11

Area frame design for agricultural surveys

Jim Cotter, Carrie Davies, Jack Nealon and Ray Roberts
US Department of Agriculture, National Agricultural Statistics Service, USA

11.1 Introduction

The National Agricultural Statistics Service (NASS) has been developing, using and analysing 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 country consists of a collection or listing of all parcels of land for the area of interest from which to sample. These land parcels can be defined based on factors such as ownership or based simply on easily identifiable boundaries as is done by the NASS.

The purpose of this document is to describe the procedures used by the NASS to develop and sample area frames for agricultural surveys. The process involves many steps, which have been developed to provide statistical and cost efficiencies. Some of the key steps are as follows:

- *Stratification.* The distribution of crops and livestock can vary considerably across a state in the United States. The precision of the survey estimates or statistics can be substantially improved by dividing the land in a state into homogeneous groups or strata and then optimally allocating the total sample to the strata. The basic stratification employed by the NASS involves: (1) dividing the land into land-use strata such as intensively cultivated land, urban areas and range land, and

Agricultural Survey Methods Edited by Roberto Benedetti, Marco Bee, Giuseppe Espa and Federica Piersimoni
 ISBN: 978-0-470-74371-3

(2) further dividing each land-use stratum into substrata by grouping areas that are agriculturally similar.

- *Multi-step sampling.* Within each stratum, the land can be divided into all the sampling units or segments and then a sample of segments selected for a survey. This would be a very time-consuming endeavour. The time spent developing and sampling a frame can be greatly reduced by: (1) dividing the land into larger sampling units called first-step or primary sampling units (PSUs); (2) selecting a sample of PSUs and then delineating the segments only for these PSUs; and (3) selecting a sample of segments from the selected PSUs.
- *Analysis.* Several decisions are made that can have an appreciable impact on the statistical and cost efficiency. These include decisions such as the land-use strata definitions, the number of substrata, the size of the sampling units, the allocation of and the method of selecting the sample necessary to guide us in these decisions.

The major area frame survey conducted by the NASS is the June Agricultural Survey (JAS). This mid-year survey provides area frame estimates primarily for crop acreages and livestock inventories. During the survey, the interviewers visit each segment in the sample, which has been accurately identified on aerial photography, and interview each person who operates land inside the boundaries of the selected segments. With the respondent's assistance, field boundaries are identified on the photography and the acreage and crop type reported for each field in the segment. Counts of livestock within each sample segment are also obtained. This area frame information is subsequently used to provide state, regional and national estimates for crop acreages, livestock inventories and other agricultural items. Naturally, the procedures used to develop and sample area frames affect the precision and accuracy of the survey statistics.

11.1.1 Brief history

Iowa State University began construction of area frames for use in agricultural surveys in 1938. The NASS began research into the use of area sampling frames in the mid-1950s to provide the foundation for conducting probability surveys based on complete coverage of the farm sector. In 1954, area frame surveys were begun on a research basis in ten states, 100 counties with 703 ultimate sampling units or segments. These surveys were then expanded over the years and made operational in 1965 in the contiguous United States.

Changes made to the area frame methodology during the 1960s and early 1970s were mainly associated with sampling methods such as land-use stratification and replicated sampling (described in detail in Section 11.5). Technological changes were incorporated during the seventies and eighties in the form of increased computerization, use of satellite imagery, use of analytical software and development of an area frame sample management system among others.

The area frame programme has grown over the past 54 years and is now conducted in 49 states with approximately 11 000 segments being visited by data collection personnel for the major agricultural survey conducted during June of each year. Today, the NASS maintains an area frame for each state except Alaska; it also maintains an area frame for Puerto Rico. The frames are constructed one state at a time and used year after year until deemed outdated. A frame is generally utilized for 15–20 years, and when it becomes outdated, a new frame is constructed to replace it. Each year, three or four states are

selected to receive a new frame. The selection of states for new frames is based on the following criteria: age of the frame, significant land-use changes, target coefficients of variance being met, and significance to the national programme.

11.1.2 Advantages of using an area frame

- *Versatility.* Since reporting units can be associated with an area of land (a sampling unit), an area frame can be used to collect data for multiple variables in one survey. For example, crop acreage, livestock, grain production and stocks, and economic data are all collected during the JAS.
- *Complete coverage.* The NASS's area frame is complete, meaning when all the sampling units are aggregated, the entire population is completely covered and every sampling unit has a known chance of being selected. The sampling units do not overlap, nor are there gaps between adjacent sampling units. This is a tremendous advantage since it provides the vehicle to generate unbiased survey estimates. Complete coverage is also useful in multiple-frame (area and list) surveys where the area frame is used to measure the degree of incompleteness of the list frame.
- *Statistical soundness.* The advantage of complete coverage combined with a random selection of sampling units is that it can provide unbiased estimates with measurable precision.
- *Non-sampling errors reduced.* Face-to-face interviews are conducted for the JAS, which generally result in better-quality data being gathered than data collected by mail or telephone. The interviewer uses an aerial photograph showing the location and boundary of the sample segment to collect data for all land within the segment boundary such as crop acreages, residential areas, and forest. If the respondent refuses to participate in the survey, or is inaccessible, the interviewer is instructed to make observations which are helpful when making non-response adjustments.
- *Longevity.* Once an area frame is constructed, it can be used year after year without having to update the sampling units. The frames can become inefficient as land use changes. However, the area frames constructed for most states last 15–20 years before they need to be replaced.

11.1.3 Disadvantages of using an area frame

- *Can be less efficient than a list frame.* If a list of farm operators can be stratified by a variable related to the survey items, it will provide greater sampling efficiency than an area frame that is stratified by land use. For example, a list frame that is stratified by peak number of cattle and calves will provide greater sampling efficiency than the area frame when estimating cattle inventory. Unfortunately, the NASS list frame does not provide 100% coverage, because of the difficulty of obtaining and maintaining a complete list of producer names, addresses, and appropriate control data. Since the area frame is a complete sampling frame, it is used to measure incompleteness in the list.

- *Cost.* An area frame can be very expensive to build and sample. New frame construction, on average, uses five full-time employees for four months per state. Also, face-to-face interviews conducted by a trained staff are also very costly.
- *Lack of good boundaries.* Although this is not a problem for most areas in the United States, it can be when building a frame in a foreign country. The importance of quality boundaries will be discussed later.
- *Sensitivity to outliers.* Because the sampling rate for the JAS is low, expansion factors are relatively high. For this reason, area frame surveys are sometimes plagued by a few 'extremely large' operations that are in sample segments. These operations can greatly distort the survey estimates. A solution to this problem is to identify all very large operations prior to the survey (special list frame) and sample them with certainty.

11.1.4 How the NASS uses an area frame

- *Acreage estimates for major commodities.* The primary focus is on corn, soybeans, winter wheat, spring wheat, durum wheat, cotton, not on list (NOL) cattle, and number of farms. The acreage for each crop is recorded for each field within the segment. These acreages are then expanded to produce an estimate of crop acreage at the state and national level.
- *Measure the incompleteness of the list.* The NASS maintains a list of farmers who operate land in the country. Because farm operations go in and out of business, the list is never complete at any given time. The JAS survey is used to find farmers who are not on the NASS's list. During the JAS data collection, the interviewer records the names of all people who operate agricultural land within each segment. These names are then checked against the NASS's list of farm operators. Those who are not present on the list are referred to as NOL. The NOL operations found during the JAS are multiplied by an expansion factor to estimate the incompleteness of the list for each state.
- *Ground truth for remotely sensed crop acreage estimates.* The crop data for each field inside the segment from the JAS are used to determine what crop spectral signatures from the satellite represent. Identified signatures are then used to classify fields throughout a state. Once the satellite data have been classified, an acreage estimate can be made for various crops grown in that state.
- *Follow-on surveys.* The NASS uses the data from the JAS for follow-on surveys such as the Objective Yield Survey where specifically cropped fields are randomly selected with probability proportional to size. These yield surveys involve making counts, measurements, and weightings of selected crops. Every 5 years additional sampling units are added to the JAS for the Agricultural Coverage Evaluation Survey (ACES). Data collected from the JAS segments, in combination with the additional ACES segments, are used to measure the completeness of the Census of Agriculture mail list. The Not on Mail List (NML) estimates are used to weight census data at the record level to produce coverage-adjusted estimates.

11.2 Pre-construction analysis

Before building a new frame, analysis is conducted to determine which states are most in need of one. Generally three to four states are selected to receive a new frame each year. Data collected from approximately 11 000 segments during the JAS is used to determine the extent to which the land-use stratification has deteriorated for each state. This involves comparing the coefficients of variation for the survey estimates of major items over the life of the frame. Typically states with the oldest frames have the highest probability of being selected. Also important is the extent to which a state contributes to the national programme for major commodities. For example, Kansas contributes approximately 20% to the national estimate for winter wheat. If it were determined that the state's JAS target coefficients of variance for winter wheat were not being met, Kansas would be likely to be selected to receive a new frame.

Once a state has been selected to receive a new frame, analysis is performed to determine the most appropriate stratification scheme to be used. Previous years' survey data are used to calculate the percentage of cultivated land in the sample segments, the average number of interviews per segment in each stratum, and the variances for important crops in each stratum. These data are used to determine the following:

Land-use strata definitions. Several land-use strata are common to all frames, including cultivated land, ag-urban, urban, and non-agricultural land. The cultivated land is divided into several strata based on the distribution of cultivation in the state. Previous years' survey data are analysed to provide information such as the percentage of cultivated land in the sample segments so that the distribution of cultivated land can be ascertained. This will help determine the number of and definition of the cultivated strata. Table 11.1 presents the land-use stratification scheme generally followed along with the codes to be used during the stratification process.

Strata 11, 12, and 20 are where the majority of cropland is present. These strata target commodities such as corn, soybeans, cotton, and wheat. In many states, strata 11 and 12 are collapsed into one stratum. The 40's strata contain less than 15% cultivation. Range and pasture land, as well as woods, mountains, desert, swampland, etc., are put into stratum 40. Cattle and other livestock operations are usually also found in stratum 40. Little to no agriculture is expected to be found in strata 31, 32, and 50. These strata are present

Table 11.1 Land-use strata codes and definitions.

Stratum	Definition
11	General cropland, 75% or more cultivated
12	General cropland, 50–74% cultivated
20	General cropland, 15–49% cultivated
31	Ag-urban, less than 15% cultivated, more than 100 dwellings per square mile, residential mixed with agriculture
32	Residential/commercial, no cultivation, more than 100 dwellings per square mile
40	Less than 15% cultivated
50	Non-agricultural, variable size segments

in all states. Stratum 31 contains dense commercial and residential areas of cities and towns. The ag-urban land in stratum 32 represents a mixture of residential and areas with the potential for agricultural activity, usually located in a band around a city, where the city blends into the rural area. Stratum 50 contains non-agricultural entities such as state and national parks, game and wildlife refuges, military installations and large airports.

These are the strata present in most states, however; adjustments may be made to the design depending on the state involved. For example, stratum 40 is often broken into two or more strata in the western states, with a special stratum for forest or desert. Also, a stratum may be added for Indian reservation land. Crop-specific strata are also used in several states to allow the opportunity to channel a sample either into, or away from, a certain area. For example, citrus strata were created in Florida. However, since an annual citrus survey conducted in Florida provides reliable estimates, the JAS is not used for citrus estimates. The citrus strata are in place to allow for a heavier sample in strata where field crops are present.

PSU and segment sizes. In the process of constructing a new area frame, all land in the selected state will be broken down into PSUs. The population of PSUs is sampled from by stratum, and the selected PSUs are further broken down into an average of six to eight segments from which one segment is chosen. This way, an entire frame need not be divided into segments, saving a tremendous amount in labour costs.

Before area frame construction can start, the sizes of the PSUs and segments must be determined. The target PSU and segment sizes are determined for each stratum based on the analysis of previous years' JAS data. The target size of the segment is determined first.

The optimum segment size for a land-use stratum depends upon a multitude of often interrelated factors such as the survey objectives, data collection costs, data variability among segments, interview length, population density, concentration of cropland, and the availability of identifiable boundaries for the segments. The segment size, which is determined in the pre-construction phase, is based on the analysis of previous years' JAS data. The target segment size varies from stratum to stratum and state to state. Table 11.2 is an example of the segment sizes per strata for a typical state.

When the PSUs in stratum 11 are broken down in this example, the resulting segments should be as close to 1 square mile as possible. The target segment size in the intensively cultivated strata (10s strata) is usually 1 square mile, with the exception of a few states where the target segment size is less. In the moderately cultivated strata (20s strata), the target segment size is typically 1–2 square miles.

Table 11.2 Target segment sizes.

Stratum	Definition	Target segment size (square miles)
11	General cropland, 75% or more cultivated	1.00
12	General cropland, 50–74% cultivated	1.00
20	General cropland, 15–49% cultivated	1.00
31	Ag-urban	0.25
32	Residential/commercial	0.10
40	Open land, <15% cultivated	2.00
50	Non-agricultural, variable size segments	PPS

The target segment size for open land strata (40s strata) varies the most. In states where good boundaries are available, the target segment size can be 1–2 square miles. In some areas (e.g. desert, mountainous, or range areas), boundaries are few and far apart. The target segment size in these areas will range from 4–8 square miles. If adequate boundaries are not available, the segments in the strata will not have a target segment size. Segment size will vary depending on available boundaries. The probability of selecting a segment is proportional to the size of the segment (PPS).

The target segment sizes in the urban and ag-urban strata are always one-tenth and one-quarter square mile, respectively. In stratum 50, the non-agricultural stratum, there is no segment size (except for states that have not received a new frame since 1985). Entities such as state parks, forests, airports, military land, etc. are placed in stratum 50. The boundary of the segment is the boundary of the entity. The probability of selecting a segment in this stratum is proportional to the size of the segment (PPS).

When determining the segment size for each stratum, the following are taken into consideration:

- *Minimize sampling variability.* Ideally the segments within a (non-PPS) substratum will be equal in size and homogeneous in terms of agricultural content to keep variance down. As the size of the segments decreases, so does the ability to delineate segments (to be discussed later) that are homogeneous with respect to the amount of cultivated land. Therefore, the sampling variability among segments increases for a given sample size.

- *Availability of boundaries.* As the size of the segments decreases, the availability of suitable boundaries also decreases. Quality boundaries are pertinent when delineating PSUs and segments. If boundaries are not available to delineate segments that are equal in size, variability may increase. Also, if poor-quality boundaries are used, the result could mean more reporting errors during the data collection phase. In highly cultivated strata, as well as the urban and ag-urban strata, quality boundaries are generally plentiful, allowing for a smaller segment size. The land in the 40s strata consists of forest, desert, range, pasture, etc. Quality boundaries in these strata are more spread apart, making a larger segment size more accommodating.

- *Data collection costs.* The interviewer must contact and interview each person operating farmland within the segment boundaries. To minimize data collection costs, the interviewer should be able to complete a segment in less than 12 hours. Research has shown that interviewers are generally able to complete an average of 3–4 interviews per segment in 12 hours.

The target segment size for a stratum will be based partly on the average number interviews per segment. In moderate to intensively cultivated strata (10s and 20s strata), farms are relative close together. A segment size of 1 square mile will result in an average of 3–4 interviews. In the lower intensively cultivated strata (40s strata), the farm operations are typically farther apart in location. The segment size in these strata can be larger and still not increase data collection costs. The target segment size in ag-urban (stratum 31), urban (stratum 32), and non-agricultural strata has no influence on data collection costs since few if any interviews are done for segments in these strata.

The target sizes of the PSUs are based on the segment size. PSUs should contain six to eight segments, so the PSU size should be about six times the segment size. PSUs that

Table 11.3 Primary sampling unit size tolerance guide.

Stratum	Cultivation	PSU size (sq miles)		
		Minimum	Target	Maximum
11	75% cultivated	1.00	6.00–8.00	9.00
12	50–75% cultivated	1.00	6.00–8.00	9.00
20	15–49% cultivated	1.00	6.00–8.00	9.00
31	Ag-urban	0.25	1.00–2.00	3.00
32	Commercial	0.10	0.50–1.00	1.00
40	>15% cultivated	4.00	20.00–24.00	36.00
50	Non-ag	–	PPS	–

are smaller than the target size will be broken down into fewer segments, and PSUs that are larger will be broken down into more segments if boundaries are available. So if a PSU in stratum 11 is only 2 square miles, it will most likely be broken down into two segments. The minimum PSU size is generally one segment. Table 11.3 is an example of the PSU size tolerance range for the target segment sizes in Table 11.2.

Once the analysis is complete and strata definitions and segment sizes are specified by the Area Frame Section (AFS), stratification will begin. After this point, the strata definitions, PSU and segment sizes are used for the life of the frame and are not changed.

11.3 Land-use stratification

The process of land-use stratification is the delineation of land areas into land-use categories (strata). 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. The stratification unit divides the land within each county into PSUs using quality physical boundaries, then assigns them to a land-use strata (defined in the pre-construction phase). Later in area frame construction, the PSUs are further divided into segments, also using quality physical boundaries. A quality physical 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 non-sampling errors to be introduced into the survey data. Also, 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. Quality boundaries include highways, roads, railroads, rivers, streams, canals, and section lines.

The stratifier breaks down all the land within each county into PSUs. The stratifiers locate the best physical boundaries and draw off PSUs as close to the target PSU size (defined in the pre-construction phase) as possible. They use the following materials to locate boundaries:

- *Topographic quadrangle maps.* Produced by the US Geological Survey (USGS), digital raster graphic maps are scanned images of USGS 1:100 000 scale topographic maps.

- *Tele Atlas data.* Produced by Dynamap, these accurate and complete digital vector data provide the NASS with an accurate map base on which to verify boundaries during the frame construction process.
- *National Agriculture Imagery Program (NAIP).* This is operated by the Farm Service Agency (FSA), which acquires one- and two-metre digital ortho-imagery during the agricultural growing seasons in the continental USA. Coverage provides approximately 20% one-metre and 80% two-metre. The NAIP imagery is used to verify boundaries during the frame construction process and has improved the accuracy of the frame. These data are also used when generating the photo enlargements used for the JAS data collection.

Simultaneously, they use the following materials to classify the PSUs into strata:

- *Satellite imagery.* Satellite imagery is derived from digital data collected by sensors aboard satellites. The AFS currently uses imagery from the LANDSAT 7 satellite. Satellite imagery is used primarily to ascertain where the cultivated areas and the pasture areas are present in a county.
- *Cropland Data Layer.* This is an agriculture specific land cover geospatial product developed by the Spatial Analysis Research Section (SARS). It is used during the frame construction process as an additional layer to assist the stratifier in isolating agriculture. It is also used to assist in isolating crop specific strata in a number of states.

Once all of the PSUs in the county have been delineated and classified into strata, the PSU identification number is attached. This is done automatically in ArcGIS9. The PSUs are numbered beginning in the upper right-hand corner and winding through the county in a serpentine fashion. Figure 11.1 shows an example of the numbering scheme. The first number is the stratum and the second is an incremental PSU number.

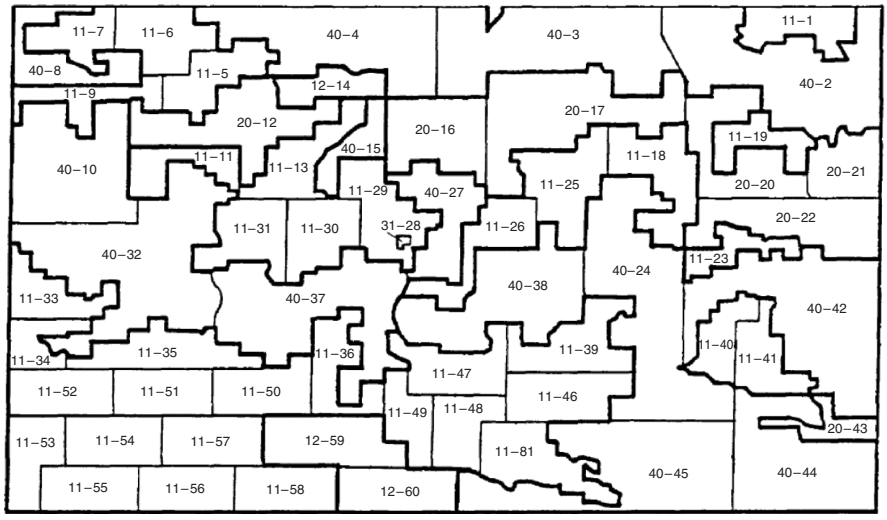

Figure 11.1 PSU ordering in a serpentine manner.

11.4 Sub-stratification

There is a further level of stratification which is applied to the frame. Sub-stratification is the process used to divide the population of sampling units within each stratum equally into categories (substrata). These substrata do not have a definition associated with them like strata do (e.g. 50% or more cultivated). Sampling units are placed into substrata based on likeness of agricultural content and, to a certain extent, location. Sub-stratification activities include ordering the PSUs, ordering the counties, calculating the number of sampling units in the strata, determining the number of substrata, and placing the sampling units into substrata.

Recall in Section 11.3 that when the stratifier completes stratification for a county, the PSUs within the county are ordered automatically in ArgGIS9. The PSUs are numbered beginning in the upper right-hand corner and winding through the county in a serpentine fashion. This ordering plays a role in creating the substrata. Once stratification is complete and all PSUs within each county are ordered, the counties are ordered by an area frame statistician. This county ordering is based on a multivariate cluster analysis of county level crop and livestock data. The purpose of cluster analysis is to group counties into clusters or groups which generally have the same overall agricultural make-up.

Figure 11.2 exhibits the county ordering used in the Pennsylvania area sampling frame. Note that in all but one instance, the ordering proceeds from one county into an adjacent county. The reason for the exception along the southern border of the state is that Somerset County is more agriculturally similar to Fulton County than the adjacent Bedford County. The county ordering need not be continuous. If the counties in one corner of the state are very similar to those in another corner, the ordering can skip across several counties. The starting point of the ordering is somewhat arbitrary, so a logical starting point would be any corner of the state. However, if the cluster analysis indicates a clear distinction between two groups of counties, it may be advantageous to start in one area and end in the other.

The county ordering 'links' the PSU ordering within each county together. In the example above, the PSU ordering for the state begins with the PSU ordering in the first

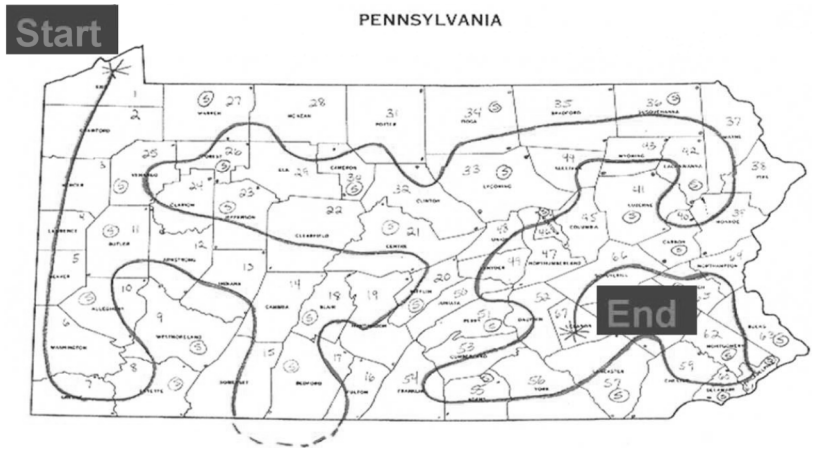

Figure 11.2 County ordering used for the Pennsylvania area frame.

county, Erie County. The PSUs ordered in the second county, Crawford county, go next in the ordering, and so on. When the ordering 'enters' a county from the west or the south, the order of the PSUs in the county is reversed. PSUs within a county are ordered by arbitrarily starting in the northeast corner of the county. Therefore, reversing the order will ensure a fairly continuous ordering of PSUs from one county to the next.

Since sampling units (not PSUs) are placed into substrata, the population of sampling units needs to be calculated for each stratum. Only PSUs that are chosen for the sample are physically broken down into sampling units or segments. However, the number of potential sampling units must be determined for all PSUs in the population in order to calculate the population of sampling units. The number of sampling units varies from PSU to PSU depending on the PSUs size. The number of sampling units for any given PSU in the strata is

$$N_{ik} = \frac{S_{ik}}{S_i},$$

where N_{ik} is the number of potential sampling units in the kth PSU from the ith land-use stratum, rounded to the nearest whole number, S_{ik} is the size of the kth PSU from the ith land-use stratum, and S_i is the target size of sampling units in the ith land-use stratum.

For example, if a PSU is 8.3 square miles in size, and the target segment size is 1.0 square miles, then the number of potential sampling units that could be delineated from the PSU would be eight. The number of potential sampling units is determined for all PSUs in the population of each land-use stratum. Then the total number of sampling units is

$$N_i = \sum_{i=1}^{k} N_{ik},$$

where N_i is the number of sampling units in the ith land-use stratum, and N_{ik} is the number of sampling units in the kth PSU from the ith land-use stratum.

After the population of segments has been determined for each stratum, the number of substrata for each land-use stratum is established. Several factors are considered in the determination, including experience with sampling frames in other states, the number of sample segments and replicates within each stratum and the degree of homogeneity among the sampling units within the various strata. Generally, the higher the intensity of cultivation and variation in crops, the higher the number of substrata relative to the sample size.

Table 11.4 is the sample design for Pennsylvania. The land-use stratum definition and target segment size are determined in the pre-construction phase. The number of sampling units for each stratum is calculated after stratification is complete. The sample size is determined by the sample allocation.

The NASS employs a concept called replicated sampling which provides several key benefits in the estimation process (described in Section 11.5). Approximately 20% of the replicates are rotated out of the sample each year with new replicates taking their place. The number of substrata is the sample size divided by the number of replicates, as is illustrated in Table 11.4.

In this example, there are 2800 sampling units in stratum 13. So 700 sampling units will go into each of the four substrata. The first 700 sampling units in the ordering within the stratum will go into substratum 1. The next 700 sampling units in the ordering within the stratum will go into substratum 2, and so on. In many cases the substrata

Table 11.4 Pennsylvania sample design showing the number of substrata and replications.

Stratum	Land-use stratum	Target segment size (sq mi)	Number of sampling units	Number of substrata	Number of replications	Sample size
13	>50% cult.	1.00	2800	4	6	24
20	15–49% cult.	1.00	17084	14	7	98
31	Ag-urban	0.25	8284	1	5	5
32	Commercial	0.10	1814	1	2	2
40	<15% cultivated	2.00	11344	8	6	48
50	Non-ag	PPS	45	1	2	2

'break' will 'split' the last PSU (some sampling units will be in one substratum, the rest in the next substratum). Each substratum will contain the same number of sampling units, except the last, which may contain slightly more or fewer than the others due to rounding. For example, the 17 084 sampling units in stratum 20 are divided into 14 substrata. The first 13 substrata will contain 1220 units and the last will contain 1224.

Sub-stratification is implemented to reduce variability in sampling units. The land-use stratification is based on the percentage of cultivation. Therefore, while the majority of the segments within a stratum may be intensely cultivated, the agricultural make-up of the segments may differ depending on the location of the segments within the state. Ordering the population of PSUs according to agricultural content will yield greater precision in the estimates for individual commodities. Sub-stratification is particularly effective in areas of intensive cultivation where cropland content varies across the state. Utilizing substrata in grazing or range strata contributes very little to reducing variance except possibly for cattle. Therefore, more substrata are used in the intensely cultivated strata as compared to the range or lightly-cultivated strata.

11.5 Replicated sampling

NASS's area frames have been sampled using a replicated design since 1974. Replicated sampling is characterized by the selection of a number of independent subsamples or replicates from the same population using the same selection procedure for each replicate. Each replicate is therefore an unbiased representation of the population.

A replicate for NASS's area frame sample design is a random sample of land areas (segments) selected within a land-use stratum. The sub-stratification within each land-use stratum has been incorporated into the sampling process to improve the sampling efficiency and the sample dispersion. Therefore, a replicate is more specifically defined as a simple random sample of one segment from each substratum in a land-use stratum.

The first segment randomly selected in each substratum in a land-use stratum is designated as replicate 1, the second segment selected from each substratum is designated as replicate 2, and so forth. The number of replicates is the same for each substratum in a given land-use stratum. Therefore, the number of sample segments in a land-use stratum is simply the product of the number of replicates and the number of substrata in

the land-use stratum,

$$n_i = r_i \cdot s_i,$$

where n_i is the number of segments in the sample for the ith land-use stratum, r_i is the number of replicates for each substratum in the ith land-use stratum and s_i is the the number of substrata in the ith land-use stratum.

Suppose, for example, we want to select a replicated sample of two replicates from a land-use stratum consisting of three substrata with ten segments in each substratum. Then the total sample size for the land-use stratum would be $n_i = r_i \cdot s_i = 2 \cdot 3 = 6$ segments, as illustrated in Table 11.5. Notice that a simple random sample of one segment is selected

Table 11.5 Replicated sampling process for a land-use stratum.

		Replicate	
Substratum	Segment	1	2
1	1		
	2		
	3		
	4		
	5		
	6		×
	7		
	8	×	
	9		
	10		
2	11		
	12		
	13		
	14		
	15		
	16		
	17		
	18		×
	19	×	
	20		
3	21		
	22		
	23		
	24		
	25		
	26	×	
	27		
	28		×
	29		
	30		

in each substratum for a replicate so that the number of sample segments in a replicate is simply the number of substrata.

The number of replicates certainly does not need to be the same in each substratum. Sometimes it may be advantageous to vary the number of replicates in the substrata for a land-use stratum. For example, if a crop is localized to a few counties in a state and greater precision is desired for data pertaining to this crop, then the sampling variance could be reduced for this crop by increasing the number of replicates in the substrata corresponding to these counties.

There are six reasons why the NASS uses replicated sampling:

1. *Sample rotation.* A sample rotation scheme is used to reduce respondent burden caused by repeated interviewing, avoid the expense of selecting a completely new area sample each year, and provide reliable measures of change in the production of agricultural commodities from year to year through the use of the ratio estimator. Sample rotation is accomplished each year by replacing segments from specified replicates in each land-use stratum with newly selected segments. Approximately 20% of the replicates in each land-use stratum are replaced annually. The sample design does not rotate exactly 20% of the segments because the number of replicates is not always a multiple of 5. To illustrate how replicated sampling simplifies the sample rotation process, Table 11.6 shows the numbering scheme for a hypothetical land-use stratum with five replicates in each of eight substrata. The first digit in the five-digit segment number represents the year the segment rotated into the sample, e.g. 50001 entered in 2005. The remaining four digits are simply unique numbers. The sample rotation in 2010 will be performed by replacing the segments in the 50000 series (replicate 1), which have been in the sample for 5 years, with segments numbered 00001, 00002,. . ., 00008. In 2011, the segments will be replaced from replicate 2 since the 60000 series would have completed its five-year sample cycle. In 2012, the segments from replicate 3 will be replaced and so forth.

2. *Methodology research.* Replicated sampling provides the capability to test alternative survey procedures or evaluate current methodology since different replicates can be assigned to the research and operational methods. For example, if there are a total of ten replicates in a land-use stratum and there is a need to compare two

Table 11.6 Replicated sampling process for a land-use stratum.

	Replicate				
Substratum	1	2	3	4	5
1	50001	60009	70017	80025	90033
2	50002	60010	70018	80026	90034
3	50003	60011	70019	80027	90035
4	50004	60012	70020	80028	90036
5	50005	60013	70021	80029	90037
6	50006	60014	70022	80030	90038
7	50007	60015	70023	80031	90039
8	50008	60016	70024	80032	90040

approaches to asking a particular question, then five replicates could be assigned to each method. The test statistic could then be easily derived using the means or totals from each replicate for each approach. Some examples of survey procedures that might be tested are different questionnaire designs and alternative interviewing approaches.

3. *Quality assurance.* Replication also facilitates quality assurance analysis by allowing data comparisons among years in order to determine if significant differences in survey processes exist over time. For example, segment sizes can readily be compared among replicates to determine if the average size and the variability in size differ significantly from year to year. If so, this may indicate that the manual procedures for delineating segments (to be discussed later) need to be reviewed.

4. *Sample management.* Replication allows easy management of the sample due to the replicate numbering scheme. This simplifies the process of designating a subsample of segments for one-time or repetitive surveys, increasing or decreasing the sample size in a land-use stratum to improve sampling efficiency, and identifying segments to be rotated out of the area frame sample. For example, replicates are added every 5 years for the ACES survey to estimate the completeness of the Census mail list.

5. *Variance estimation.* Replicated sampling provides a simple, unbiased method for estimating the sampling variance using replicate means or totals. The NASS estimates the sampling variance for agricultural surveys using the sub-stratification design rather than replicate totals. However, replicate totals are sometimes used for variance and covariance estimation to simplify multivariate statistical analysis in research studies. The benefit of using replicate totals to estimate the sampling variance is most pronounced in underdeveloped countries where a computer facility or the necessary statistical software is not available.

6. *Rotation effects.* Replication readily provides the vehicle for evaluating sample rotation effects. Rotation effects are defined as the impact on survey data resulting from the number of years a segment has been in the sample. The NASS has a five-year rotation process which permits replicate totals to be compared for segments in the sample from one to five years.

11.6 Sample allocation

The area frame sample is used to collect data on a wide range of agricultural items such as crop acreages, livestock inventories and economic data. Therefore, the allocation of the sample across states and within states to the land-use strata is extremely important. The NASS evaluates optimum allocations of the sample to obtain the most precision in the major survey estimates for a given budget. The number of sample segments allocated to each land-use stratum and state depends on factors such as the average data collection cost per segment in each stratum, the variability of the data in each stratum resulting from the intensity and diversity of agriculture, the total number of segments or land area in each stratum, and the importance of the state's agriculture relative to the national agricultural statistics programme.

An optimum sample allocation to the land-use strata is generated for each of the most important agricultural survey items (univariate) and for all of the important commodities considered simultaneously (multivariate). These important commodities include corn, soybeans, cotton, winter wheat, spring wheat, durum wheat, number of farms, and NOL cattle. The allocations are evaluated not only from an area frame perspective but also from a multiple frame point of view where the area frame is used to measure the incompleteness in the list frame. Finally, optimum allocations are conducted at the national, regional, and state levels to assess the allocations at the various inference levels.

The NASS places the most importance on the multivariate optimum allocation for the area frame non-overlap estimates at the state level since it is important to provide useful statistics at the state level. Adjustments are made to this sample allocation to improve the precision of the regional and national estimates without seriously hindering the precision levels for the states. Minor adjustments to the optimum allocation are also made to provide a multiple of five replicates in each stratum to simplify the sample rotation process and to protect against the impact of outliers by not allowing the sampling rate to be too small in a stratum, e.g. 1 in 750 segments.

The optimum allocation of a sample for multi-purpose surveys can be viewed as a problem in convex programming. An iterative, nonlinear programming algorithm is used to provide the univariate and multivariate optimum sample allocations for the area frames. The algorithm is guaranteed to converge to the optimum solution. A brief description of the multivariate sample allocation model follows.

Suppose each of the j survey items, $1 \leq j \leq p$, from the p selected survey items must satisfy the constraint

$$\text{var}(\hat{Y}_j) \leq v_j,$$

where $\text{var}(\hat{Y}_j)$ is the estimated sampling variance for the jth survey total, and v_j is the desired or target sampling variance for the jth survey total. Assume the cost function

$$C(x) = \sum_{i=1}^{l} a_{ij} c_i n_i = \sum_{i=1}^{l} a_{ij} \frac{c_i}{x_i},$$

where c_i is the average cost per segment in the ith land-use stratum, n_i is the number of sample segments in the ith land-use stratum, l is the number of land-use strata, and $x_i = 1/n_i$ with $n_i \geq 1$. The problem then reduces to minimizing the cost function subject to the constraints

$$\sum_{i=1}^{l} a_{ij} x_i \leq 1, \qquad 1 \leq j \leq p,$$

where $a_{ij} = N_i^2 s_{ij}^2 / (v_j + \sum_{i=1}^{l} N_i s_{ij}^2)$, s_{ij}^2 is the square of the standard deviation for the jth survey item in the ith land-use stratum and N_i is the number of segments in the ith land-use stratum. The nonlinear algorithm iteratively finds the intersection between $A_k = \{x : C(x) = k\}$ for fixed values of k, and $F = \{x : a_j' x \leq 1\}$. The intersection is the optimal solution. Experience has shown that the program converges rapidly to the optimal solution.

Given this allocation model, the input for the model is generated as follows:

- The average cost per segment for each land-use stratum, c_i, is estimated by having the interviewers keep time records during field work.

Table 11.7 Number of segments in the area frame sample, 2008.

State	Number of segments	State	Number of segments
Alabama	236	Nebraska	473
Arizona	118	Nevada	26
Arkansas	342	New Hampshire	10
California	404	New Jersey	48
Colorado	267	New Mexico	124
Connecticut	8	New York	96
Delaware	23	North Carolina	319
Florida	100	North Dakota	420
Georgia	290	Ohio	220
Idaho	148	Oklahoma	335
Illinois	401	Oregon	194
Indiana	264	Pennsylvania	179
Iowa	452	Rhode Island	8
Kansas	487	South Carolina	119
Kentucky	189	South Dakota	395
Louisiana	249	Tennessee	334
Maine	32	Texas	1120
Maryland	61	Utah	69
Massachusetts	12	Vermont	21
Michigan	145	Virginia	179
Minnesota	393	Washington	267
Mississippi	298	West Virginia	66
Missouri	383	Wisconsin	219
Montana	316	Wyoming	53

- The population counts that are calculated after the stratification process.
- The desired sampling variance for the estimated total of each item, v_j, is established by the AFS after consultation with others in the NASS.
- The square of the standard deviation, s_{ij}^2, for the jth item in the ith land-use stratum is estimated using the previous two years' survey data.

The area frame sample allocations among and within states are evaluated periodically to determine if a reallocation of the sample is worthwhile. The sample allocations among the 48 states for 2008 are shown in Table 11.7.

11.7 Selection probabilities

There are two methods for selecting the ultimate sampling unit or segment – equal and unequal selection. Which method is used depends on the availability of adequate boundaries for segments. If good boundaries are plentiful so that segments can be made approximately the same size within a land-use stratum, then segments are selected with equal probability. If adequate boundaries are not available, then unequal probability of selection is used since segment sizes are allowed to vary greatly in order to ensure easily identifiable segment boundaries.

The use of unequal selection probabilities is restricted to the non-agricultural stratum in area frames developed since 1985 and to some open land strata (40s strata) in some western states. In all other land-use strata in the USA, equal probability of selection is used. About 96% of the approximately 11 000 segments in the area frame sample are selected based on the equal probability of selection method.

The probability expressions for equal and unequal probability of selection will now be derived in the context of the NASS's area frame design. These expressions provide the statistical foundation for area frame sampling.

11.7.1 Equal probability of selection

A two-step procedure is used to select sample segments from the selected PSUs when selection probabilities are equal. Recall that the number of segments delineated within the selected PSU depends on the size of a PSU. The number of segments in a PSU is simply the total area of the PSU divided by the target (desired) segment size for the land-use stratum in which the PSU has been stratified. This quotient is rounded to the nearest integer since fractional segments are not allowed. For example, if a PSU in an intensively cultivated stratum is 7.1 square miles and the target segment size is 1.0 square mile, then the number of segments for the PSU is seven.

1. A sample of PSUs ia selected within each substratum in a given land-use stratum. Selection is done randomly, with replacement, with probability proportional to the number of segments in the PSU. That is, the probability of selecting the kth PSU in the jth substratum from the ith land-use stratum is

$$P(A_{ijk}) = \frac{N_{ijk}}{N_{ij}},$$

   where A_{ijk} is the kth PSU in the jth substratum from the ith land-use stratum, N_{ijk} is the number of sampling units (segments) in the kth PSU from the jth substratum in the ith land-use stratum, and N_{ij} is the number of sampling units (segments) in the jth substratum from the ith land-use stratum.

2. After the sample of PSUs is drawn, each selected PSU is divided into the required number of segments. This step involves randomly selecting a segment with equal probability from the selected PSU. That is, the probability of selecting the mth segment given that the kth PSU was selected from the jth substratum in the ith land-use stratum is

$$P(B_{ijkm}|A_{ijk}) = \frac{1}{N_{ijk}},$$

   where B_{ijkm} is the mth segment in the kth PSU from the jth substratum and the ith land-use stratum. Therefore, the unconditional probability of selecting the mth segment in the kth PSU from the jth substratum in the ith land-use stratum is

$$P(B_{ijkm}) = P(A_{ijk}P(B_{ijkm}|A_{ijk}) = \frac{N_{ijk}}{N_{ij}} \cdot \frac{1}{N_{ijk}} = \frac{1}{N_{ij}}.$$

   Therefore, all sampling units within a given substratum in a land-use stratum have an 'equal' probability of selection using the two-step selection procedure. This

Table 11.8 Selection probabilities for the two-step procedure.

PSU	Number of segments in PSU	$P(A_{ijk})$	$P(B_{ijkm}\|A_{ijk})$	$P(B_{ijkm})$
1	2	2/40	1/2	1/40
2	3	3/40	1/3	1/40
3	5	5/40	1/5	1/40
4	6	6/40	1/6	1/40
5	7	7/40	1/7	1/40
6	8	8/40	1/8	1/40
7	9	9/40	1/9	1/40

fact is illustrated in Table 11.8 for a hypothetical substratum with seven PSUs. This table shows the number of required segments in each PSU, the probability of selecting each PSU, $P(A_{iik})$, the probability of selecting a segment given the PSU was selected, $P(B_{ijkm}|A_{ijk})$, and the unconditional probability of selecting a segment in the PSU, $P(B_{ijkm})$. Notice that the unconditional selection probability is the same for all segments, as previously stated.

11.7.2 Unequal probability of selection

PSUs are selected with unequal probability in less cultivated strata (40s strata) in some western states and in the non-agricultural stratum (stratum 50) for states receiving a new area frame since 1985. This type of selection is performed because adequate boundaries are not available in these areas to draw off segments of approximately the same size. The probability of PSU selection in these strata is proportional to its size (PPS). In PPS strata, the PSUs are not further broken down into segments. Therefore, the PSU and segment are synonymous. The probability of selecting the kth PSU in the jth substratum from the ith land-use stratum is

$$P(A_{ijk}) = \frac{S_{ijk}}{S_{ij}},$$

where A_{ijk} is the kth PSU in the jth substratum from the ith land-use stratum, S_{ijk} is the size (in acres) of the kth PSU in the jth substratum from the ith land-use stratum, and S_{ij} is the size (in acres) of the jth substratum in the ith land-use stratum.

The selection probabilities for all situations encountered during the sampling process have now been formulated. The expansion factor or weight assigned to each segment to expand the survey data to population totals is derived from these selection probabilities. The expansion factor for a segment in a substratum is simply the inverse of the product of the probability of selection for the segment and the number of segments in the sample for the substratum,

$$e_{ijm} = \frac{1}{p_{ijm} n_{ij}},$$

where e_{ijm} is the expansion factor for the mth segment in the jth substratum and the ith land-use stratum, p_{ijm} is the probability of selecting the mth segment in the jth substratum from the ith land-use stratum, n_{ij} is the number of segments or replicates in the sample for the jth substratum in the ith land-use stratum.

11.8 Sample selection

The procedures used to select the area frame samples will be described in this section for the equal and unequal probability of selection methods.

11.8.1 Equal probability of selection

Recall that a two-step selection procedure is followed when segments are selected with equal probability. The first step is PSU selection. An SAS program is run which uses the selection probabilities discussed in the previous section to select the chosen PSUs. The program creates a listing of all chosen PSUs.

Personnel in the sample selection unit break down the chosen PSUs that have equal probability of selection into segments in ArcGIS9. NAIP photography is used because it provides valuable detail in terms of land use and availability of boundaries. Three criteria are followed when delineating segments using aerial photography in order to control the total survey error (non-sampling errors and sampling variability):

- Use the most permanent boundaries available for each segment so that reporting problems during the data collection phase caused by ambiguous boundaries will be minimized.
- Create segments that are as homogeneous as possible with respect to agricultural content. Since crop types are generally not distinguishable on the aerial photography, homogeneity is usually based on the amount of cultivated land. This criterion reduces the sampling variability among segments in a given substratum.
- Choose boundaries so that the size of each segment is as close to the target segment size as practical. Deviations from the target size as large as 25% are permitted to satisfy the first two criteria. This criterion, like the second, helps control sampling variability.

After the required number of segments has been delineated for a selected PSU, the segments are automatically numbered in ArcGIS9. Then one segment is chosen at random also in ArcGIS9.

11.8.2 Unequal probability of selection

Recall that PSUs selected with unequal probability in less cultivated strata (40s strata) in some western states and in the non-agricultural stratum (stratum 50) because adequate boundaries are not available in these areas to draw off segments of approximately the same size. The probability of PSU selection in these strata is proportional to its size (PPS). PSUs in PPS strata vary in size and are not broken down further. Because the PSU and segment are one in the same, the sample selection unit reviews the boundaries identified by the stratification unit.

Once the chosen PSUs with equal probability of selection are broken down into segments and the boundaries of chosen PSUs with unequal probability of selection are reviewed, the sample is prepared for data collection. A $24' \times 24'$ image of each segment ($8'$ = one mile scale) is printed onto Kodak photographic paper. This printout is used to collect data for all land inside the segment boundary.

11.9 Sample rotation

As mentioned earlier, the NASS uses a five-year rotation scheme for the sample segments. Rotation is accomplished by replacing segments from specified replicates within a land-use stratum with newly selected segments. Preferably, the number of replicates is a multiple of 5 to provide a constant workload for sample selection and preparation activities in the AFS and for data collection work in the state offices. Naturally, instances occur when the number of replicates is not a multiple of 5, especially in urban, commercial, and non-agricultural strata where the sample size is small (usually two replicates).

Table 11.9 illustrates how the replicates are rotated over a five-year cycle (2008–2012) for different numbers of replicates. If a land-use stratum has two replicates, the segments in replicate 1 will be replaced with all new segments in 2010 and will stay in the sample for 5 years. In 2015, the segments in replicate 1 will be replaced again. Likewise, new segments will rotate into replicate 2 in 2011 and 2016. No segments rotate into or out of the sample in the years in between (2012, 2013, 2014). If a stratum has five replicates, then the segments in one replicate are replaced with new segments each year which is 20% of the sample.

All segments are not in the sample exactly 5 years as has been implied. Segments from the first and last four years of an area frame's life are in the sample less than 5 years, as shown in Table 11.10. This table presents the rotation cycle for an area frame assuming a 20-year life and, for simplicity, five replicates in each land-use stratum.

The national area frame sample size is approximately 11 000 segments. The total number of segments rotated each year is approximately 3000. This results from an average of 800 segments being selected for new area frames (except in census years when no states receive new frames) and about 2200 segments being selected based on a 20% rotation of the remaining 15 000 or so segments. Therefore, approximately 27% of the national area frame sample is based on newly selected segments each year.

Table 11.9 Rotation of replicates depending upon the number of replicates.

	Year				
Number of replicates	2008	2009	2010	2011	2012
2			1	2	
3			1	2	3
4	4	1		2	3
5	4	5	1	2	3
6	4	5,6	1	2	3
7	4,7	5,6	1	2	3
8	4	5	1,6	2,7	3,8
9	4,9	5	1,6	2,7	3,8
10	4,9	5,10	1,6	2,7	3,8
11	4,9	5,10	1,6,11	2,7	3,8
12	4,9	5,10	1,6,11	2,7,12	3,8
13	4,9	5,10	1,6,11	2,7,12	3,8,13
14	4,9,14	5,10	1,6,11	2,7,12	3,8,13
15	4,9,14	5,10,15	1,6,11	2,7,12	3,8,13

Table 11.10 Rotation cycle for a 20-year period assuming five replicates in the stratum.

Year	Replicates																			
2009	1	2	3	4	5															
2010		2	3	4	5	1														
2011			3	4	5	1	2													
2012					4	5	1	2	3											
2013					5	1	2	3	4											
2014						1	2	3	4	5										
2015							2	3	4	5	1									
2016								3	4	5	1	2								
2017									4	5	1	2	3							
2018										5	1	2	3	4						
2019											1	2	3	4	5					
2020												2	3	4	5	1				
2021													3	4	5	1	2			
2022														4	5	1	2	3		
2023															5	1	2	3	4	
2024																1	2	3	4	5
2025																	2	3	4	5
2026																		3	4	5
2027																			4	5
2028																				5

11.10 Sample estimation

This final section will briefly discuss the approaches used to estimate agricultural production with an area frame sample of segments. The NASS uses two area frame estimators, namely the closed and weighted segment estimators. Both require that the interviewer collect data for all farms that operate land inside each segment. (A farm is defined to be all land under one operating arrangement with gross farm sales of at least $1000 a year.) The portion of the farm that is inside the segment is called a tract. The interviewer draws the boundaries of each tract on the photo enlargement, accounting for all land in the segment.

When an interviewer contacts a farmer, the closed segment approach requires that the interviewer obtain data only for that part of the farm within the tract. For example, the interviewer might ask about the total number of hogs on the land in the tract. The most common uses of the closed segment estimator are to estimate crop acreages and livestock inventories. An interviewer accounts for all land in each tract by type of crop or use and for all livestock in the tract. The main disadvantage of the closed segment estimator arises when the farmer can only report values for the farm rather than for a tract which is a subset of the farm. For example, 'How many tractors do you own?' can only be answered on a farm basis. Thus, the closed segment estimator is not applicable for many agricultural items. Economic items and crop production are two major examples which farmers find difficult or impossible to report on a tract basis.

The weighted segment estimator, by contrast, does not have this limitation. It can be used to estimate all agricultural characteristics, which is a major advantage for this

estimator. The weighted segment approach requires that the interviewer obtain data on the entire farm. For example, the interviewer would ask about the total number of hogs on all land in the farm. Using the weighted segment approach, the interviewer obtains farm data for each tract, but these farm data are weighted. The weight used by the NASS is the ratio of tract acres to farm acres.

Suppose the following situation occurs for a specific farm: tract acres = 10, farm acres = 100, hogs on the tract = 20, and hogs on the farm = 40. The closed segment value of number of hogs would be 20, and the weighted segment value would be $40 \cdot 10/100 = 4$.

When estimating survey totals and variances for these estimators, segments can be treated as a stratified sample with random selection within each substratum. The formulas for each of the three estimators can be described by the following notation. For some characteristic, Y, of the farm population, the sample estimate of the total for the closed segment estimator is

$$\hat{Y}_c = \sum_{i=1}^{l} \sum_{j=1}^{s_i} \sum_{k=1}^{n_{ij}} e_{ijk} y_{ijk},$$

where l is the number of land-use strata, s_i is the number of substrata in the ith land-use stratum, n_{ij} is the number of segments sampled in the jth substratum in the ith land-use stratum, e_{ijk} is the expansion factor or inverse of the probability of the selection for the kth segment in the jth substratum in the ith land-use stratum,

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

where f_{ijk} is the number of tracts in the kth segment, jth substratum, and ith land-use stratum, and t_{ijkm} is the tract value of the characteristic Y for the mth tract in the kth segment, jth substratum, and ith land-use stratum.

The weighted segment estimator would also be of the same form,

$$\hat{Y}_w = \sum_{i=1}^{l} \sum_{j=1}^{s_i} \sum_{k=1}^{n_{ij}} e_{ijk} y_{ijk},$$

except that

$$y_{ijk} = \begin{cases} \sum_{i=1}^{f_{ijk}} a_{ijkm} y_{ijkm} & \text{if } f_{ijk} > 0, \\ 0 & \text{if } f_{ijk} = 0, \end{cases}$$

where a_{ijkm} is the weight for the mth tract in the kth segment, jth substratum, and ith land-use stratum.

The following weight is currently in use:

$$a_{ijkm} = \frac{\text{tract acres for the } m\text{th tract}}{\text{farm acres for the } m\text{th tract}}.$$

The precision of an estimate can be measured by the standard error of the estimate. An estimate becomes less precise as the standard error increases. Given the same number of segments to make an estimate, weighted segment estimates are usually more precise than

closed segment estimates. For both estimators, the formula for the sampling variance can be written as

$$\text{var}(\hat{Y}) = \sum_{i=1}^{l} \sum_{j=1}^{s_i} \frac{1 - 1/e_{ij}}{1 - 1/n_{ij}} \sum_{k=1}^{n_{ij}} (y'_{ijk} - y'_{ij.})^2,$$

where $y'_{ijk} = e_{ij} y_{ijk}$ and $y'_{ij.} = (1/n_{ij}) \sum_{k=1}^{n_{ij}} y'_{ijk}$. The standard error is then

$$\text{se}(\hat{Y}) = \sqrt{\text{var}(\hat{Y})}.$$

In closing, research into non-sampling errors associated with these estimators has shown that the closed estimator, when applicable, is generally the least susceptible to non-sampling errors. The closed segment estimator is much relied on for NASS's area frame surveys, and the weighted segment estimator is the most used for multiple-frame surveys where the area frame is only used to measure the incompleteness in the list frame.

11.11 Conclusions

The JAS is an annual area frame survey conducted by the NASS to gather agricultural data such as crop acreages, cost of production, farm expenditures, grain yield and production, and livestock inventories. The JAS provides estimates for major commodities, including corn, soybeans, winter wheat, spring wheat, durum wheat, cotton, NOL cattle, and number of farms. The JAS also provides measurement of the incompleteness of the NASS list frame, provides ground truth data to verify pixels from satellite imagery, and serves as a base for the NASS's follow-on surveys. The development and implementation of an area frame requires several steps, including dividing all land into PSUs, classifying the PSUs into strata and substrata, sampling the PSUs, dividing the chosen PSUs into segments, and randomly selecting a segment from each chosen PSU. While area frames can be costly and time-consuming to build and sample, the importance of the results from the JAS justify the effort.