Area Frame Design for Agricultural Surveys

Carrie Davies

United States Department of Agriculture

National Agricultural Statistics Service

Research and Development Division
Washington DC 20250

RDD Research Report Number RDD-09-xx

June 2009

AREA FRAME DESIGN FOR AGRICULTURAL SURVEYS

PREFACE

Area frames have been and will continue to be the foundation to the agricultural statistics program of the National Agricultural Statistics Service (NASS). The methodology applied by NASS to develop and sample area frames is of interest to domestic and international users of NASS's agricultural statistics. This interest has motivated the need to prepare a document that describes the current area frame procedures in NASS. Therefore, the purpose of this document is to provide a thorough presentation of the present area frame development and sampling procedures employed by NASS.

There are two chapters to this document. The first chapter was originally written by Jim Cotter from NASS in 1987 and updated by Carrie Davies from the Area Frame Section, Research and Development Division, National Agricultural Statistics Service, U.S. Department of Agriculture. Also, the Stratification Materials Section was written by Ray Roberts from the Area Frame Section in NASS in 2009. Chapter I describes the procedures NASS uses to develop new area frames.

The second chapter was written by Jack Nealon from NASS in 1987 and updated by Carrie Davies in 2009. Chapter II describes the area frame sampling methodology. Together, these chapters will provide the reader with a greater understanding and appreciation for the area sampling frame methods currently in use by NASS.

ACKNOWLEDGEMENT

The original authors would like to thank Fred Vogel and Bob Tortora for their suggestions and support.

The reviser would like to thank the following for their important contributions to sections of this document:

Denise Abreu
Belinda Bigelow
Terry Broz
Beth Edwards
Herb Eldridge
Peter Gajary
Rebecca Henson
Elizabeth Ward-Kelly
Andrea Lamas
James Lee
Paula Myers
Ray Roberts

TABLE OF CONTENTS

CHAPTER I. AREA FRAME DEVELOPMENT

Jim Cotter
Carrie Davies
Ray Roberts

Abstract

Area frames are the backbone to the agricultural statistics program of the National Agricultural Statistics Service (NASS). The purpose of this report is to describe the procedures currently used by NASS to develop and sample area frames for agricultural surveys.

Key Words: Area frame; stratification; replicated sampling.

INTRODUCTION

Overview

The National Agricultural Statistics Service (NASS) has been developing, using and analyzing area sampling frames since 1954 as a vehicle for conducting surveys to gather information regarding crop acreage, cost of production, farm expenditures, grain yield and production, livestock inventories and other agricultural items. An area frame for a land area such as a state or country consists of a collection or listing of all parcels of land for the area of interest from which to sample from. These land parcels can be defined based on factors such as ownership or based simply on easily identifiable boundaries as is done by NASS.

The purpose of this document is to describe the procedures used by 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:

- 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 NASS involves: (1) dividing the land into land-use strata such as intensively cultivated land, urban areas and range land, and (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 endeavor. 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.

- Quality Assurance: Care must be taken to ensure that no land is omitted from the frame (unless by design), that no land area is included more than once and that the land is properly stratified into the land-use strata. Once stratification is complete, the resulting PSUs should be reviewed to verify boundaries and proper classification of PSUs into strata. Also, a quality assurance edit should be run to identify errors such as PSUs with area outside the tolerance range and mislabeled PSUs. Catching errors using quality assurance measures will ensure that sound statistical rules are followed to reduce sampling errors and minimize non-sampling errors.

The major area frame survey conducted by 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.

This chapter of the document will begin by briefly outlining the history of area frames in NASS and then will discuss the analysis conducted in the pre-construction stage. The source materials used in the stratification process will be discussed and then the stratification process itself. Other topics covered will be the construction of primary sampling units, the digitization process, the use of sub-stratification and the costs incurred in designing and building an area frame.

Brief History

Iowa State University began construction of area frames for use in agricultural surveys in 1938. NASS began research into the use of area sampling frames in the mid-1950's 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 sixties and early seventies were mainly associated with sampling methods such as land-use stratification and replicated sampling (described in detail in the second chapter of this report). 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.

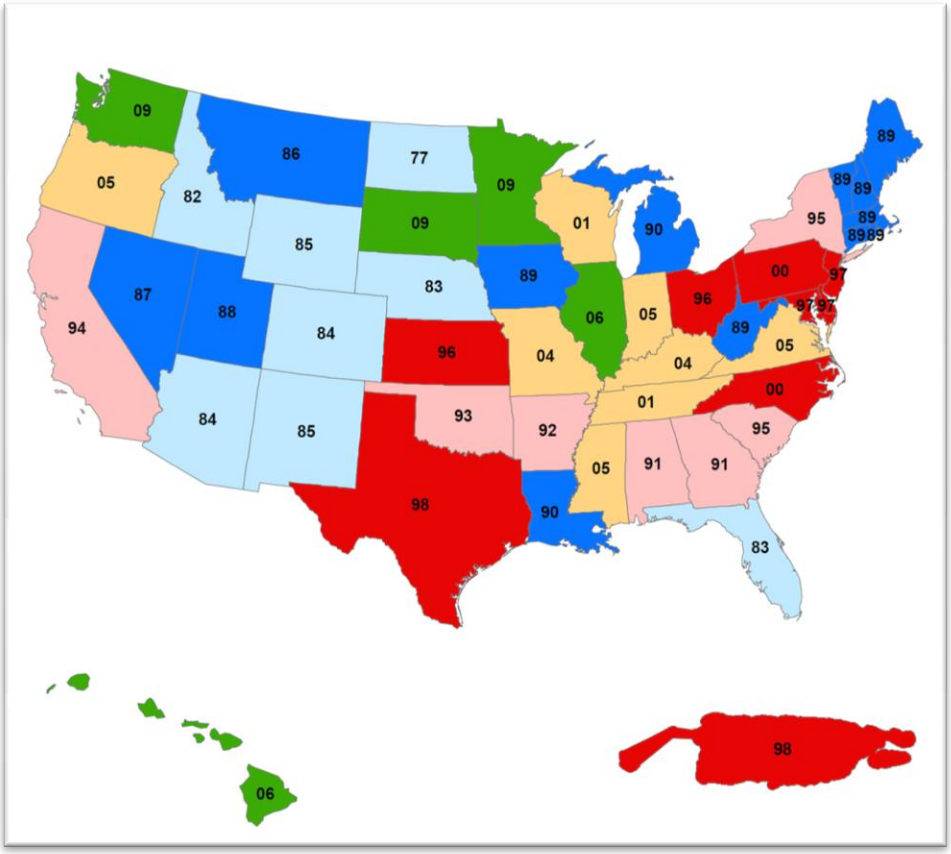
The area frame program has grown over the past 54 years and is now conducted in 49 states with approximately 11,000 segments being visited by data collection personnel for the major agricultural survey conducted during June of each year. Readers interested in a more detailed overview of the historic developments of area sampling frames in NASS are referred to a paper by Fecso, Tortora and Vogel (see Bibliography at the end of this chapter).

Year of Implementation

NASS maintains an area frame for each state except Alaska. NASS also maintains an area frame for Puerto Rico. The frames are constructed one state at a time and used year after year until the frame is deemed outdated. Frames generally are utilized for 15 to 20 years, and when they become outdated, a new frame is constructed to replace it. Each year, three to four states are selected to receive a new frame. The decision of choosing which states are selected for new frames is based on the following criteria: Age of the frame, significant land use changes, target CV's being met, and significance to the national program.

Figure 1.1 displays the year of implementation of NASS's currently used area frames. The number in each state represents the year the current frame was implemented (i.e. Texas' frame was implemented in 1998).

Figure 1.1: Year of Implementation for States that Maintain and Area Frame



Advantages to Using an Area Frame

1) 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.

2) Complete Coverage: 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.

3) Statistically Sound: The advantage of complete coverage combined with a random selection of sampling units is that it can provide unbiased estimates with measurable precision.

4) 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, forest, etc. 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.

5) 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 to 20 years before they need to be replaced.

Disadvantages to using an Area Frame

1) 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, NASS's list frame does not provide 100 percent coverage, because of the difficulty in 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.

2) 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.

3) 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.

4) Sensitive 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.

How NASS Uses an Area Frame

1) 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.

2) Measure the Incompleteness of the List: 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 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 NASS's list of farm operators. Those who are not present on the list are referred to as Not On List or NOL. The NOL operations found during the JAS are multiplies by an expansion factor to estimate the incompleteness of the list for each state.

3) Ground Truth for Remotely Sensed Crop Acreage Estimates: The crop data for each field inside the segment from the JAS is 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 has been classified, an acreage estimate can be made for various crops grown in that state.

4) Follow on Surveys: 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 weighings of selected crops.

   Every five 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, is 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.

PRE-CONSTRUCTION ANALYSIS

Prior to building a new frame, analysis is conducted to determine which states are most in need of a new frame. 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 program for major commodities. For example, Kansas contributes approximately twenty percent to the national estimate for winter wheat. If it were determined that the state's JAS target coefficients of variance for winter wheat was 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 percent 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. This data is used to determine the following:

1) The 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 analyzed to provide information such as the percent 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 1.1 presents the land-use stratification scheme generally followed along with the codes to be used during the stratification process.

Table 1.1 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.

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 percent 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 found in stratum 40.

Little to no agriculture is expected to be found in Strata 31, 32, and 50. These strata are present 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.

2) <u>The PSU and Segment Sizes</u>: 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 are 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 labor 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 target segment size is the same for all segments within each stratum except for segments in PPS (Probability of Selection Proportional to Size) strata like stratum 50. An explanation of how the optimal segment size per stratum is determined is discussed in detail in Chapter II. Table 1.2 is an example of the segment sizes per strata for a typical state.

Table 1.2 Target Segment Sizes

Stratum	Definition	Target Segment Size (square miles)
11	General Cropland, >75% 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 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 are smaller than the target size will be broken down into less 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 two square miles, it will most likely be broken down into two segments. The minimum PSU size is generally one segment. Table 1.3 is an example of the PSU size tolerance range for the target segment sizes in Table 1.2.

Table 1.3 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	---

Once the analysis is complete and strata definitions and segment sizes are specified by the Area Frame Section (AFS), the proposal is sent to the state's Field Office (FO) for comments and suggestions. Once the specifications have been agreed upon, 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. For a list of strata definitions and segment sizes for each state, refer to the Area Frame Design Information document.

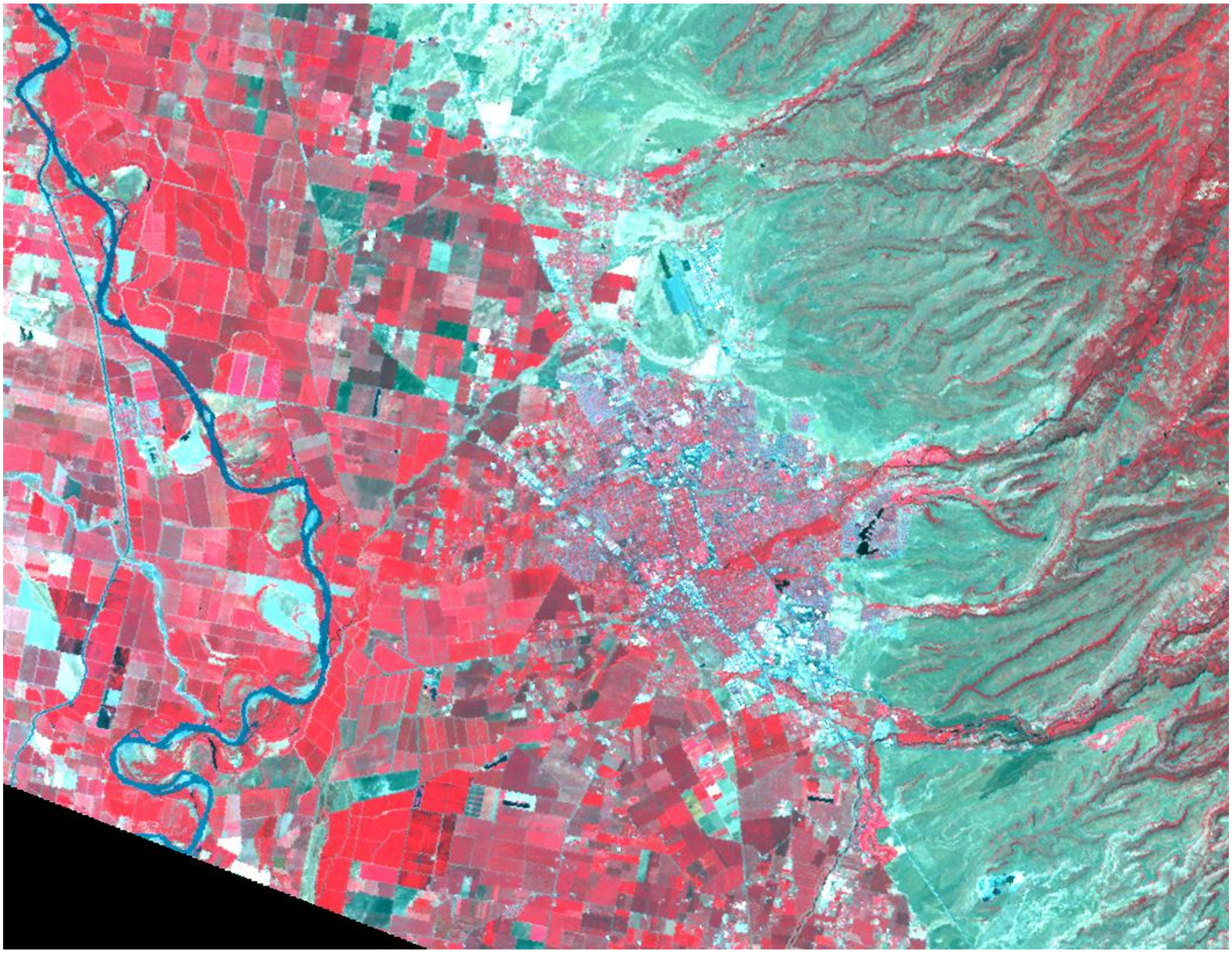
STRATIFICATION MATERIALS

This section describes the different types of data which are used to construct the state area frame.

Satellite Imagery: Satellite imagery is derived from digital data collected by sensors aboard satellites. AFS currently uses imagery from the LANDSAT 7 satellite. The spacecraft completes 14 orbits per day and covers the entire Earth every 16 days. Landsat 7 is equipped with Enhanced Thematic Mapper Plus (ETM +). The sensor has seven bands that simultaneously record reflected or emitted radiation from the Earth's surface in the blue-green (band 1), green (band 2), red (band 3), near-infrared (band 4), mid-infrared (bands 5 and 7), and the far-infrared (band 6) portions of the electromagnetic spectrum. A panchromatic band 8, with a high resolution of 15m exists as well. TM band 2 can detect green reflectance from healthy vegetation, and band 3 is designed for detecting chlorophyll absorption in vegetation. TM band 4 is ideal for near-infrared reflectance peaks in healthy green vegetation, and for detecting water-land interfaces. TM band 1 can penetrate water for bathymetric (water depth) mapping along coastal areas, and is useful for soil-vegetation differentiation, as well as distinguishing forest types. The two mid-infrared bands on TM are useful for vegetation and soil moisture studies, and discriminating between rock and mineral types. The far-infrared band on TM is designed to assist in thermal mapping, and for soil moisture and vegetation studies. The stratification unit typically uses bands 4, 3, and 2.

Satellite imagery is currently not used for boundary identification. Instead, it is used when delineating agricultural strata since it provides a very recent picture of the degree of cultivation. Different land uses will show different color signatures on the imagery. The use of satellite imagery will be further discussed in the land-use stratification section of this chapter. Figure 1.2 is a Landsat 7 image showing bands 4, 3, and 2.

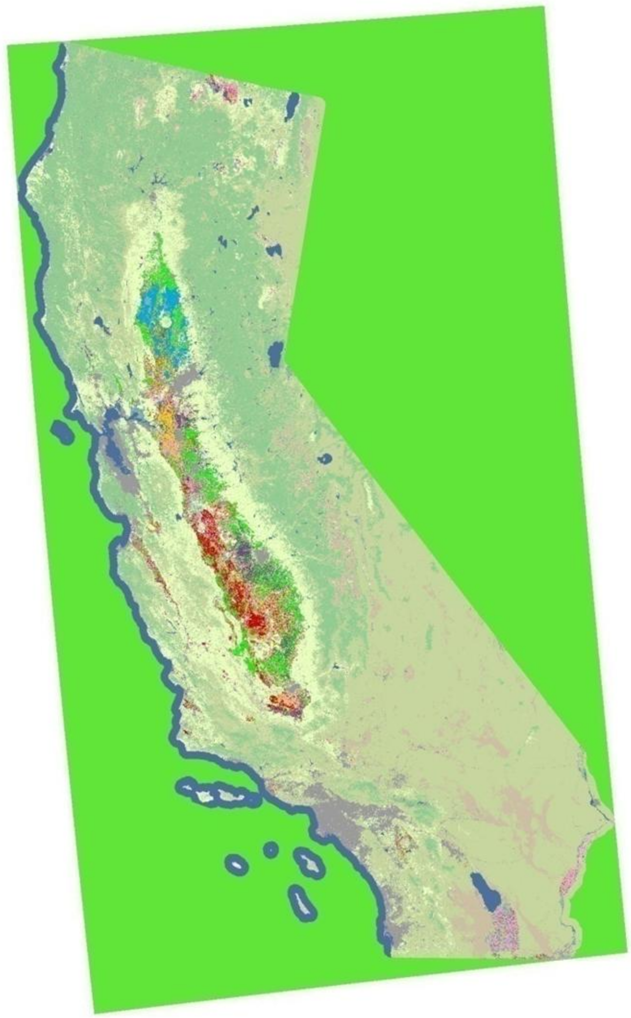
Figure 1.2: Landsat Image Bands 4, 3, and 2



Cropland Data Layer (CDL): The CDL is an agriculture specific land cover geospatial product developed by the Spatial Analysis Research Section (SARS). SARS uses a number of inputs to develop a CDL. Vector inputs include: NASS JAS segments and Farm Services Agency (FSA) Common Land Units (CLU) data. Raster inputs include: USGS NLCD, IRS Resourcesat-1 raw AWiFS Spring/Summer time, NASA Terra MODIS 16 day NDVI, USGS NLCD 2001 Impervious and Canopy, and USGS NED Elevation.

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. Figure 1.3 is the CDL for California.

Figure 1.3: California's Cropland Data Layer



National Agriculture Imagery Program (NAIP): This program is operated by the Farm Service Agency (FSA). The FSA acquires one and two meter digital ortho imagery during the agricultural growing seasons in the continental U.S. Coverage provides approximately 20 percent one meter and 80 percent two meter.

The one meter product is accurate to within five meters of Digital Ortho Quarter Quads (DOQQ's) used to digitize CLU's and other datasets. The one meter product is intended to provide an updated ortho base procured on a three to five year cycle.

The two meter product is accurate within ten meters of DOQQ's intended for aerial compliance and other programs not requiring high spatial accuracy.

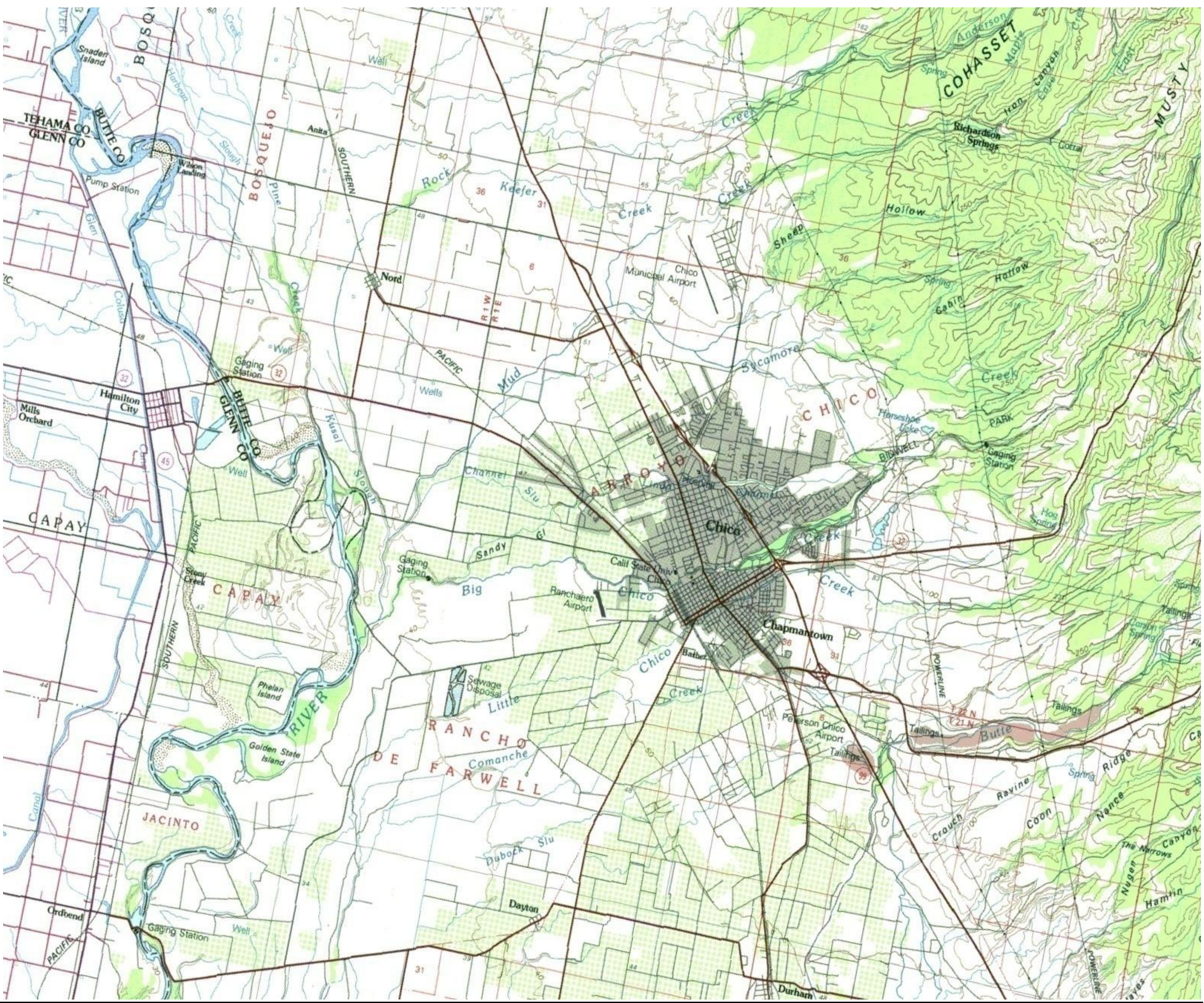
NASS acquires the data via download from FSA. The downloaded data is a compressed county mosaic in MRSID format.

The NAIP imagery is used to verify boundaries during the frame construction process and has improved the accuracy of the frame. This data is also used when generating the photo enlargements used for the JAS. Figure 1.4 is a NAIP image for part of Butte County, California.

Figure 1.4: NAIP Image, Butte County, CA

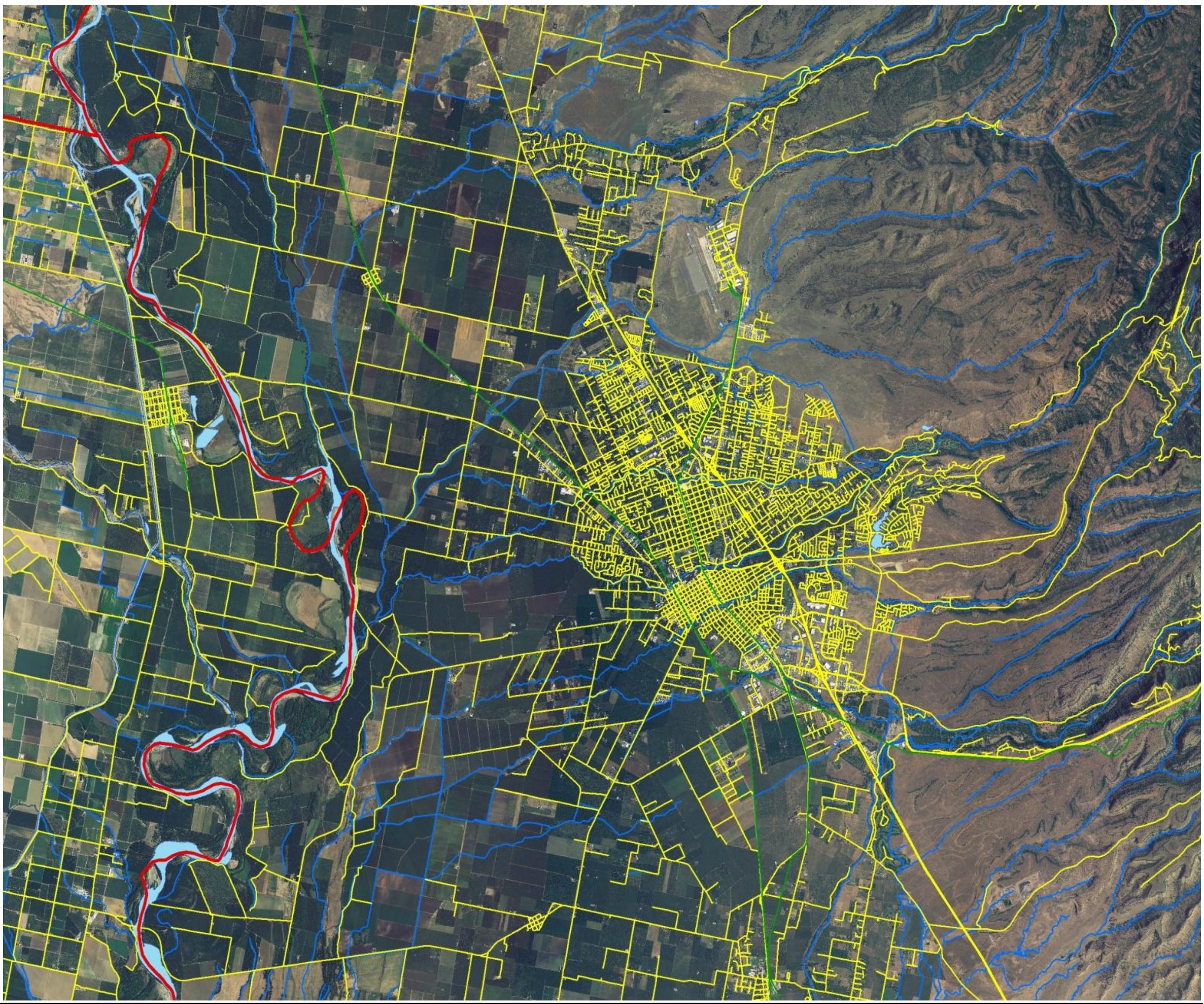
Topographic Quadrangle Map (Quad): Produced by the U.S. Geological Survey (USGS), Digital Raster Graphic (DRG) maps are scanned images of USGS 1:100,000 scale topographic maps. Figure 1.5 is a portion of a digital raster topographic map for Butte County, California. This type of digital map is used as a base map during stratification. It shows political boundaries and physical features such as roads and streams. The digital maps are also used to generate county copy maps. These maps are used to assist enumerators in locating JAS segments.

Figure 1.5: Digital Raster Topographic Map for Butte County, CA



Tele Atlas Data: Produced by Dynamap, this Tele Atlas data provides accurate and complete digital vector data. This data provides NASS with an accurate map base on which to verify boundaries during the frame construction process. Figure 1.6 shows an example of Tele Atlas data over NAIP imagery for part of Butte County, CA. The yellow lines are roads, blue lines are waterways, green lines are railroads, and the red lines are the county boundaries.

Figure 1.6: Tele Atlas Data over NAIP imagery, Butte County, CA



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 preconstruction 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.

Given that the area frame is to be used over a period of 15-20 years and represents a major investment, the best and most permanent boundaries must be used. Roads and rivers make good strata boundaries, while intermittent streams and field edges do not and should rarely be used. The following is a list of geographic features which represent strata boundaries. The list is ranked by quality from highest to lowest:

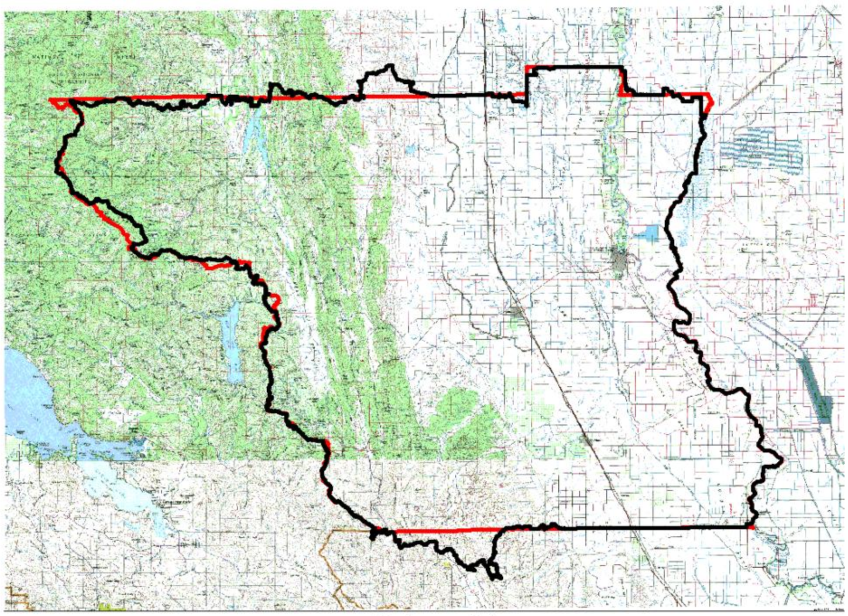
1) Paved highways.
2) Secondary all-weather roads.
3) Local farm to market roads.
4) Railroads.
5) Permanent rivers and streams.
6) Visible section lines.
7) Irrigation canals.
8) Intermittent streams and rivers or prominent water courses that carry water during and immediately following rains.
9) Trails and internal roads.

Most of the boundaries used in the U.S. are 1 through 5. Occasionally, 6 through 9 are used in areas where boundaries may not be present like open range land, swamps, or deserts.

The stratification of a state is performed by county. Each county is assigned to a stratifier who is responsible for all work done on that county. Before stratification can begin, a county outline is created using physical boundaries that follow, as closely as possible, the political county boundary. Since the county outline may be different from the county's political boundary, a

PSU may contain land in two counties. While this may occur across county boundaries, state boundaries must be used whether the political boundary is physical or not (i.e. PSUs cannot contain land in two neighboring states). Figure 1.7 shows the political county boundary (in red) and the physical boundary chosen for the county outline (in black) for Colusa County, California.

Figure 1.7: County Outline for Colusa County, CA



Colusa, CA

Once the county outline is created, the stratifier breaks down all the land within the county outline into PSUs. The stratifiers view topographical maps and aerial photography downloaded in ArcGIS9 to locate the best physical boundaries and draw off PSUs as close to the target PSU size (defined in the preconstruction phase) as possible.

Simultaneously, they use Landsat satellite imagery to classify the PSUs into strata. Satellite imagery is used primarily to ascertain where the cultivated areas and the pasture areas are present in a county. Geographic features and their image colors or signatures include:

- <u>Crops</u>: Different crops will correspondingly produce different color signatures, depending on the growth stage. For example, as the wheat crop emerges, a reddish appearance occurs and becomes a deeper red as growth continues. A ripe wheat field will have a greenish-yellow color, becoming greener just prior to harvest. At harvest, the field will appear white or tan. Most small grain crops will follow this general pattern. Crop calendars are provided to the stratifier showing approximate planting, growing and harvesting intervals for major crops.

- Water appears blue to black. The cleaner and deeper the water, the darker the blue color. Lighter colors are usually the result of high turbidity, not due to temperature.

- Cities and other urban areas appear light blue-gray to bright blue. Concrete is usually white or very light blue. Asphalt is very dark blue or black.

- Clouds are white. The shadows should be black and have the same shape as the clouds next to them. Distances from the clouds to their respective shadows should be similar for all clouds in a given area.

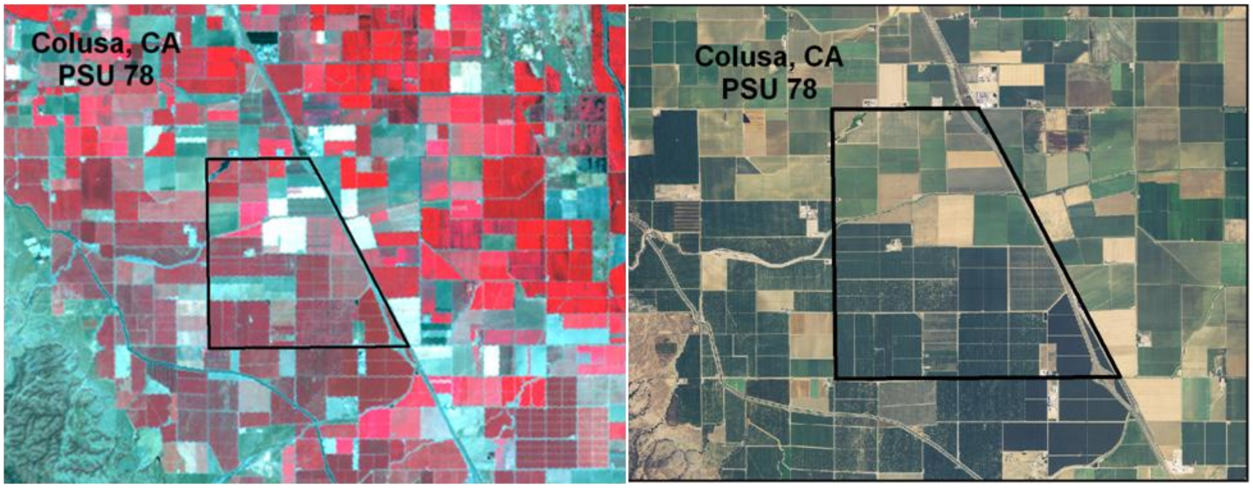
- Bare soil - Soil colors vary dramatically and can be shades of green, gray, blue or brown. Moist soils are darker than dry soils. Bare rock is often brighter than more developed soils.

- Green, growing vegetation will appear pink to red on the imagery.

These color signatures are only general guides. Variations in the signatures can occur depending on the type of crop or what time of the growing season the imagery was taken. Therefore, no single color key will work for all scenes. The Landsat imagery usually covers the most recent growing season, providing a very recent look at the area.

The stratification unit must make subjective decisions on creating PSUs and placing them in their respective strata. For example, stratum 11 may be defined as greater than 75 percent cultivated with a target PSU size of six to eight square miles. The stratifiers use the satellite imagery to identify areas of the county that are highly cultivated and belong in stratum 11 and which areas are less cultivated and belong in other strata. Once an area of the county is deemed greater than 75% cultivated, the stratifier uses the topographical map and the aerial photography to locate physical boundaries and creates the PSU in ArcGIS9. Figure 1.8 is the satellite image (left) and the aerial photograph (right) of a PSU created in stratum 11 in California.

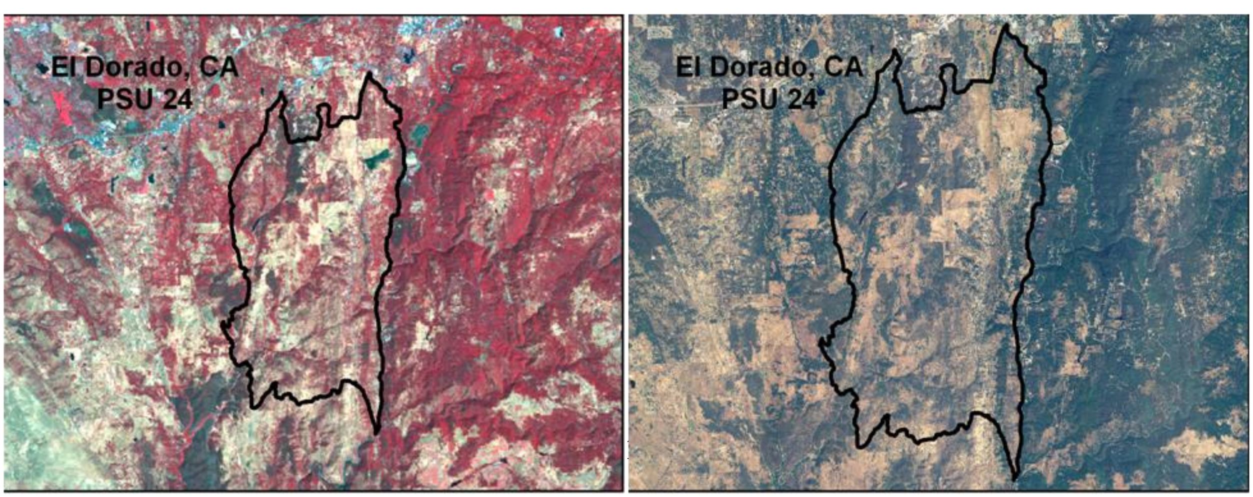
Figure 1.8: PSU created in Stratum 11



Ideally, the created PSU would be six to eight square miles in size, with all quality physical boundaries, and greater than 75 percent cultivation within the PSU. Additionally, the cultivation would be equally distributed within the PSU so that if the PSU is selected to be further broken down, the resulting segments would also adhere to the greater than 75 percent cultivated definition. In the above example, PSU 78 is 7.4 square miles in size with roads used for boundaries on all four sides. Most of the PSU is cultivated with the exception of roads, buildings, etc. The fields show up in different shades of red and green on the satellite imagery depending on the crop grown.

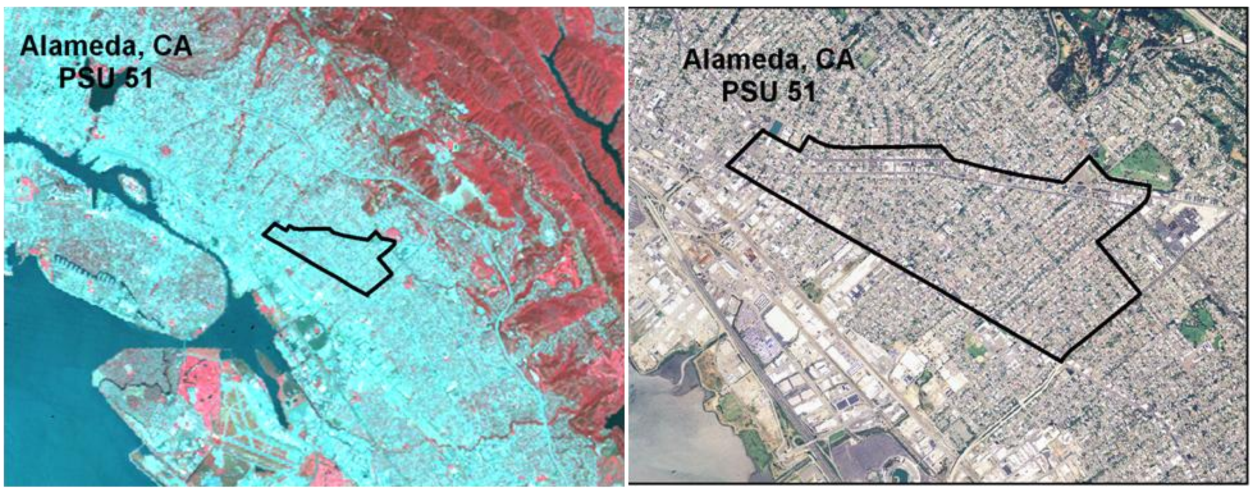
In lesser cultivated areas like stratum 40, the PSUs may contain pasture, woods, desert, mountains, etc. The PSU size in stratum 40 will most likely be larger (10 to 12 square miles) since there are not as many boundaries available in these areas. Figure 1.9 is the satellite image (left) and the aerial photograph (right) of a PSU created in stratum 40 in California. The red areas are woodland and the tan areas may be pasture. This PSU is 24.6 square miles in size.

Figure 1.9: PSU created in Stratum 40



To stratify urban (stratum 32) and ag-urban areas (stratum 31), all cities and towns are located using satellite imagery. The bright blue color indicates urban areas. The PSUs are created in ArcGis9 using the aerial photography and topographic maps for boundary identification. Figure 1.10 is the satellite image (left) and the aerial photograph (right) of a PSU created in stratum 32 in California. The size of this PSU is .98 square miles.

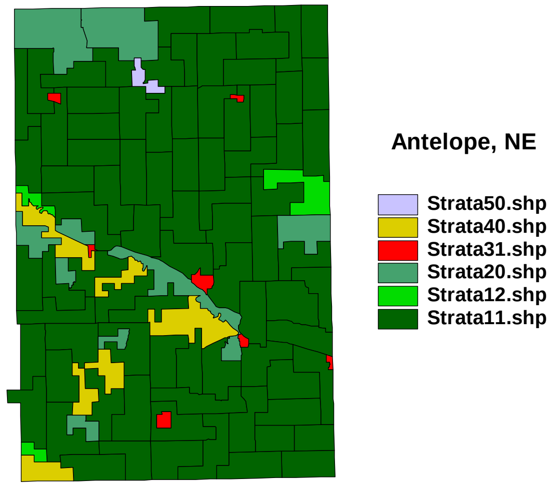
Figure 1.10: PSU created in Stratum 32



Stratum 50 includes national and state parks, wildlife refuges, military installations and airports. The topographic maps are used to identify these entities. Since there is no target segment size in this stratum, the boundary of the entity is used as the boundary of the PSU. Also, the PSUs are not broken down into segments in stratum 50. So when a state park is found, its boundary is not only the PSU boundary, but also the segment boundary. If that segment is chosen, the interviewer would visit the state park and verify that no agricultural activity is present in the entire state park.

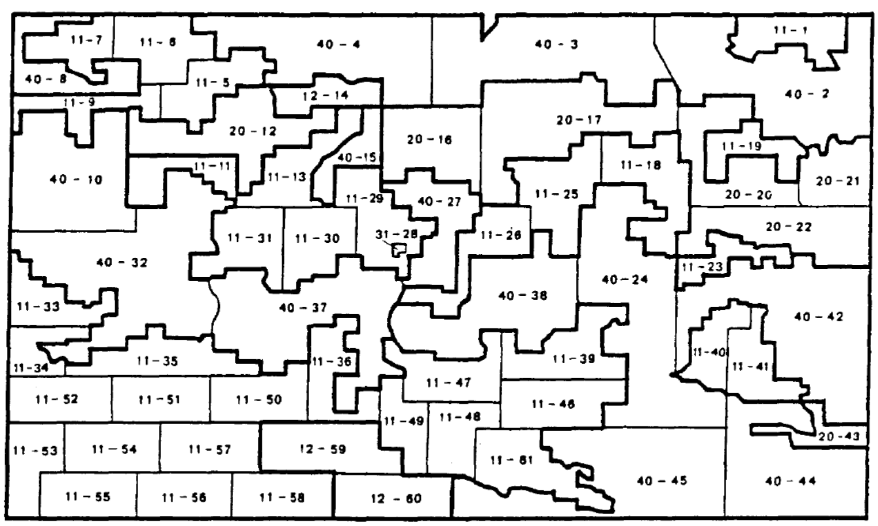
Figure 1.11 shows the stratification of Antelope County in Nebraska. All land inside the county outline has been drawn off into PSUs in black. Each PSU has been assigned to a stratum. This county is highly cultivated (in green) with a small amount of residential areas (in red).

Figure 1.11: Stratification of Antelope County, NE



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 1.12 shows an example of the numbering scheme. The first number is the stratum and the second is an incremental PSU number.

Figure 1.12: PSU Ordering in a Serpentine Manner



QUALITY ASSURANCE

Because the stratification process is open to subjectivity, quality assurance measures are in place to ensure a high quality product. The following checks are performed during the stratification of a state:

1) County Outlines Check: County outlines are established before stratification begins. If the political boundary of a county is not also a physical boundary, the closest physical boundary is used to create a county outline (see Land-Use Stratification for more detail). The reviewer checks to ensure that the most appropriate physical boundaries were chosen to create the county outline. Also, the reviewer verifies that the same physical boundary was used for neighboring counties that share a boundary.

2) PSU Check: Once all PSUs in a county are created and classified into strata, they are examined by the reviewer. When disagreements occur, the reviewer and the stratifier will communicate to resolve any differences.

3) Spot Check: Once the reviewer has finished checking all PSUs in the county, the Lead Cartographer-Technician performs a "spot" check on the work that the stratifier and reviewer have performed.

4) Quality Assurance Edit Check: The Quality Assurance Edit (or Q-edit) is a program that detects PSUs that are outside of the tolerance range for size. Corrections are made to PSUs outside of the tolerance range if boundaries are available. Also, PSU areas for each county are summed and compared against the official county size. County areas are allowed to vary ±25.0 percent from the published area size. The same procedure is performed for the state area. The accumulated state area is only allowed to vary ±0.5 percent from the published area size. Since the stratification is never allowed to cross state boundary lines, only a small amount of error is acceptable.

The PSU areas are then summed to calculate the digitized area for each stratum. The area of the PSU divided by the target segment size for the stratum is equal to the total number of ultimate sampling units (segments) in that PSU. Summing the number of segments will yield the total number of segments in the stratum, which will be utilized in the sample allocation and selection processes (described in detail in the second chapter of this report). Table 1.4 shows the results of that calculation for each stratum in Nebraska.

Table 1.4 Stratum Level Data for Nebraska

Land-Use Stratum	Target Segment Size	Number of PSUs	Average PSU Size	Digitized Area (Sq Mi)	Number of Segments
11	1.00	3,557	8.5	30,338	30,274
12	1.00	1,135	7.8	8,812	8,819
20	2.00	979	9.7	9,597	4,779
31	0.25	640	1.0	642	2,562
32	0.10	90	1.9	169	1,693
40	4.00	1,158	23.8	27,831	6,940
50	1.00	43	4.16	180	183
Total		**7,602**		**77,569**	**55,250**

SUB-STRATIFICATION

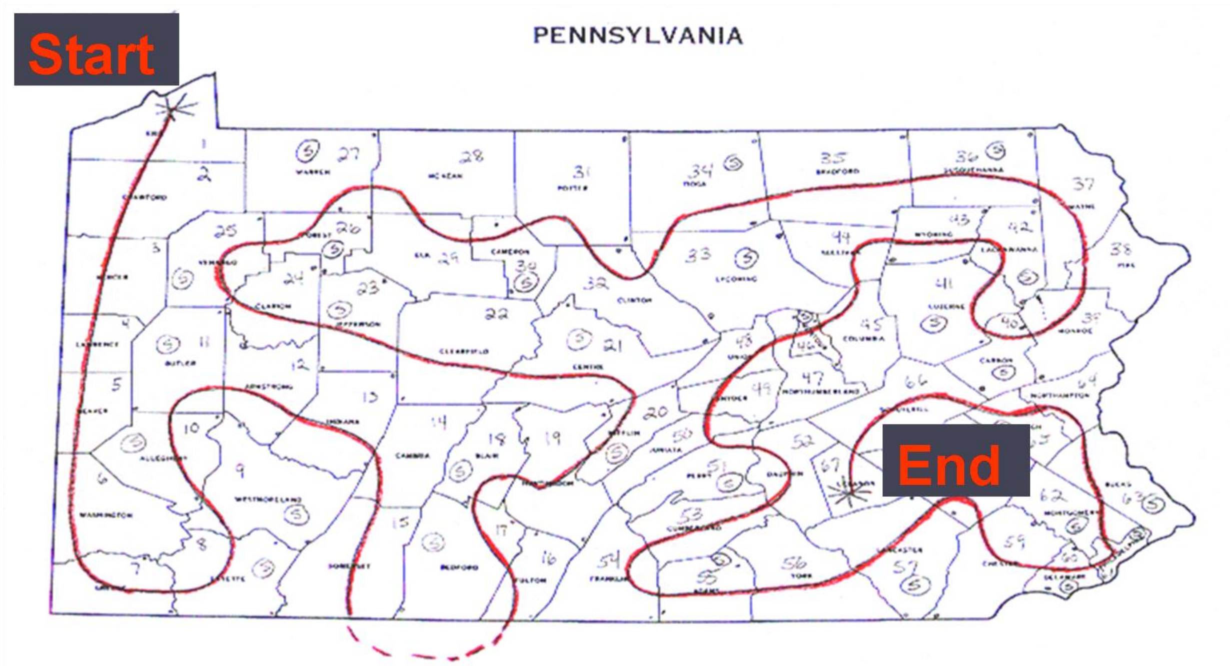
There is a further level of stratification which is applied to the frame. Sub-stratification is the process used to divide the population of sampling units within each stratum equally into categories (substrata). These substrata do not have a definition associated with them like strata do (i.e. 50% or 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 the Land-Use Stratification section 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 makeup.

Figure 1.13 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.

Figure 1.13 County Ordering Used for the Pennsylvania Area Frame



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 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 primary sampling units 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 strata. 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 PSU’s size. The number of sampling units for any given PSU in the strata is:

$$N_{ik} = \left(\frac{S_{ik}}{S_i}\right)$$

where N_{ik} = the number of potential sampling units in the k^{th} PSU from the i^{th} land-use stratum, rounded to the *nearest whole number*,

S_{ik} = the size of the k^{th} PSU from the i^{th} land-use stratum, and

S_i = the target size of sampling units in the i^{th} 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 sampling units. 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 = the number of sampling units in the i^{th} land-use stratum, and

N_{ik} = the number of sampling units in the k^{th} PSU from the i^{th} 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 1.5 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 (described in Chapter II).

NASS employs a concept called replicated sampling which provides several key benefits in the estimation process which are described in the second chapter of this report. Approximately 20 percent of the replicates are rotated out of the sample each year with new replicates taking their place. As a rule of thumb, five replicates are generally used if the sample size for a stratum is less than 50 segments, five or ten replicates if the sample size for a stratum is between 50 and 100 segments, and ten replicates are used if the sample size is between 100 and 200 segments.

The number of substrata is therefore simply the sample size divided by the number of replicates, as is illustrated in Table 1.5.

Table 1.5 Pennsylvania Sample Design Showing the Number of Substrata and Replications

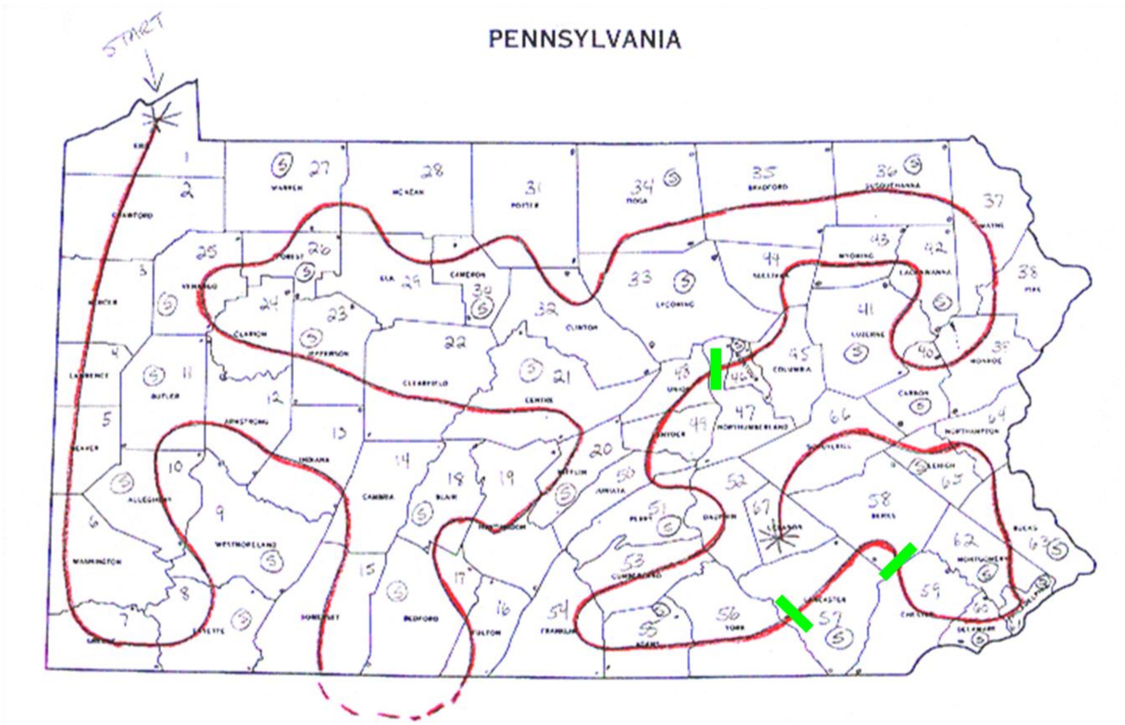
Stratum	Land-Use Stratum	Target Segment Size (Sq Mi)	Population Number of Sampling Units	Number of Substrata	Number of Replications	Sample Size
13	>50% Cult.	1.00	2,800	4	6	24
20	15-49% Cult.	1.00	17,084	14	7	98
31	Ag-Urban	0.25	8,284	1	5	5
32	Residential	0.10	1,814	1	2	2
40	<15% Cult.	2.00	11,344	8	6	48
50	Non-Ag	PPS	45	1	2	2

In this example, there are 2,800 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 "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 less 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 1,220 units and the last will contain 1,224.

Figure 1.14 is a visual of where the substrata breaks fall in stratum 13. Stratum 13 is the highest cultivated stratum and is broken down into four substrata. The green lines are where the substrata breaks occur. Notice that the first substratum covers the largest chunk of land. This is because the western and northern part of the state has much less agriculture than the south-eastern part of the state.

Figure 1.14: Substrata Breaks in Stratum 13 in Pennsylvania



Sub-stratification is implemented to reduce variability in sampling units. The land-use stratification is based on the percent of cultivation. Therefore, while the majority of the segments within a stratum may be intensely cultivated, the agricultural makeup of the segments may differ depending on the location of the segments within the state. Ordering the population of primary sampling units 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.

BIBLIOGRAPHY

Ciancio, N.J., Rockwell, D.A., and Tortora, R.D. (1977): An Empirical Study of Area Frame Stratification. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.

Cotter, J., and Nealon J. (1987): Area Frame Design for Agricultural Surveys. U.S. Department of Agriculture, National Agricultural Statistics Service. Washington, D.C.

Fecso, R. and Johnson, V. (1981): The New California Area Frame: A Statistical Study. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 22. Washington, D.C.

Fecso, R., Tortora, R.D., and Vogel, F.A. (1986): Sampling Frames for Agriculture in the United States. Journal of Official Statistics. Vol. 2, No.3. Statistics Sweden.

Geuder, J. (1983): An Evaluation of SRS Area Sampling Frame Designs: Ordering Count Units and Creating Paper Strata. U.S. Department of Agriculture, Statistical Reporting Service Staff Report No. AGES830715. Washington, D.C.

Geuder, J. (1984): Paper Stratification in SRS Area Sampling Frames. U.S. Department of Agriculture. Statistical Reporting Service, SF&SRB Staff Report No.79. Washington, D.C.

Hanuschak, G. and Morrissey. K. (1977): Pilot Study of the Potential Contributions of Landsat Data in the Construction of Area Sampling Frames. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.

Houseman, E. (1975): Area Frame Sampling in Agriculture. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 20. Washington, D.C.

Huddleston, H. (1976): A Training Course in Sampling Concepts for Agricultural Surveys. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 21. Washington, D.C.

Jessen, R.J. (1942): Statistical Investigation of a Sample Survey for Obtaining Farm Facts, Iowa Agricultural Experiment Station Research Bulletin 304.

Moore, J. (2008): National Agricultural Statistics Service Area Frame Construction and Sampling. U.S. Department of Agriculture, National Agricultural Statistics Service. Washington, D.C.

Pratt, W. (1974): The Use of Interpenetrating Sampling in Area Frames. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.

U.S. Department of Agriculture (1983): Scope and Methods of the Statistical Reporting Service. Publication No. 1308. Washington, D.C.

U.S. Department of Agriculture (1984): Area Frame Analysis Package. Statistical Reporting Service Staff Report. Washington, D.C.

U.S. Department of Agriculture (1987): International Training: Area Frame Development and Sampling. National Agricultural Statistics Service. Washington, D.C.

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

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

CHAPTER II: AREA FRAME SAMPLING

Carrie Davies
Jack Nealon

INTRODUCTION

The purpose of this chapter on area frame sampling is to provide an in-depth presentation of NASS's area frame sampling concepts and procedures. The chapter will begin by discussing NASS's use of replicated sampling. Next, the procedures used to determine the size of the ultimate sampling unit (segment) and the allocation of the sample across and within states are described. The probability models are then detailed for the various methods of selecting the sample segments. An overview of the office procedures applied to select and prepare the sample segments for agricultural surveys are also discussed. In addition, the sample rotation scheme and the costs associated with the sampling process are described. Finally, the estimation approaches used to estimate agricultural production from area frame surveys will be discussed briefly.

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, which was described in the first chapter, 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. That is,

$$n_i = r_i \times s_i$$

where n_i = the number of segments in the sample for the i^{th} land-use stratum,

r_i = the number of replicates for each substratum in the i^{th} land-use stratum,

s_i = the number of substrata in the i^{th} 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 $n_i = r_i \times s_i = 2 \times 3 = 6$ segments, as illustrated in Table 2.1. Notice that a simple random sample of one segment is selected in each substratum for a replicate so that the number of sample segments in a replicate is simply the number of substrata.

Table 2.1: Replicated Sampling Process for a Land-Use Stratum

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

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 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 percent of the replicates in each land-use stratum are replaced annually. The sample design does not rotate exactly 20 percent of the segments because the number of replicates is not always a multiple of five.

   To illustrate how replicated sampling simplifies the sample rotation process, Table 2.2 shows the numbering scheme for a hypothetical land-use stratum with five replicates in each of eight substrata. The first digit in the four 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 five 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.

Table 2.2: Replicated Sampling Process for a Land-Use Stratum

Substratum	Replicate				
	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

3) 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 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.

4) 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.

5) 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 five years for the ACES survey to estimate the completeness of the Census mail list.

6) Variance Estimation: Replicated sampling provides a simple, unbiased method for estimating the sampling variance using replicate means or totals. 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.

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

SEGMENT SIZE

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 preconstruction 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 2.3 is an example of the segment sizes per strata for a typical state.

Table 2.3: Target Segment Sizes

Stratum	Definition	Target Segment Size (square miles)
11	General Cropland, >75% 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

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

The target segment size for open land strata (40's strata) varies the most. In states where good boundaries are available, the target segment size can be one to two square miles. In some areas, like desert, mountainous, or range areas, boundaries are few and far apart. The target segment size in these areas will range from 4 to 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).

For a complete listing of the target segment sizes by stratum for each state, refer to the "Area Frame Design Information" document.

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

1) 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.

2) 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 40's strata consists of forest, desert, range, pasture, etc. Quality boundaries in these strata are more spread apart, making a larger segment size more accommodating.

3) 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 to 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 (10's and 20's strata), farms are relative close together. A segment size of one square mile will result in an average of 3 to 4 interviews. In the lower intensively cultivated strata (40's 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 little if any interviews are done for segments in these strata.

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. 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 program.

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 not on list (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.

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 multipurpose 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 following constraint:

$$V(\hat{Y}_j) \leq v_j$$

Where $V(\hat{Y}_j)$ = the estimated sampling variance for the j^th survey total, and

v_j = the desired or target sampling variance for the j^th survey total.

Assume the following cost function:

$$C(x) = \sum_{i=1}^{l} c_i n_i = \sum_{i=1}^{l} c_i / x_i$$

Where c_i = the average cost per segment in the i^th land-use stratum,

n_i = the number of sample segments in the i^th land-use stratum,

l = the number of land-use strata, and

$x_i = 1/n_i$; $n_i \geq 1$

The problem then reduces to minimizing the cost function subject to the constraints that:

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

Where $a_{ij} = \frac{N_i^2 S_{ij}^2}{v_j + \sum_{i=1}^{l} N_i S_{ij}^2}$

S_{ij}^2 = the square of the standard deviation for the j^th survey item in the i^th land-use stratum,

N_i = the number of segments in the i^th 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:

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

2) The population counts that are calculated after the stratification process.

3) The desired sampling variance for the estimated total of each item, v_j, is established by the Area Frame Section after consultation with others in NASS.

4) The square of the standard deviation, S_{ij}^2, for the j^{th} item in the i^{th} 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 2.4.

Table 2.4: 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

There were 10,912 segments in the national area frame sample. This represents approximately 0.6 percent of the total segments in the population for the nation. The allocation of the sample across land-use strata naturally concentrates the majority of the sample segments in the cultivated strata. Approximately 54 percent of the national sample is in the intensively cultivated strata (10's), 26 percent in the moderately cultivated strata (20's), 16 percent in the less cultivated strata (40's), 3 percent in the ag-urban strata (30's), and 1 percent in the non-agricultural strata (50). Table 2.5 displays the sample allocation by stratum for Minnesota.

Table 2.5: 2009 Sample Allocation for Minnesota

Stratum	Stratum Definition	Number of Reps	Number of Substrata	Number of Segments
11	>75% Cultivated	20	11	220
12	50-74% Cultivated	20	4	80
20	15-49% Cultivated	5	10	50
31	Ag-Urban	2	1	2
32	Commercial	2	1	2
40	<15% Cultivated	5	8	40
50	Non-Agricultural	2	1	2
Total				**396**

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 (40's strata) in some western states. In all other land-use strata in the U.S., equal probability of selection is used. About 96 percent 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 NASS's area frame design. These expressions provide the statistical foundation for area frame sampling.

Equal Probability of Selection: A two-step procedure is used to select sample segments from the selected PSUs when selection probabilities are equal. Recall from the first chapter 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.

Step One: The first step involves the selection of a sample of PSUs 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 k^{th} PSU in the j^{th} substratum from the i^{th} land-use stratum is:

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

where A_{ijk} = the k^{th} PSU in the j^{th} substratum from the i^{th} land-use stratum,

N_{ijk} = the number of sampling units (segments) in the k^{th} PSU from the j^{th} substratum in the i^{th} land-use stratum, and

N_{ij} = the number of sampling units (segments) in the j^{th} substratum from the i^{th} land-use stratum.

Step Two: After the sample of PSUs is drawn, each selected PSU is divided into the required number of segments. The second step of the two-step sampling process involves randomly selecting a segment with equal probability from the selected PSU. That is, the probability of selecting the m^{th} segment given the k^{th} PSU was selected from the j^{th} substratum in the i^{th} land-use stratum is:

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

where B_{ijkm} = the m^{th} segment in the k^{th} PSU from the j^{th} substratum and the i^{th} land-use stratum,

A_{ijk} and N_{ijk} are previously defined.

Therefore, the unconditional probability of selecting the m^{th} segment in the k^{th} PSU from the j^{th} substratum in the i^{th} land-use stratum is:

$$P(B_{ijkm}) = P(A_{ijk})P(B_{ijkm}|A_{ijk})$$

$$= \frac{N_{ijk}}{N_{ij}} \; \frac{1}{N_{ijk}}$$

$$= \frac{1}{N_{ij}}$$

where N_{ij} is as previously defined.

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 fact is illustrated in Table 2.6 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.

Table 2.6: 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

Unequal Probability of Selection: PSUs are selected with unequal probability in less cultivated strata (40's 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 k^{th} PSU in the j^{th} substratum from the i^{th} land-use stratum is:

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

where A_{ijk} = the k^{th} PSU in the j^{th} substratum from the i^{th} land-use stratum,

S_{ijk} = the size (acres) of the k^{th} PSU in the j^{th} substratum from the i^{th} land-use stratum, and

S_{ij} = the size (acres) of the j^{th} substratum in the i^{th} 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. That is,

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

where e_{ijm}= the expansion factor for the m^{th} segment in the j^{th} substratum and the i^{th} land-use stratum,

p_{ijm} = the probability of selecting the m^{th} segment in the j^{th} substratum from the i^{th} land-use stratum,

n_{ij} = the number of segments or replicates in the sample for the j^{th} substratum in the i^{th} land-use stratum.

SAMPLE SELECTION

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

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. A SAS program is run which uses the selection probabilities discussed in the last section to select the chosen PSUs. The program creates a listing of all chosen PSUs.

Personnel in the Sample Select Unit break down the chosen PSUs that have equal probability of selection into segments in ArcGIS9. The National Agricultural Imagery Program (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). These criteria are:

1) 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.

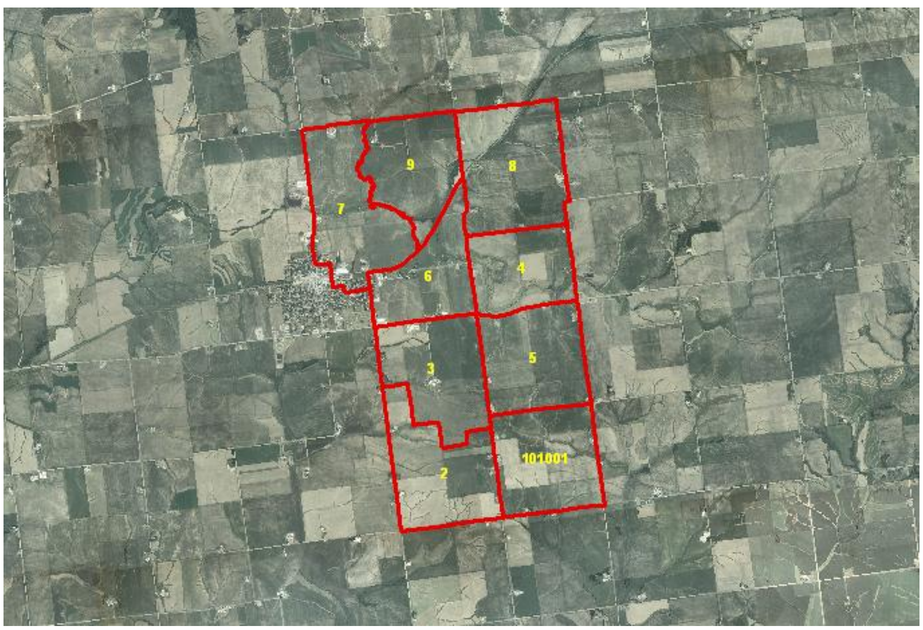
2) Create segments that are as homogeneous as possible with respect to agricultural content. Since crop types are generally not distinguishable on the aerial photography,

homogeneity is usually based on the amount of cultivated land. This criterion reduces the sampling variability among segments in a given substratum.

3) 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 percent 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. Figure 2.1 shows a chosen PSU that has been broken down into segments.

Figure 2.1: Chosen PSU Broken Down into Segments



The PSU in Figure 2.1 is 8.8 square miles in size. The target segment size is 1.0 square miles, so the PSU was broken down into 9 segments using physical boundaries. Then the segments were numbered and a segment was randomly selected in ArcGIS9. The selected segment is in the bottom right-hand corner of the PSU labeled 101001 (according to NASS's segment id scheme). It is 1.01 square miles in size. Once the PSU is broken down and a segment is selected, it is then checked by a reviewer to identify and resolve any errors.

Unequal Probability of Selection: Recall that PSUs selected with unequal probability in less cultivated strata (40's 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 Select 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 to go to the field office.

SAMPLE PREPARATION

Once the sample is selected for a state, the chosen segments must be reviewed, prepared, and printed. These activities are performed by the Sample Preparation Unit. As was the case with the sample selection activities, the work to prepare the samples is carefully reviewed upon completion so that high quality materials are made available to the interviewers.

The process begins with the edit/review phase, when the chosen segment is put through a rigorous review and quality assurance program. This includes examining the segment's boundaries closely, ensuring that high quality, identifiable boundaries were used. Also, areas may be added or subtracted so that the segment meets the stratum definition (percent cultivated). The chosen segment is carefully edited to one and two meter resolution NAIP photography using ArcGIS9 and server technology. The finished version of the chosen segment is subsequently re-projected in ArcGIS9 to obtain final coordinate and acreage values.

The chosen segment is then prepared for the plotting phase, which includes the manual adjustment to contrast, resolution, and colors using bands 3, 2, and 1 on the most recent NAIP imagery. All pertinent boundary labels are utilized as are necessary adjustments to the heading (county name and segment number) and north arrow. Two inset maps are added to the image to help the interviewer locate the chosen segment during data collection. These are:

- The County Outline Map, which shows where the general location of the segment in the county, and

- The Segment Location Map, which shows a closer view of the segment and nearby roads and highways to plan a route to the segment.

The layout window with the fully prepared final version of the chosen segment, as well as accompanying inset segment location map insets, is then captured and saved before sending the digital layout to the plotter for printing. A 24" x 24" image of the segment (8"=one mile scale) plus the two inset maps are then printed onto Kodak photographic paper. Figure 2.2 is an example of the image printout of a chosen segment.

Figure 2.2: Chosen Segment Printout

Once all images for the segments in the sample are printed, they are inventoried through a conclusive quality control phase. They are sorted by state, alphabetized, and checked for errors. The final phase of preparation is the encapsulated lamination of each printout using three millimeter water resistant laminates. The finished laminated printouts are then carefully packed and sent to their respective state office destinations to be used for data collection.

The final activity performed in the sample preparation process for all segments is the creation of a segment-level data file called the area frame master. The master contains pertinent sampling information on each segment such as the land-use stratum, substratum, segment and replicate numbers, the expansion factor and digitized acreage. This information is later used along with the survey data for computing the survey statistics and standard errors.

SAMPLE MAINTENANCE

NASS has an active program for reviewing the area frame sample to uncover any problems that might affect the quality of the survey results. This review consists of four components which will now be described.

1) Problem Segments: Two situations sometimes surface for which corrective measures are taken to control non-sampling errors during the data collection phase. The AFS can take corrective action for these situations provided strict statistical standards are followed so that potential biases resulting from the actions are negligible. Segments that fit either one of the following two situations are referred to as "problem" segments:

    - Too Many Tracts: If a segment has more than 20 tracts of which ten or more are agricultural and data collection time exceeds 12 hours, it is considered a problem segment. To rectify this problem, the Sample Preparation Unit divides the segment into halves, thirds, or quarters with input from the Field Office. When the segment is divided, the resulting parcels should be as equal in the following as possible:

        - Agricultural use of land,
        - Number of agricultural and non-agricultural tracts, and
        - Total Area.

      One of the parcels is then randomly selected with equal probability of selection for use in future surveys. The selection probability for the partial segment is then multiplied by 1/p. That is, the expansion factor is multiplied by p.

    - Boundary Problem: The segment's boundaries on the photo enlargement do not follow features that are identifiable by the respondent which causes reporting errors. Action is taken only if the boundaries can be adjusted without introducing bias. Small adjustments to the boundaries (less than five percent of the land area in the segment) can only be made if:

        - Neither a land-use stratum boundary nor a PSU boundary is affected.

        - Minor acreage adjustments resulting from changing a boundary are offsetting to the extent possible. All changes are documented and analyzed for each state.

2) Imagery Printout Replacement: Each year, field office personnel have the opportunity to request new printouts for segments having out of date imagery. Interviewing problems can occur due to major land use changes such as new housing developments. State offices also request replacements for segments with damaged or lost imagery.

3) Post-Survey Analysis: The AFS statisticians use SAS Software to produce graphical and statistical analyses of each area frame and area frame sample using survey data. The major graphical and statistical analyses provided for a state are:

   - Percentage of the segments in each land-use stratum satisfying the stratum definition. This information is used to evaluate the area frame stratification.

   - Comparison of the reported, digitized, and target segment sizes in each land-use stratum.

   - Descriptive statistics for major agricultural commodities for each land-use stratum.

   - Distribution of the survey estimates in each land-use stratum for each of the five sample rotation groups.

   - Survey estimates and coefficients of variation at the substratum, land-use stratum and state levels.

   - These analyses sometimes uncover non-sampling errors, or suggest improved sample allocations and design alternatives which could benefit future surveys and area frame development and sampling activities.

4) Historic Information: Pertinent historic information about each area frame and area frame sample is maintained in an automated system. Basically there are three types of information for each state. These are:

   - General information, which is provided for each land-use stratum such as the stratum definition, the number of substrata, the total number of segments in the stratum, and the number of segments in the sample.

   - Survey estimates and coefficients of variation since 1979 at the state and land-use stratum levels, which are available for key survey items. Also, information on the number of operations, agricultural operations and resident farm operations is provided for each land-use stratum. Finally, non response counts and survey estimates since 1979 are given for each of the five sample rotation groups.

- Historic information, which is provided on the number of problem segments submitted each year by each state and the number for which corrective action was taken.

This information is useful for providing a statistical history to assess if a state needs a new or updated area frame and to provide information on the sample for evaluation purposes.

SAMPLE ROTATION

As mentioned earlier, 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 five 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 five, especially in urban, commercial, and non-agricultural strata where the sample size is small (usually two replicates).

Table 2.7 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 one will be replaced with all new segments in 2010 and will stay in the sample for five years. In 2015, the segments in Replicate one will be replaced again. Likewise, new segments will rotate into replicate two 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 percent of the sample.

Table 2.7: Rotation of Replicates Depending Upon the Number of Replicates

Number of Replicates	Year				
	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

All segments are not in the sample exactly five 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 five years as shown in Table 2.8. 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.

Table 2.8: 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	1	
2026																		3	4	5	1	2
2027																			4	5	1	2
2028																				5	1	2
Number of Years in Sample	1	2	3	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	3

The national area frame sample size is approximately 11,000 segments. The total number of segments rotated each year is approximately 3,000. 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 2,200 segments being selected based on a 20 percent rotation of the remaining 15,000 or so segments. Therefore, approximately 27 percent of the national area frame sample is based on newly selected segments each year.

SAMPLE ESTIMATION

This final section will briefly discuss the approaches used to estimate agricultural production with an area frame sample of segments. 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 $1,000 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 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 x (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 sub-stratum. 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 = the number of land-use strata,

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

n_{ij} = the number of segments sampled in the j^{th} substratum in the i^{th} land-use stratum,

e_{ijk} = the expansion factor or inverse of the probability of the selection for the k^{th} segment in the j^{th} substratum in the i^{th} land-use stratum,

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

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

t_{ijkm} = the tract value of the characteristic, Y for the m^{th} tract in the k^{th} segment, j^{th} substratum, and i^{th} 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

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

a_{ijkm} = the weight for the m^{th} tract in the k^{th} segment, j^{th} substratum, and i^{th} land-use stratum.

The following weight is currently in use:

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

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:

$$V(\hat{Y}) = \sum_{i=1}^{l} \sum_{j=1}^{s_i} \frac{\left(1 - {}^{1}/_{e_{ij}}\right)}{\left(1 - {}^{1}/_{n_{ij}}\right)} \sum_{k=1}^{n_{ij}} \left(y'_{ijk} - y'_{ij\bullet}\right)^2$$

where $y'_{ijk} = e_{ij} y_{ijk}$,

$$y'_{ij\bullet} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y'_{ijk}$$

The standard error is then:

$$SE(\hat{Y}) = \{V(\hat{Y})\}^{1/2}$$

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' 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.

BIBLIOGRAPHY

Bethel, J. (1986): An Optimum Allocation Algorithm for Multivariate Surveys. U.S. Department of Agriculture, Statistical Reporting Service, SF&SRB Staff Report No. 89. Washington, D.C.

Cotter, J., and Nealon J. (1987): Area Frame Design for Agricultural Surveys. U.S. Department of Agriculture, National Agricultural Statistics Service. Washington, D.C.

Fecso, R. and Johnson, V. (1981): The New California Area Frame: A Statistical Study. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 22. Washington, D.C.

Ford, B.L. (1975): Rotation Group Effects in SRS Surveys. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.

Ford, B.L., Nealon, J.P., and Tortora, R.D. (1986): Area Frame Estimators in Agricultural Surveys: Sampling Versus Non-sampling Errors. Agricultural Economics Research. Vol. 38, No.2. Journal of the U.S. Department of Agriculture. Washington, D.C.

Geuder, J. (1981): An Evaluation of the New York Area Sampling Frame. U.S. Department of Agriculture, Statistical Reporting Service, Staff Report No. AGES810929. Washington, D.C.

Geuder, J. (1984): Paper Stratification in SRS Area Sampling Frames. U.S. Department of Agriculture. Statistical Reporting Service, SF&SRB Staff Report No.79. Washington, D.C.

Houseman, E. (1975): Area Frame Sampling in Agriculture. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 20. Washington, D.C.

Huddleston, H. (1976): A Training Course in Sampling Concepts for Agricultural Surveys. U.S. Department of Agriculture, Statistical Reporting Service Publication SRS No. 21. Washington, D.C.

Moore, J. (2008): National Agricultural Statistics Service Area Frame Construction and Sampling. U.S. Department of Agriculture, National Agricultural Statistics Service. Washington, D.C.

Nealon, J.P. (1984): Review of the Multiple and Area Frame Estimators. U.S. Department of Agriculture, Statistical Reporting Service, SF &SRB Staff Report No. 80. Washington, D.C.

Pratt, W. (1974): The Use of Interpenetrating Sampling in Area Frames. U.S. Department of Agriculture, Statistical Reporting Service Staff Report. Washington, D.C.

U.S. Department of Agriculture (1983): Scope and Methods of the Statistical Reporting Service. Publication No. 1308. Washington, D.C.

U.S. Department of Agriculture (1984): Area Frame Analysis Package. Statistical Reporting Service Staff Report. Washington, D.C.

U.S. Department of Agriculture (1987): International Training: Area Frame Development and Sampling. National Agricultural Statistics Service. Washington, D.C.

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

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