Within each triangle, three sets of 95 parallel lines were constructed. Each set of parallel lines connected the 95 points on the one side of the triangle with their corresponding points on another side of the triangle. This process of constructing intersecting sets of parallel lines created the grid within the base hexagon. Further, these intersecting lines created regular hexagons around each grid point. Of the 28,000 points so constructed, 12,600 fell within the conterminous United States.

These form the baseline grid for the EMAP frame. However, the procedures easily lend themselves to creating additional grid points within specified hexagons whenever higher density sampling is required. From this grid baseline, various tiers of samples can be constructed.

Tier 1 Samples: Regular hexagons were formed using a grid point as the center, using the intersecting lines creating the grid point as radii, and forming sides so that the resulting regular hexagon has an area of approximately 40 sq. km. The 12,600 hexagons thus constructed form the first stage sample of primary sampling units (PSUs) of the EMAP area frame and are called the Tier 1 sample. This sample incorporates approximately 1/16th of the area of the United States. Landscape descriptions are made of each sampled PSU, and each PSU is then partitioned into resource units (those areas occupied by a single resource or land use class).

Tier 2 Samples: These samples are generally resource based. A specific resource is identified for study. PSUs containing that resource type are identified, and subsampled if appropriate. Details of the subsampling procedures were still in design stage when the design report was published. (Overton, et. al). Agricultural cropland is one major resource type of interest

Area Frame of the National Resource Inventory

The National Resources Inventory was last conducted in 1982, and is a comprehensive study of the United States' natural resources. This endeavor is the latest in a series of national inventories conducted by the Soil Conservation Service of the United States Department of Agriculture, which have been conducted every 9-10 years since 1958. The 1982 Inventory was a joint effort between the Soil Conservation Service, the Statistical Laboratory at Iowa State University, and the U.S. Forest Service. The purpose of the Inventory was to provide statistically reliable data on land use, conservation treatment needs, erosion, and other conservation issues at various substate levels defined by either political or natural boundaries. Once again an area frame was developed to sample for this "area based" information.

A full description of the stratified, two stage design of this area frame is contained in USDA (1987). The universe of interest consisted of all nonfederal lands in the conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands. The 3,300 counties in this geographic area served as the sampling base for the process.

Within each county the total surface area was stratified geographically, and land in some counties (where irrigation is important to agriculture) were also stratified according to broad resource and ownership conditions. Many small strata were constructed. In 34 states, the strata were 2-mile by 6-mile rectangular-shaped pieces of land corresponding to 12 sections. In states not covered by the public land survey system, the stratification was based on either latitude-longitude lines or the Universal Transverse Mercator projection. Always strata were constructed on a county by county basis.

Within each stratum, a two-stage area sample was drawn. The primary sampling unit was an area of land which forms a square, one-half mile on each side, containing 160 acres. In Western states some PSU's were 40- or 640-acre squares (the smaller units among irrigated land and the larger among large tracts of range land or forest). In the northeastern U.S., PSU's are 20 seconds of latitude by 30 seconds of longitude and range in size from 97 acres to 114 acres. In Louisiana and northern Maine, the PSUs are 1/2 kilometer squares (61.8 acres), while in Arkansas they are square kilometers of land. The number of PSU's selected in a given stratum depended on the variability of the county relative to land use and soil patterns, size of the county, and projected workload of data collectors. The entire sample consists of approximated 350,000 PSUs, which comprise a 3.5 percent sample of the nonfederal land area of the U.S.

Within each PSU, three point samples were selected. (Exceptions: two selected in 40-acre PSU, and one in Arkansas, Louisiana, and northern Maine). The process of selecting points assured both a random selection and a spread across the PSU. Soil Conservation Service employees collected data for each sample. Some information was collected for the entire PSU (such as area in farmsteads, enumeration of ponds, lakes, steams). Other information relating to soil type, land use, and erosion potential were collected at and for the point sights.

GPS Natural Boundaries
Others no natural Boundaries

COMPARISON OF FRAMES

This final section summarizes and focuses the detail presented earlier in the paper by comparing and contrasting the four frames in terms of a) the purpose for which each was built and the universe over which it can provide inference, b) the sampling units used, and c) the stratification of the sampling units and what that says about estimation efficiency.

The purpose of the NASS area frame is to serve as a sampling base for producing agricultural statistics, both as a single frame and in multiple frame methodology. It provides complete coverage of all land area within the conterminous United States and Hawaii. The purpose of the Statistics Canada frame is almost identical to that of the NASS frame, except that it is used exclusively in the multiple frame context. The Canadian list frame has a higher coverage of farms than the NASS list frame, and therefore the area frame has less impact on the estimating program. It provides complete coverage of all Canadian provinces except Newfoundland. The focus of the EMAP frame and the NRI frame is environmental. Because the land and water used for agricultural production represent a significant portion of total natural resources of the United States, both frames can be used to target farm establishments. For the NRI frame, agricultural land is intended to be its main focus. It provides complete coverage of all nonfederal land in the conterminous United States plus Hawaii, Puerto Rico and the Virgin Islands. The EMAP frame is designed to focus on many different environmental resources. It provides complete coverage of all land area and water masses within the conterminous United States.

The basic sampling unit for the NASS frame is the segment, generally one square mile in size, which has natural boundaries and may be irregularly shaped. Statistics Canada uses rectangular cells, generally 3 x 2 kilometers in size, which were defined using the Universal Transverse Mercator projection rather than natural boundaries. In the west, they follow segment lines. The basic sampling units for the NRI frame are the PSUs and the three point samples selected within each sampled PSU. The PSUs are square areas, one-fourth square mile in size, that do not follow natural boundaries. The EMAP frame uses 40 sq. km hexagons as the basic sampling unit. These where built using a grid system, and do not follow natural boundaries. In three of the four frames the lack of natural boundaries in defining the sampling unit causes more difficulty during data collection, and increases the chance of enumeration errors.

The NASS frame is built individually for each state, and population units are stratified by general land use categories and sub-stratified geographically within each state. It uses a two stage design with heavier sampling rates in intensive agricultural strata. This provides relatively efficient estimates of major agricultural production items. The area frame used by Statistics Canada is first stratified geographically and then by a measure of agricultural activity obtained from the Agricultural Census. It is a single stage design, and like the NASS frame, samples areas of intensive agriculture more heavily. The use of a single stage design and availability of Census data for stratification has the potential for making this frame the most efficient of the four for targeting farm establishments. The NRI frame is stratified geographically, but has no other stratification to target agricultural activity. This probably leads to some lack of efficiency in estimating agricultural items. The EMAP frame serves many different purposes so it is designed to spread samples geographically, but has no stratification. It is probably the least efficient for targeting farm establishments.

REFERENCES

Bush, Jeffrey. 1993. "Ranking the States for Area Frame Development." Staff Report Number SMD 93-01. Washington, D.C.: U. S. Department of Agriculture, National Agricultural Statistics Service.

Cotter, Jim and Jack Nealon. 1987. "Area Frame Design For Agricultural Surveys." Staff Report. Washington, D.C.: U. S. Department of Agriculture, National Agricultural Statistics Service.

Goebel, J. Jeffrey, Mark Reiser, and Roy D. Hickman. 1985. "Sampling and Estimation in the 1982 National Resources Inventory." *Proceedings of the American Statistical Association Meetings.*

Gordon, Daniel K.. 1985. "An Investigation Of Thematic Mapper Satellite Imagery For Inventorying Fruit Trees In New York." Thesis presented at Cornell University.

Ingram, S. and G. Davidson. 1983. "Methods Used in Designing the National Farm Survey." *Proceeding of the Section on Survey Research Methods, American Statistical Association.*

Julien, C. and F. Maranda. 1990. "Sample Design of the 1988 National Farm Survey." *Survey Methodology* 16, 117-129.

Marx, Robert W. 1984. "Developing An Integrated Cartographic/Geographic Data Base For The United States Bureau Of The Census." Washington, D.C.: United States Department of Commerce, Bureau of the Census.

Mergerson, James W. 1989. "Area Frame Sampling: Sample Allocation." Internal Documentation. Washington, D.C.: United States Department of Agricultural, National Agricultural Statistics Service.

Nealon, John Patrick. 1984. "Review of the Multiple and Area Frame Estimators." Staff Report Number 80. Washington, D.C.: United States Department of Agriculture, Statistical Reporting Service.

Nealon, Jack. 1990. Revised. "Statistical Standard For Area Frame Problem Segments." Internal Documentation. Washington, D.C.: United States Department of Agriculture, National Agricultural Statistics Service.

Overton, W. Scott, Dennis White, and Don L. Stevens. 1990. *Design Report for EMAP*. EPA/600/3-91/053. Washington, D.C.: U. S. Environmental Protection Agency, Office of Research and Development.

Theberge, Alain, and John G. Kovar. 1993. "The Design of the Canadian Area Farm Survey." Florence: Presented at the 49th Session of the International Statistical Institute.

U.S. Department of Agriculture. 1987. *Basic Statistics 1982 National Resources Inventory.* Statistical Bulletin Number 756. Washington, DC.: U. S. Department of Agricultural, Soil Conservation Service.

U.S. Department of Agriculture, Iowa State University. 1987. *National Resource Inventory: A Guide For Users of 1982 NRI Data Files.* Unpublished report.

U.S. Department of Agriculture. 1992. "Agricultural Surveys Supervising and Editing Manual, Section 3." Internal Documentation. Washington, D.C.: National Agricultural Statistics Service.