Survey Research Methods (2017)
Vol. 11, No. 3, pp. 289-306
doi:10.18148/srm/2017.v11i3.6725

ISSN 1864-3361
http://www.surveymethods.org

Integration of the Cropland Data Layer Based Automatic Stratification Method Into the Traditional Area Frame Construction Process

Claire G. Boryan
USDA National Agricultural Statistics Service
Washington D.C., U.S.A.

Zhengwei Yang
USDA National Agricultural Statistics Service
Washington D.C., U.S.A.

A new automatic stratification method utilizing United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) geospatial Cropland Data Layers was recently implemented in NASS operations. Recent research findings indicated that using the automated stratification method significantly improved area sampling frame stratification accuracies in intensively cropped areas (>15% cultivation) and overall stratification accuracies when compared to traditional stratification based on visual interpretation of aerial photography or satellite data, while reducing the cost of area frame construction (Boryan et al., 2014). Though the new automated stratification method has improved stratification efficiency, objectivity, and accuracy in the intensively cropped areas, it inherits the Cropland Data Layer classification errors and has lower accuracies in low or non-agricultural areas. This implies that the automated stratification process is not a perfect solution to directly replace the NASS traditional stratification method for area frame construction operationally. This paper describes a hybrid approach: an operational area frame construction process that integrates the automated stratification results with manual editing/review methods. New 2014–2015 NASS area frames for South Dakota, Oklahoma, Arizona, New Mexico, Georgia, Alabama and North Carolina were successfully built using the new integrated operational process. The improvement measures used to assess the traditional, automated and hybrid methods for area frame construction include: 1) area frame stratification accuracy; 2) mean stratum primary sampling unit size, mean stratum percent cultivation and stratum standard deviations; 3) the variances of key estimators; and 4) labor cost. The seven updated area frames delivered significant improvements in objectivity, operational efficiency, and frame accuracy, based on 2013–2015 June Area Survey reported data.

Keywords: Area sampling frame, automated stratification, cropland data layer, cultivated layer, land cover-based stratification, land use estimation

1 Introduction

Area sampling frames are used for a variety of surveys ranging from crop acreage and yield, soils, livestock, as well as forest and natural resource inventories and are the foundation of the agricultural statistics program of the National Agricultural Statistics Service (NASS) and many other statistical survey programs around the world (Arroway, Abreu, Lamas, Lopiano, & Young, 2010; Cotter, Davies, Nealon, & R., 2010; Ford, Nealon, Tortora, et al., 1986; Nusser & House, 2009; Vogel, 1995). Area sampling frames or area frames have been used in NASS since 1954 as a primary tool for conducting surveys to gather diverse agricultural information, notably planted acreage of major crops, economic, chemical use and other information.

NASS's primary area frame based survey is the June Area Survey in which approximately 11,000 one square mile sample segments are visited by survey enumerators at the beginning of each growing season to collect crop type and acreage information. Estimates of crop acreage and livestock inventories are based on these data. The June Area Survey is a component of data collection and estimation for more than 14 official NASS statistical publications. In recent years, NASS has become increasingly relied upon to provide timely, accurate and useful statistical information for public use. The wide range of crop, livestock, and economic statistics published in the NASS official reports helps to reduce the risk of instability for agricultural operations. The June Area Survey plays a critical role in many of the NASS programs (USDA FSA, 2017b).

NASS also uses area frames for follow-on surveys such as the Objective Yield Survey. Further, every five years, additional sampling units or segments are added to the June

Contact information: Claire G. Boryan, USDA National Agricultural Statistics Service, 1400 Independence Ave, Washington, D.C. 20250, USA, (email: Claire.Boryan@nass.usda.gov).

Area Survey for the Agricultural Coverage Evaluation Survey. Data collected from the June Area Survey and Agricultural Coverage Evaluation Survey segments, are used to measure the completeness of the Census of Agriculture mail list. For these applications and more, area frames are considered "the backbone to the agricultural statistics program of the National Agricultural Statistics Service"(Cotter et al., 2010).

The accuracy of NASS survey statistics depends on the quality of the NASS area frames and therefore the techniques used in their construction. The NASS area frames are based on a stratification of U.S. land cover which classifies land into different agricultural intensity groups or strata based on percent cultivation in a given land parcel. The traditional stratification of land cover has been conducted using visual interpretation of aerial or satellite data, and topographic maps since the 1950s. This popular method requires a manual, subjective and labor intensive process. The traditional stratification method does not utilize existing land cover data such as the geospatial Cropland Data Layer (Boryan, Yang, Mueller, & Craig, 2011; Han, Yang, Di, & Mueller, 2012) in an automated, objective and efficient manner to stratify area frames.

Statistical offices around the world commonly use subjective procedures, similar to NASS's traditional method to stratify area frames for agricultural surveys (Cotter et al., 2010; Cotter & Tomczak, 1994; FAO, 1998; Hanuschak & Morrissey, 1977). The European Union's Monitoring of Agriculture with Remote Sensing project used visual interpretation techniques to conduct stratification for the Land Use and Cover Area Frame Statistical Survey (LUCAS) and proposed using a regular grid of square segments instead of physical boundaries for segment delineation to reduce cost (Gallego, Delincé, & Carfagna, 1994). The LUCAS sample units are points defined as 3m diameter circles (Gallego & Delincé, 2010). The original objective of the LUCAS was to provide annual European crop estimates. The Italian Agricultural Survey (AGRIT) program uses an area frame created from a regular grid of points with a resolution of 500m. Point sampling is used for data collection for the multi-purpose survey to derive combined estimates of crop acreage, yields and land use (Benedetti, Piersimoni, & Postiglione, 2015; Carfagna & Gallego, 2005). The Utilization du territoire (TER-UTI) is an annual survey conducted by the French Ministry of Agriculture, in which a grid area frame and sample points are used to collect land use and land cover information as well as to derive an estimate of land cover change (Benedetti et al., 2015). Workneh, Tylka, Yang, Faghihi, and Ferris (1999) used an area frame for evaluation of the prevalence of brown stem rot in the north central U.S.. The European Union's Monitoring Agriculture through Remote Sensing group also used an area frame for improvement of agricultural ground survey estimates (Tsiligirides, 1998). Faulkenberry and Garoui compared the utility of frequently used estimators based on an area frame utilized in agricultural surveys (Faulkenberry & Garoui, 1991). Pradhan (2001) used an area frame for development of crop area estimation at a regional level in the Islamic Republic of Iran.

Geospatial data such as aerial photography and satellite data are commonly used to stratify area frames, delineate primary sampling units, delineate segments, identify point positions, conduct data collection, and identify segment locations, using visual interpretation and other manual techniques. Martinez (2013) provides further details on the subjective use of geospatial data for area frame construction by countries around the world.

Utilizing remotely derived geospatial land cover data for stratification of an area frame has been gradually investigated since 1990. The European Coordination of Information on the Environment (CORINE) Land Cover 2000 map was tested for use in stratification in Spain, but difficulties were encountered when multiple Corine Land Cover polygons crossed individual segment boundaries causing designation of segments to multiple strata. The Corine Land Cover data were evaluated as a covariable to define strata based on an agricultural intensity index (Gallego, Carfagna, & Peedell, 1999). Hansen and Wendt tested using United States Geological Survey National Gap Analysis Program classifications for stratification of USDA Forest Service's Forest Inventory Analysis plots in Indiana and Illinois (Hansen & Wendt, 2000). They noticed the increased precision of forest inventory estimates. Dunham et al. further evaluated stratifications using an automated method based on the 1992 National Land Cover Data Set and using visual analysis of photo imagery in western Oregon, and concluded that forest inventory estimation accuracies were similar but cost was reduced using the automated land cover based approach (Dunham, Weyermann, & Azuma, 2002). Liknes, Nelson, and McRoberts (2004) used land cover classification results from Moderate Resolution Imaging Spectroradiometer data for Forest Inventory Analysis stratification and found that the results were inferior to those based on National Land Cover data due to a coarser spatial resolution.

Perry (2000) used visual interpretation of Landsat Thematic Mapper data and soil maps for substratification of the NASS Arkansas area frame. A manual reordering procedure was conducted to group area frame primary sampling units into substrata which resulted in a reduction in sampling variance for major crops when compared with the traditional NASS serpentine ordering process. Implementation was impractical due to the statistical expertise and manual labor required to conduct reordering (Perry, 2000). Perry and Gentle (2000) further developed an automated procedure based on simulated annealing, which was tested in the intensive agricultural land use category of the Arkansas area frame. Variances were further reduced but implementation remained "impractical" based on available resources.

McRoberts, Wendt, Nelson, and Hansen (2001) conducted post stratification forest area estimation for the states of Indiana, Iowa, Minnesota and Missouri based on the 1992 National Land Cover Data set (Vogelmann et al., 2001) classes and observation of forest inventory plots. They concluded that the National Land Cover Data set 1992 provided an effective means of post stratification, which resulted in reduced estimated variances for estimators of forest land area over post stratification based on visual interpretation (McRoberts et al., 2001; McRoberts & Wendt, 2002). Likens et al. further evaluated the use of the USDA NASS 2005 Wisconsin Cropland Data Layer and the 1992 National Land Cover Data for post stratification of Forest Inventory Analysis plots. The results indicated that the 2005 Cropland Data Layer outperformed the 1992 National Land Cover Data for post stratification (Liknes, Nelson, Gormanson, & M. Hansen, 2009). These studies indicated that utilizing geospatial land cover classification data for area frame stratification might result in improved estimates and significantly reduce stratification cost. However, utilizing remotely derived geospatial land cover data for stratification of an agricultural area frame has not been used in operations. Using a combination of automatic and manual procedures for area frame stratification in an operational setting in order to exploit the benefits of both automatic and manual procedures is yet to be investigated.

In recent years, gains have been made in the quantity and quality of geospatial data and technology available at low or no cost to the public (ESA, 2017a, 2017b; Han et al., 2012; QGIS, 2017; USGS, 2017b). Based on the previously described research and these recent technological advancements, improvements in area frame accuracy and a reduction in labor cost appeared feasible by adopting a more automated area frame stratification method based on moderate resolution (30 meter) geospatial data. Consequently, in an effort to improve the NASS area frame stratification procedures, Boryan and Yang used U.S. Cropland Data Layer (30 meter) geospatial cropland cover data derived from remotely sensed satellite imagery to automatically and objectively stratify land cover in the U.S. based on percent cultivation (Boryan, Yang, Di, & Hunt, 2014). The new automated method stratified the U.S. state level area sampling frame by automatically calculating percent cultivation at the primary sampling unit level based on the NASS Cultivated Data Layer, a composite data product derived from five years of Cropland Data Layers (Boryan, Yang, & Di, 2012). Boryan and Yang found that the automated Cropland Data Layer stratification method significantly improved stratification accuracies in intensively cropped areas and performed less well in non-agricultural areas as compared with traditional stratification based on visual interpretation for five test states. Though this method has improved stratification efficiency, objectivity, and accuracy in the intensively cropped areas, it has lower accuracies in low or non-agricultural areas. This implies that the Cropland Data Layer based automated stratification process is not a perfect solution to directly replace the traditional NASS stratification method for area frame construction.

This paper presents a technical route for integrating the new Cropland Data Layer based automated stratification method into the NASS area frame construction operational process. The newly integrated operational stratification process utilizes results of both traditional and automatic Cropland Data Layer stratification methods and further refines area frames with manual editing and review procedures. Area sampling frame refinement includes reducing primary sampling unit sizes and/or redefining primary sampling units to improve stratum accuracy and homogeneity. The remainder of the paper is organized as follows. Section 2 provides background on area frames and stratification research. Section 3 defines the research scope and data used in the study. Section 4 describes the methods and the technical route for integrating the automated stratification method into the NASS traditional operational area frame construction process, implementation details, and the metrics for assessing the experimental results. Section 5 presents the experimental results and highlights gains in accuracy and efficiency that have been achieved with the adoption of the new automated stratification method. Finally, this paper summarizes the benefits achieved by adapting the new automated stratification method, based on the Cropland Data Layer, operationally for NASS area frame construction.

2 Background

2.1 NASS Area Sampling Frames

NASS's area frame land use stratification divides land area using physical boundaries on the ground (roads, railroads and rivers) into broad land-use categories and is known to improve efficiency for statistical sampling and estimation. In the construction of a NASS area frame, general cropland (based on percentage cultivation), agriculture/urban, residential/commercial, and non-agriculture are the commonly identified land covers. The strata completely partition the land area within a state. The agricultural stratum definitions vary between states depending on the type and intensity of agricultural production. Table 1 illustrates NASS typical land-use stratification codes and definitions. It should be noted that states with strata 11 and 12 do not have a stratum 13. All are included on Table 1 because our test includes states that use strata 11 and 12, and other states that use stratum 13. Once stratum definitions are assigned, all land is subdivided into primary sampling units, which are designed to reduce labor cost in random sampling. Only selected primary sampling units are further subdivided into segments or sample units, and a segment is randomly selected from each allocated primary sampling units for enumeration (Cotter et al., 2010). Stratification at the primary sampling unit level implies that all segments in a given primary sampling unit belong to the same stratum. However, enumerating the entire primary sampling unit is cost prohibitive. Therefore, operationally in NASS, one segment is randomly selected to represent the entire primary sampling unit. In reality, however, the selected segments may not necessarily have the same percent cultivation range as the primary sampling unit stratum definition due to land cover heterogeneity in a primary sampling unit. This indicates that sampling from segments adds increased variability. Therefore, stratum homogeneity is critical for the performance of the NASS area frames.

Manual inspection and editing can improve upon the autostratification results and consequently the quality of the area frame by reducing errors that result from 1) primary sampling unit heterogeneity, 2) changes in land cover over time, and 3) errors in the Cropland Data Layer classifications. First, it is common that primary sampling units, particularly in stratum 12 (>50-75% cultivated) and stratum 20 (15-50% cultivated) are not evenly divided into homogeneous segments. The cultivated land tends to be clustered. Consequently, when the primary sampling units are automatically stratified, although accurate (meet stratum definition) at the primary sampling unit level, they cannot be evenly divided into homogeneous segments. Second, boundary and population changes occur over time that require manual editing thus enabling the area frame to more accurately reflect current conditions. Third, the Cropland Data Layers are not 100% accurate although they are highly accurate (85-95% for major crops in large production states) (Boryan et al., 2011). However, local areas can be impacted by Cropland Data Layer errors of omission and commission that can be identified and accounted for through manual inspection, particularly in low cultivation areas.

2.2 NASS Cropland Data Layer

The NASS geospatial Cropland Data Layers are annually updated 30-56 meter raster-formatted, geo-referenced, crop-specific land cover classifications as shown in Fig 1. Rulequest Research's See5 Decision Tree software is used to perform supervised classifications of satellite imagery for all 48 conterminous states. Currently, the satellite images used for Cropland Data Layer production include Landsat 8 and Disaster Monitoring Constellation satellite data. Digital elevation, percent canopy, and percent impervious data are used as auxiliary data for Cropland Data Layer classification. The USGS National Land Cover Data and USDA Farm Service Agency Common Land Unit & Administrative 578 data sets are used for non-crop and crop type training and validation ground truth data. The Cropland Data Layer thematic map includes over 110 different crop categories. The Cropland Data Layer classifications were first produced at the state level in 1997 with one state. The NASS Remote Sensing Estimation Program has expanded to include production of Cropland Data Layers for all 48 US conterminous states for 2008-2016. Crop mapping accuracies for major crops in large production states, in historic Cropland Data Layers, range from 85-95% (Boryan et al., 2011; USDA FSA, 2017c). The 1997–2016 archive of Cropland Data Layer is publically available from NASS's online geospatial application CropScape (Han et al., 2012).

2.3 NASS Cultivated Layer

The Cropland Data Layers provide the unique opportunity to use annually produced and highly accurate crop specific geospatial data for automated area frame stratification. State and national scale 30 m raster cultivated land cover information can be derived from the NASS Cropland Data Layers. For operational purposes, five years of Cropland Data Layers are combined to create a national scale "Cultivated Layer" which can be applied directly for the area frame stratification. The methodology used to create the NASS Cultivated Layers is outlined in (Boryan et al., 2012). The NASS Cultivated Layer is a highly accurate characterization of cultivated land across the 48 conterminous states in the U.S. Unlike the original Cropland Data Layers, which include more than 100 different crop categories, the Cultivated Layer includes only two categories (non-cultivation and cultivation). The NASS 2016 Cultivated Layer was validated using 2012-2016 Farm Service Agency Common Land Unit data (Table 2).

3 Scope And Data

The integration of Cropland Data Layer based automated stratification into the NASS area frame construction process is studied for improving its operational use. The new operational procedures are currently being used to construct all State area sampling frames. In this paper, the integrated process which includes both automated stratification and manual primary sampling unit editing and review are described. The study area for the new area frame construction process assessment include the states of South Dakota, Oklahoma, Arizona, New Mexico, Georgia, Alabama and North Carolina as illustrated in Fig. 2. These seven state area frames were constructed previously using traditional procedures, and revised from 2014–2015 as new area frames using the new integrated area frame construction process which hybridizes the automated and manual processes. The selection of the seven states for area frame revision was based on an analysis and ranking of the following variables with the associated weights: 1) age of the area frame (0.056); 2) June Area Survey segments not meeting area frame primary sampling unit strata definition (0.292); 3) land in farms acreage change (0.102); 4) crop coefficients of variation ranking based on the state's relative variability and importance to the national corn, soybean, spring wheat, winter wheat or cotton planted acreage estimates (0.259); 5) potential improvement using the Cropland Data Layer (0.213); and 6) variance of data

Table 1
Typical Land-Use Stratification Codes And Definitions Represented In The NASS Area Sampling Frames

Land-Use Strata Codes	Codes & Strata Definitions
11	General Cropland, greater than 75% cultivated.
12	General Cropland, 51-75% cultivated.
13	General Cropland , greater than 50% cultivated.
20	General Cropland, 15-50% cultivated.
31	Ag-Urban, residential mixed with agriculture, more than 100 dwellings per square mile.
32	Residential/Commercial, more than 100 dwellings per square mile, no cultivation.
40	Less than 15% cultivated (e.g. rangeland/forest).
50	Non-agricultural (e.g. military bases, airports, national and state parks).
62	Water

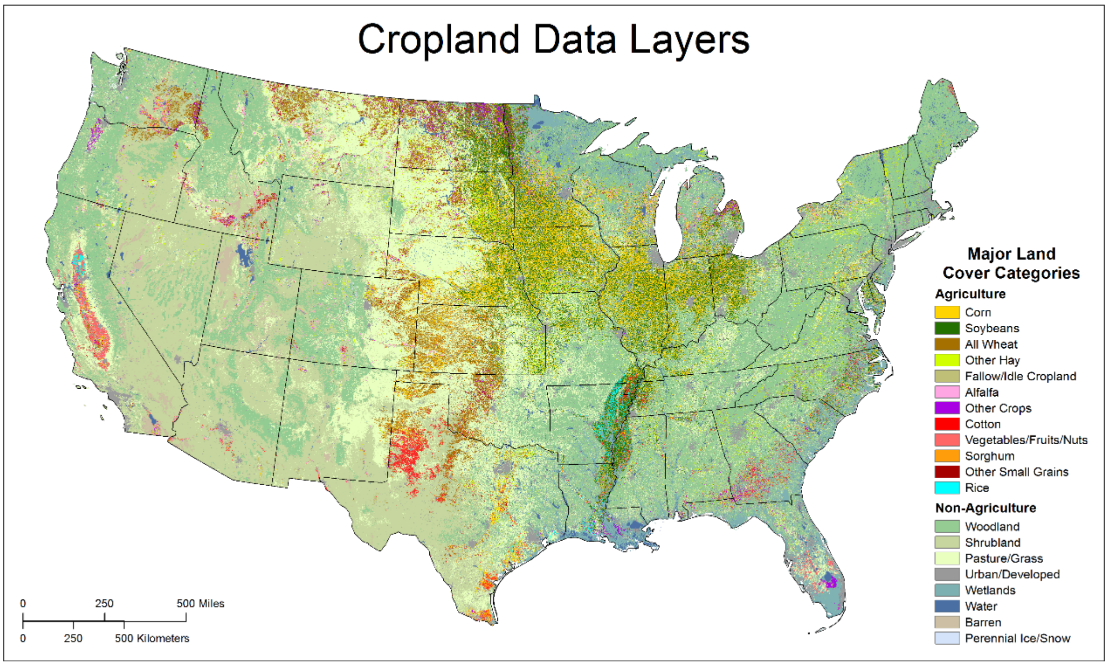

Figure 1. NASS Cropland Data Layers

Table 2
NASS 2016 cultivated layer accuracies

Categories	Producer Accuracy	Omission Error	Conditional Kappa	User Accuracy	Commission Error	Conditional Kappa
Non - Cultivation	89.98%	10.02%	0.891	88.61%	11.39%	0.876
Cultivation	98.96%	1.04%	0.876	99.10%	0.09%	0.891

Source: USDA FSA (2017e).

from the "not on list" (NASS list frame) components of multiple frame indications which were evaluated for number of farms, land in farms and cropland (0.077). The order in which the State area frames were revised was based on their ranking score. The seven states selected for revision had the top seven scores (Cotter et al., 2010; Davies, Hunt, & Willen, 2014).

As part of the revision process, the primary sampling unit boundaries in the original area frames, in ESRI ArcGIS shape file format, are used as the basis for the automated stratification procedure. The original area frame construction procedure includes 1) manually dividing a state's land cover into small land parcels (primary sampling units) which are six to eight square miles in size for highly cultivated land and 20 square miles in size for low agricultural areas and 2) defining stratum definitions for all primary sampling units by visual interpretation of satellite or aerial imagery. An example of primary sampling unit parcels is shown in Fig. 3 (left). As part of the new hybrid process, the NASS state Cultivated Layer, is applied directly to the existing area frames' primary sampling units for stratification, using the automated stratification method for percent cultivated calculation. The NASS state Cultivated Layers are illustrated in the green background in the seven states in Fig. 2. The automated stratification procedure is illustrated in Fig. 3 with the NASS Cultivated Layer (Fig.3 - left) underlying primary sampling units with calculated percent cultivation values. Fig 3 (right) illustrates the strata labels from Table 1 that are applied based on the automated percent cultivation calculation.

A variety of additional ancillary data are used in the area frame manual review and editing process. Single year Cropland Data Layers are used to manually evaluate land cover changes. Landsat 8 satellite imagery acquired during the growing season (June, July and August) are used for visual crop or cultivated land identification (USGS, 2017a). National Agricultural Imagery Program (NAIP) aerial photography, one meter aerial data acquired by the USDA Farm Service Agency, is also used for visual identification of crops. The most recent acquisitions of the NAIP data are used for editing primary sampling unit boundaries(USDA FSA, 2017a). United States Geological Survey's 7.5 minute series quadrangles (1:24,000 scale) and (1:100,000 scale) topographic maps provide information regarding the location of streams, roads, section lines, park boundaries, etc. These topographic maps are used to confirm primary sampling unit boundaries (USGS, 2017c). USDA Farm Service Agency's Common Land Unit data are standardized Geographic Information Systems data layers, which capture agricultural information about the nation's farms and fields that are established to support farm commodity, conservation programs and disaster response. Common Land Unit data are individual contiguous farming parcels which are updated every growing season when farmers report crop type and acreage

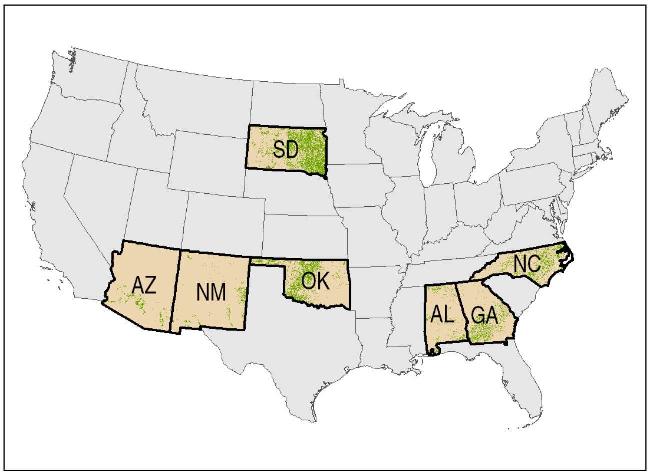

Figure 2. The area frames for South Dakota (SD), Oklahoma (OK), Arizona, (AZ), New Mexico (NM), Alabama (AL), Georgia, (GA) and North Carolina (NC) were revised and implemented from 2014–2015 using the new Cropland Data Layer-based automated methodology in combination with manual editing techniques.

for their fields to over 2,300 Farm Service Agency county offices. The Farm Service Agency Common Land Unit data are used to spot check and confirm areas under cultivation (Heald, 2002; USDA FSA, 2017e). U.S. Census Block data are used to identify population centers (U.S. Census Bureau, 2017). State-specific ancillary data sets are also used to identify and confirm population centers. State Geographic Information Systems layers, Bing Maps (2017), and Google Earth Street View Data (2017) are used to identify and confirm primary sampling unit boundaries (roads, fences).

4 Methods

To increase area frame accuracy further, the newly developed automated stratification method was integrated with manual editing/review procedures and resulted in a new hybrid operational process, which enabled significant accuracy improvements in both high and low agricultural areas. In this new operational process, existing state area frames are revised instead of starting over from scratch and the Cropland Data Layer automated stratification method is used as the primary stratification tool. In the process, existing area frame stratification definitions are compared with the results of automated stratification and the primary sampling units that have potentially changed stratum definitions are identified and targeted for manual inspection. While automated stratification alone achieves higher accuracies and improved stratum homogeneity than traditional stratification in the highly cultivated strata (Boryan et al., 2014), further improved accuracies can be achieved with additional manual inspection and editing. Primary sampling units are purposely reduced

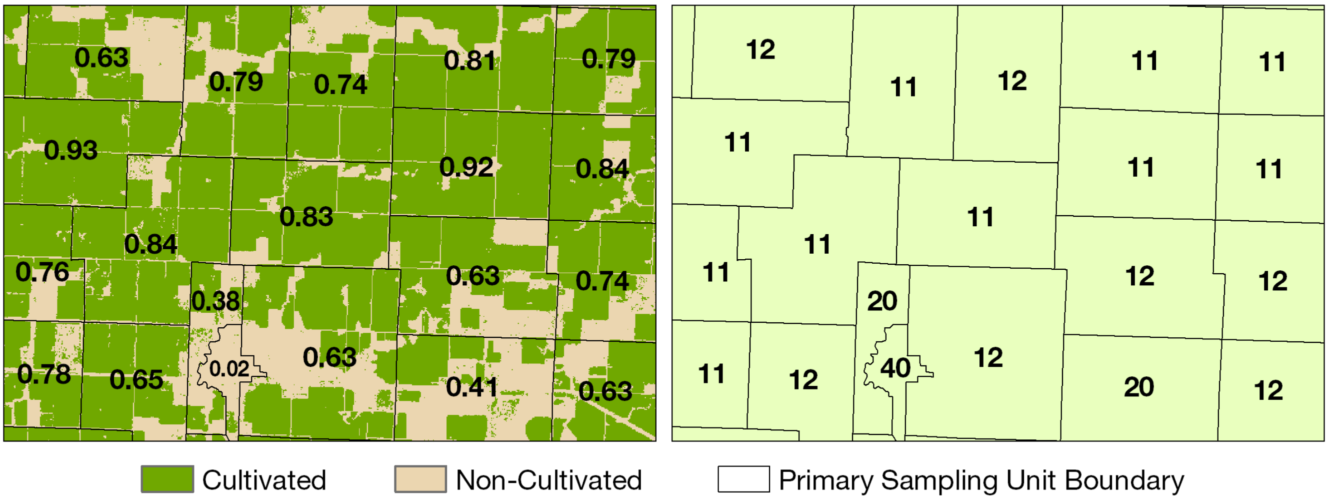

Figure 3. *Left:* Area frame primary sampling units with automatically calculated percent cultivation values, overlaying the NASS Cultivated Layer. *Right:* Primary sampling units are labeled with a stratum category based on the percent cultivation calculation and state specific stratum definitions

in size to improve segment level accuracy, and stratum definitions are changed based on the auto-stratification results and manual review. This new hybrid process integrates the new automated stratification method followed by a targeted manual stratification refinement.

The flowchart as shown in Fig. 4 describes the hybrid operational area frame construction process, and integrates automated stratification as the primary stratification tool. The existing state area frames, produced using the traditional stratification method, are reused. The existing state area frame primary sampling unit boundaries are used with automated stratification to produce revised area frames. The manual area frame refinement procedures are followed, after comparing the new results with the existing traditional area frame, and the identified stratum definition-changed primary sampling units are edited as needed. In Fig. 4, the data and steps used to perform automated stratification are highlighted in blue and the manual review and editing steps are highlighted in green. The detailed steps to conduct the hybrid operational process are as follows:

1. Automated Stratification Steps (Blue boxes in Fig. 4)

   1. Derive a state level Cultivated Data Layer from the most recent five years of Cropland Data Layers by grouping all crop categories into one cultivated (crop) category and assigning the corresponding pixels with a value of "1" while grouping the remaining categories into one non-crop category and assigning the corresponding pixels a value of "0".
   2. Load an individual area sampling frame primary sampling unit boundary layer
   3. Load a state Cultivated Data Layer
   4. Overlay an area frame primary sampling unit boundary on the Cropland Data Layer-based Cultivated Layer.
   5. Compute percent cultivation of each area sampling frame primary sampling unit by counting the total number of pixels with value "1" (cultivated) and the total number of all pixels within the primary sampling unit boundary. The percent cultivated is given by the number of "1" pixels divided by total number of pixels.
   6. Determine the primary sampling unit stratum by checking the stratum definition look-up table to map the computed percent cultivation to a defined stratum, and label the primary sampling unit with a corresponding stratum number as a primary sampling unit boundary attribute.
   7. Determine stratum definitions for all area frame primary sampling units in the state by repeating steps 2-6 for every primary sampling unit (Boryan et al., 2014).

2. Manual/Editing and Review Steps (Green boxes in Fig. 4)

   1. Identify and examine area frame primary sampling units to identify those that do not match the automated strata results and exceed size tolerances.
   2. Identify primary sampling unit boundary and population changes using ancillary data sources including the original traditional area frame, Farm Service Agency Common Land Unit data, Census data, Landsat 8 satellite imagery and National Agricultural Imagery Program data.

3. Identify areas of Cropland Data Layer omission and commission error using the previously mentioned ancillary data sources.
4. Edit primary sampling units based on steps 8-10. Manually edit the original area frame primary sampling units using ESRI's Arc GIS based on the results of steps 8-10.

3. Automated Stratification Steps continued (Blue boxes in Fig. 4)

   1. Compute percent cultivation of each area sampling frame primary sampling unit on the newly updated area frame as a final review step to identify primary sampling units in which the automated stratification results do not match with the current stratum definitions.
   2. Define strata based on the percent cultivation calculation. This step includes conducting a final review of these non-matching primary sampling units to determine the appropriate stratum definition for each.
   3. Revised Area Sampling Frame based on Hybrid Method is complete.

This integrated automated stratification and manual hybrid process is currently being used in NASS area frame operations.

5 Results And Discussion

This section presents and discusses the assessment results for three area frame stratification methods (traditional, automated and hybrid). The improvement metrics include: 1) stratification accuracy; 2) mean stratum percent cultivation range, mean stratum standard deviations; and mean stratum primary sampling unit size 3) the variances of key estimators and 4) area frame construction labor cost.

5.1 Stratification Accuracy

South Dakota, Oklahoma, Arizona, New Mexico, Georgia, Alabama, and North Carolina area frames were successfully revised from 2014–2015 using the new hybrid area frame construction method. As an example, the top of Fig. 5 illustrates the revised South Dakota area frame, which was stratified using the hybrid method. The three maps at the bottom of Fig. 5 illustrate a zoomed area of the original South Dakota area frame created using the traditional method (bottom left), the result of using automated stratification (bottom center) and the revised South Dakota area frame created using the hybrid method (bottom right). There are significant differences in the high cultivation areas between the two stratified frames from the traditional and automated stratification methods. As shown in Fig. 5 (bottom center), the stratification created using the automated method identifies more variation and detail in highly cultivated areas (dark and medium green colors), while the traditional method yields more variation in areas of the low cultivation (light green and tan colors) as shown in Fig. 5 (bottom left). The stratification generated with the automated method could be used as an update of the South Dakota area frame. However, the area frame can be further improved by reducing primary sampling unit sizes. Therefore, with this integrated new hybrid process, the manual procedures described above are applied to the frame derived with the automated stratification method. The final revised hybrid frame provides more detailed information in both high and low cultivation areas. Further, as stratum primary sampling unit homogeneity has been improved and primary sampling unit sizes reduced, the new hybrid method results in overall higher stratification accuracy.

The dot chart as shown in Fig. 6 compares the accuracies by state and method. For accuracy, we use the percent of segments (based on data collected in the June Area Survey) that matched their primary sampling unit strata definition. The years indicate the June Area Survey validation data used to obtain the accuracies. In Fig 6, the top dot chart identifies the state accuracies accuracies obtained using the traditional method. The middle chart shows the state accuracies obtained using the automated method and the bottom chart shows the state accuracies obtained using the hybrid method.

An average 15 percentage point accuracy improvement (Difference between Automated and Traditional Accuracy) is achieved based on the automated stratification results with June Area Survey reported data as *in situ* validation. Oklahoma and Arizona have larger initial improvements using the automated stratification, while New Mexico, Georgia and South Dakota have lower accuracy improvements directly related to the automated stratification alone. On average, an additional 15 percentage point increase (Difference between Hybrid and Automated Accuracy) in state level accuracies is achieved with the addition of manual editing to reduce primary sampling unit sizes and the manual review to identify areas which are impacted by Cropland Data Layers errors of omission or commission. An overall average 30 percentage point area frame accuracy improvement for the seven new area frames provides strong evidence for the improvements that can be directly attributed to the integration of the Cropland Data Layer – based automated method into the NASS operational process.

These new and updated area frames described in this paper are currently in use operationally in NASS for the June Area Survey. The revised Oklahoma area frame was first used operationally in 2014. The following state area frames were first used for the June Area Survey in the year in parenthesis including: Arizona (2014), New Mexico (2014), Georgia (2014), South Dakota (2015), Alabama (2015), North Carolina (2015). These revised area frames will continue to be used each year to select the June Area Survey sample until they are revised again. Currently, the NASS area frames are

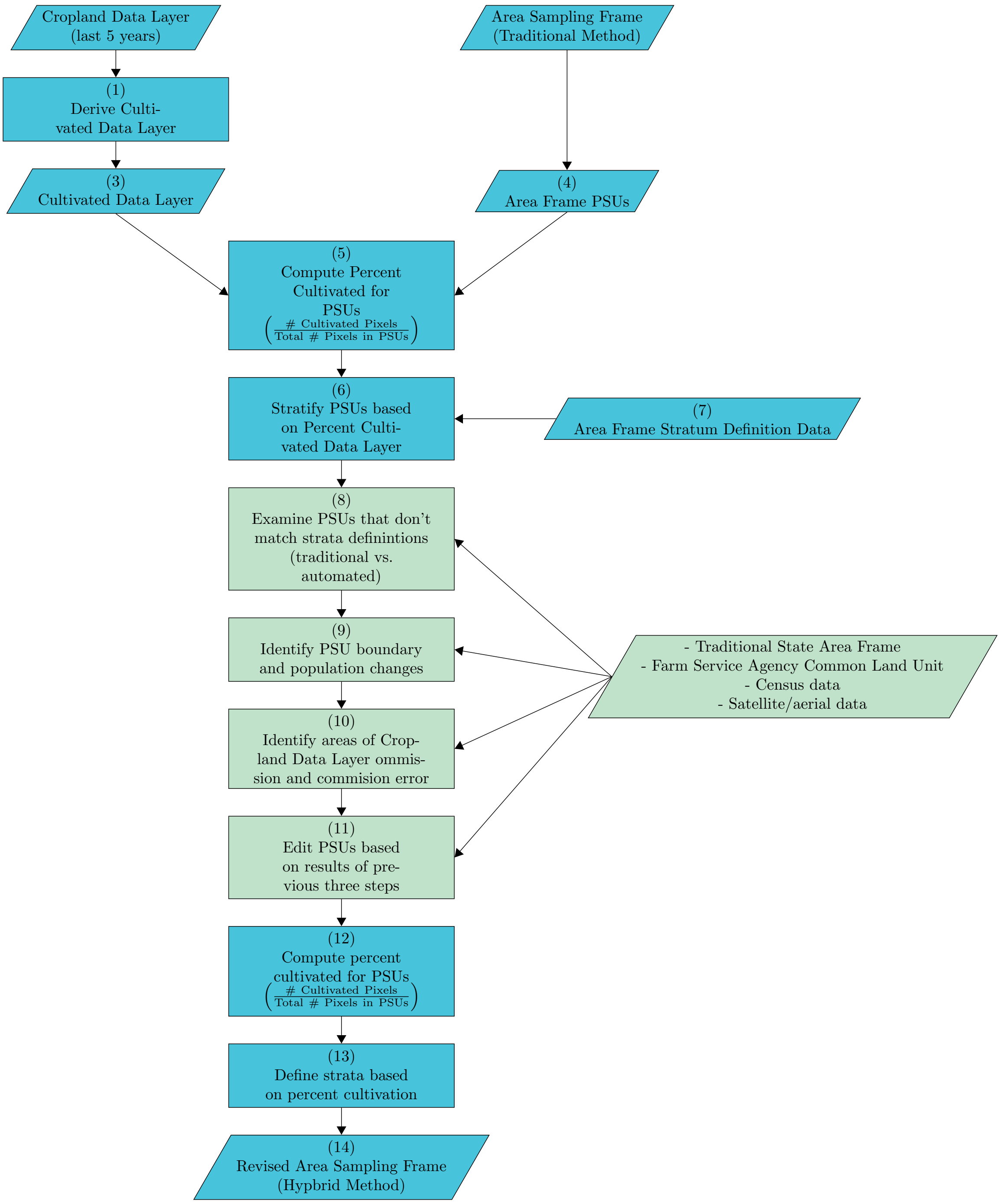

Figure 4. Flowchart illustrates the hybrid method: integrating Cropland Data Layer automated stratification of existing State area frames with a manual editing/review process. The automated stratification steps and data (#1-#7; #12-#14) are identified in the blue boxes. The manual editing/review steps and data (#8-#11) are identified in the green boxes. The numbers, in parenthesis, associated with each step correspond to steps 1-14 described in Section 4 (Methods).

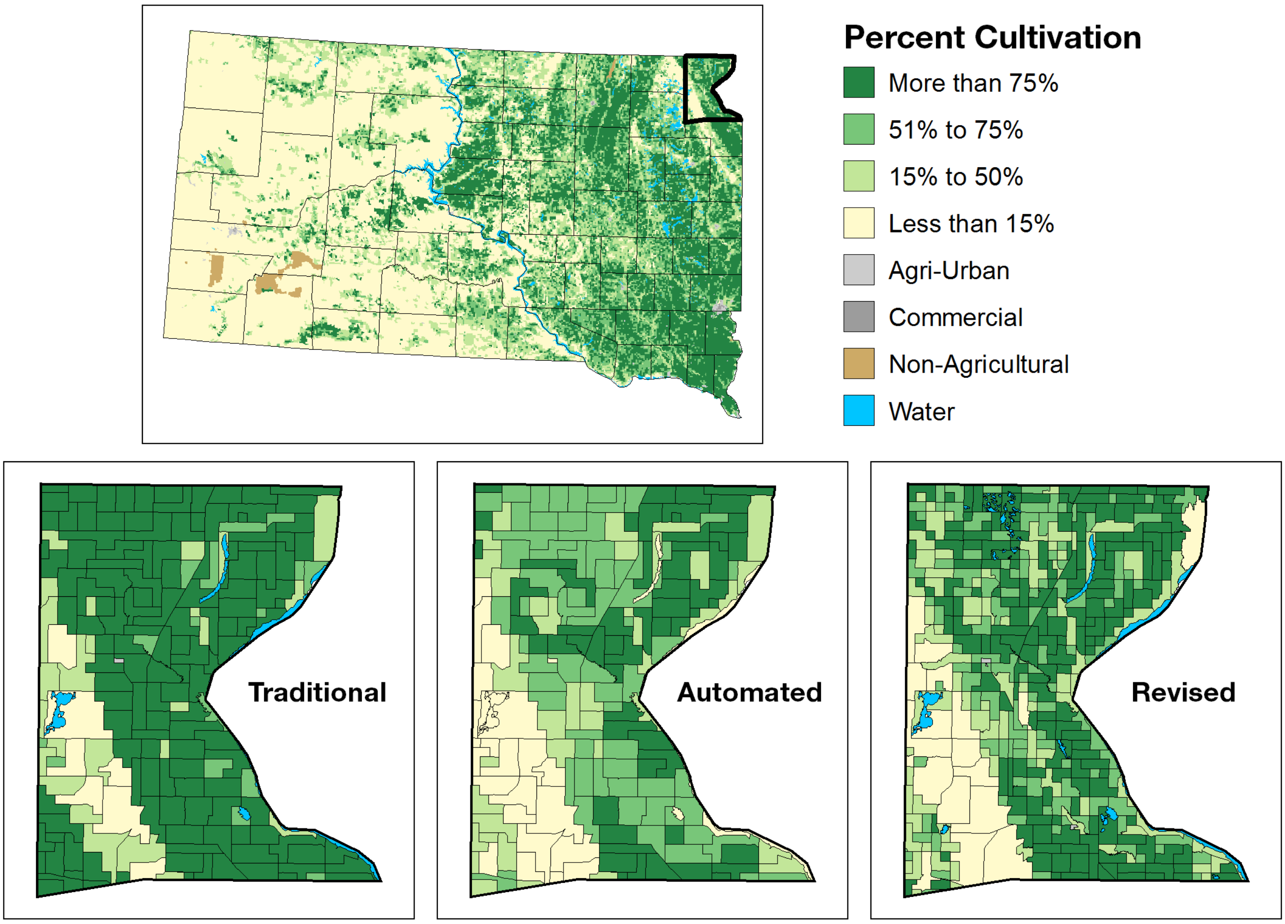

Figure 5. Revised South Dakota area frame (top) created using the new hybrid process. Zooms of Roberts County, South Dakota. The stratification on the bottom left was derived using the traditional method (visual analysis). The bottom center stratification was derived using the Cropland Data Layer; automated method applied to the original frame. The bottom right stratification (Revised) was derived using the new hybrid process which further improves area frame accuracy.

revised, on average, every 10-20 years.

5.2 Mean Stratum Percent Cultivation Range, Standard Deviations and Primary Sampling Unit Size

Area Sampling Frame improvements are further assessed based on mean stratum percent cultivation (1) stratum standard deviations (2) to assess consistency with stratum definitions and stratum homogeneity and mean stratum primary sampling unit size. These mean stratum percent cultivation and stratum standard deviation are defined by Equations 1 and 2 respectively:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \tag{1}$$

where x_i is the percent cultivation value calculated for each primary sampling unit, and n is the number of primary sampling units in a stratum.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

where x_i, n and $\bar{x}$ are defined as above.

The percent cultivation value for each primary sampling unit is based on the total number of cultivated pixels in a primary sampling unit divided by the total number of pixels in the primary sampling unit. The stratum mean percent cultivation is the average of these values for all primary sampling units within each stratum. The standard deviation calculations show how much the primary sampling unit percent cultivation calculations vary from the mean percent cultivation for the all primary sampling units in a stratum. The standard deviation is an indication of the variability in primary sampling unit percent cultivation within each stratum. A lower standard deviation indicates improved stratum homogeneity.

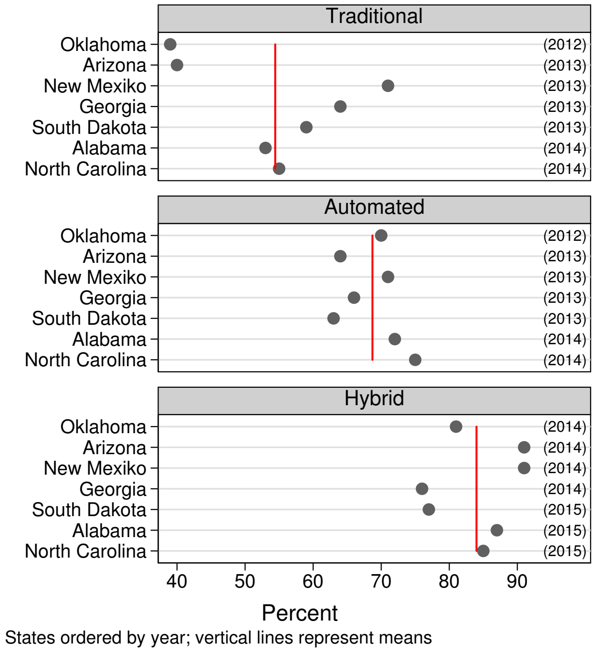

Figure 6. Traditional, automated and hybrid area frame state level accuracies. The year indicates the June Area Survey validation data used to obtain the accuracies.

Table 3 identifies the mean primary sampling unit sizes, mean percent cultivation and standard deviations for the three frames. The average primary sampling unit sizes are reduced in the hybrid area frames for all strata, but reductions are largest for the cultivated strata (11, 12, 13 and 20). This manual reduction in primary sampling unit sizes is intended to improve area frame accuracy and primary sampling unit homogeneity.

A significant result of integrating the new automated method into NASS operations, is ensuring that stratum mean percent cultivation is consistent with the stratum definition ranges. For example, strata 11, 12 and 20 in the traditional Oklahoma area frame, strata 11 and 12 in the traditional South Dakota area frame and strata 13 and 20 in the traditional Alabama area frame do not have mean strata percent cultivations that are consistent with their corresponding strata definition ranges. In Table 3, the stratum percent cultivation values, which are outside of the bound of the stratum definition ranges, are identified with an (a). After revision, all seven area frames have mean percent cultivations calculated at the primary sampling unit level and averaged at the stratum level which are consistent with the corresponding strata definition ranges. After the automated stratification results are applied, mean percent cultivation, for all strata, are not only more consistent but are more closely centered at the median of the strata definition ranges. This improvement is maintained after manual editing is conducted in the revised area frames.

Further, after the automated stratification process is applied, the majority of stratum level standard deviations, based on percent cultivation, are significantly lower when compared with the traditional area frames, with the exception of Alabama (stratum 20) which has the same standard deviation and Oklahoma (stratum 40) and Alabama (stratum 40) which have marginally higher standard deviations. The standard deviation improvements for 20 out of 23 strata in the seven area frames indicate that utilizing the new automated stratification method significantly improves stratum homogeneity.

Manual primary sampling unit editing and review retains the improved stratum homogeneity provided by the automated results, except in the low cultivation areas (stratum 40), where the standard deviations are slightly higher in the hybrid frames. While maintaining significantly improved levels of stratum homogeneity, the most significant impact of manual editing and review is an average of 15 percentage point improvement in area frame accuracy (percent of segments matching the primary sampling unit definition) that directly results from the manual procedures. The combined accuracy improvements based on the use of the automated stratification method (15 percentage points) and manual editing/ and review (15 percentage points) results in significantly improved area frame accuracies for NASS operational use.

5.3 The Variances of Key Estimators

A comparison of the variances of key estimators is another technique used to evaluate the traditional and automated stratification methods. Since the revised area frames are in current use, the crop estimate coefficients of variation (CVs) obtained from the June Area Survey can be compared based on the traditional area frame with the June Area Survey crop estimate CVs obtained using the revised hybrid area frames. Table 4 illustrates June Area Survey estimate CVs for the seven states included in this study, with major crops that are "critical" for setting the U.S. corn, soybeans, wheat and cotton estimates, as well as states that have June Area crops that are greater than 1,000,000 acres in planted area. Alabama, Arizona and New Mexico do not have any crops that fit this criteria. Since the actual June Area Survey estimates are considered confidential and not released to the public, the NASS 2016 official published estimates are included as a reference of crop acreage (USDA FSA, 2017d). Six out of nine June Area Survey crop estimate CVs are lower using the hybrid area frames, which demonstrates that the hybrid method results in June Area Survey estimates with improved precision. However, it should be indicated that the sample of nine CVs is not large enough to draw a statistical conclusion. Moreover, to derive traditional area frames for estimate and variance assessment simulation is impractical due to pro-

Table 3
Comparison of traditional, automated and hybrid area frame stratum level mean size, mean percent cultivation and standard deviations [a,b]

		Traditional area sampling frames			Automated area sampling frames			Hybrid area sampling frames		
Stratum	State	Mean PSU in square miles	Mean % cultivation	Standard deviation	Mean PSU in square miles	Mean % cultivation	Standard deviation	Mean PSU in square miles	Mean % cultivation	Standard deviation
11	Oklahoma	5.22	50.5[a]	30.6	5.83	83.2	5.4	3.91	84.7	6.3
	South Dakota	5.91	67.5[a]	17.9	6.08	83.5	5.5	2.64	84.1	7.0
12	Oklahoma	4.07	21.3[a]	21.7	4.10	63.0	7.2	3.09	62.1	8.2
	South Dakota	4.73	44.6[a]	18.8	5.76	63.3	7.2	2.22	59.7	8.0
13	Arizona	6.23	71.3	30.3	5.59	81.2	14.9	3.00	90.0	14.0
	New Mexico	6.08	53.7	30.8	6.21	75.1	14.8	1.58	87.3	24.7
	Georgia	6.41	54.1	12.4	6.62	60.4	7.7	2.88	61.7	9.7
	Alabama	6.88	32.5[a]	18.2	7.09	58.7	7.1	2.72	65.4	8.8
	North Carolina	5.79	60.1	16.8	5.93	63.1	12.6	2.56	66.2	12.9
20	Oklahoma	5.36	8.0[a]	11.4	6.01	30.3	10.3	3.85	10.4	
	South Dakota	9.25	20.1	15.6	6.41	33.0	10.2	2.48	32.0	9.7
	Arizona	4.68	36.6	28.2	6.40	29.3	10.1	1.97	57.8	28.1
	New Mexico	4.87	21.5	19.9	6.14	31.2	10.4	3.26	39.2	19.8
	Georgia	6.04	25.7	11.9	6.61	30.2	9.4	2.98	32.5	9.1
	Alabama	6.55	10.6[a]	9.3	6.88	26.8	9.3	2.72	32.4	9.0
	North Carolina	7.31	25.7	14.8	7.35	29.4	9.7	2.88	32.4	8.0
40	Oklahoma	7.81	3.6	3.5	7.75	2.5	3.7	7.35	4.8	3.9
	South Dakota	19.17	2.7	5.6	13.75	4.1	4.4	19.94	2.4	3.3
	Arizona	25.22	0.9	3.7	26.61	0.8	2.4	24.74	2.4	8.6
	New Mexico	23.31	0.5	2.3	20.56	0.1	1.0	24.09	0.9	3.7
	Georgia	14.00	3.4	5.1	15.00	2.6	3.6	12.44	3.7	4.9
	Alabama	15.24	1.2	2.1	10.35	2.5	3.6	12.89	2.7	4.6
	North Carolina	11.62	3.8	5.0	7.05	4.0	4.3	10.83	5.8	5.6

[a] Indicates the stratum percent cultivation values, which are outside of the bound of the stratum definition ranges
[b] Mean percent cultivation and standard deviations based on percent cultivation derived using the NASS 2014 Cultivated Layer.

Table 4
Estimated coefficient of variation (CV) comparison – traditional vs. hybrid area frame

State	Crop	NASS Official Published Estimate 2016 (Acres)	June Area Survey CV Traditional Frame	June Area Survey CV Hybrid Frame
South Dakota				
	Corn - Planted	5,600,000	4.9 (a)	4.9 (c)
	Soybeans	5,200,000	5.0 (a)	5.4 (c)
	Spring Wheat - Planted	1,080,000	17.6 (a)	**14.6** (c)
	Winter Wheat - Planted	1,180,000	17.9 (a)	**13.5** (c)
Oklahoma				
	Winter Wheat - Planted	5,000,000	5.6 (a)	**4.7** (c)
North Carolina				
	Corn - Planted	1,020,000	13.9 (b)	**8.1** (c)
	Soybeans - Planted	1,700,000	6.7 (b)	**6.2** (c)
	Upland Cotton	280,000	14.3 (b)	**12.9** (c)
Georgia				
	Upland Cotton	1,190,000	7.4 (a)	7.5 (b)

June Area Survey data used to obtain estimated coefficients of variation – (a) - 2013, (b) – 2014, (c) – 2015

hibitive cost.

5.4 Oklahoma and South Dakota Area Frames: State and Stratum Accuracy Improvements

Table 5 illustrates the state and stratum level accuracies of the traditional and the hybrid Oklahoma area frames based on 2012 June Area Survey reported data. The traditional area frame's cultivated strata are performing poorly, with accuracies ranging from 10-35%, and the non-cultivated strata are well identified with high accuracies. As shown in Table 5, large improvements in stratum level accuracy, based on the 2014 June Area Survey segment data are reported for all cultivated strata. These gains in accuracy are attributed to the integration of the new automated stratification method combined with manual editing/review. Results in Table 6 for South Dakota are similar.

The hybrid Oklahoma area frame more accurately reflects current conditions, based on percent cultivated land, with a 42 percentage point improvement in state level area frame accuracy (Fig 7 and Table 5). Fig. 7 (top left) illustrates the traditional Oklahoma state area frame and Fig. 7 (top right) illustrates the hybrid Oklahoma area frame updated using the new integrated process. The dark and mid green shades identify primary sampling units that are defined as stratum 11, 12 and 20. As shown in Fig. 7 (top right), the hybrid Oklahoma area frame has significantly less land cover identified as highly cultivated (>75% cultivation) than the traditional Oklahoma frame illustrated in Fig. 7 (top left). In the traditional Oklahoma area frame there are many land areas in eastern Oklahoma that are over identified as stratum 11 and 12.

The hybrid South Dakota area frame is much more accurate (Table 6), particularly in the highly cultivated strata, than the traditional South Dakota area frame with accuracy improvements in stratum 11 (28 percentage point) and stratum 12 (39 percentage point) based on 2013 (traditional frame accuracy) and 2015 (hybrid frame accuracy) June Area Survey reported data. Figure 8 illustrates the traditional South Dakota state area frame (top left) and the hybrid South Dakota area frame (top right). A zoom of the South Dakota traditional stratification (bottom left) and a zoom of the same area in the hybrid South Dakota area frame (bottom right) illustrate the reduction in primary sampling unit sizes and larger amount of detail after revision.

5.5 Labor Cost

Overall, with the integration of the new automated stratification method into NASS area frame operations, new state area frames are being built with improved objectivity, efficiency and accuracy at reduced cost. For example, the traditional Oklahoma area frame was constructed in 4,552 employee hours, while the hybrid frame, required 1,980 employee hours. In general, the reduction in labor cost for other states will be similar. States with more intensive agricultural production may have more cost savings, due to less manual editing, while states with less agriculture may cost relatively more due to a larger manual editing effort.

Table 5
Traditional Oklahoma area frame and hybrid Oklahoma area frame stratum level accuracies based on June area survey reported segment data

Stratum	Percent Cultivated	Trad. June Area Survey 2012 Segments	Trad. June Area Survey 2012 Accuracy (%)	Hybrid June Area Survey 2014 Segments	Hybrid June Area Survey 2014 Accuracy (%)	Overall Accuracy improv. (%)
11	> 75%	150	35	112	84	49
12	51% – 75%	60	10	60	58	48
20	15% – 50%	70	24	60	73	49
31	Agri-Urban	2	100	2	100	0
31	Commercial	2	100	2	100	0
40	< 15%	55	96	99	95	−1
50	Non Agricultural	2	100	2	100	0
State		341	39	337	81	42

Table 6
Traditional South Dakota area frame and hybrid South Dakot area frame stratum level accuracies based on June area survey reported segment data

Stratum	Percent Cultivated	June Area Survey 2013 Segments	June Area Survey 2013 Accuracy (%)	June Area Survey 2015 Segments	June Area Survey 2015 Accuracy (%)	Overall Accuracy improvement (%)
11	> 75%	210	60	112	88	28
12	51% – 75%	60	32	60	71	39
20	15% – 50%	70	53	60	57	4
31	Agri-Urban	2	100	2	100	0
31	Commercial	2	100	2	100	0
40	< 15%	50	92	30	97	5
50	Non Agricultural	2	50	2	100	50
State		396	59	337	77	18

6 Conclusion

This paper presents a new area frame construction operational process for agricultural surveys, which has been implemented in USDA National Agricultural Statistics Service operations. The new area frame construction operational process incorporates a new Cropland Data Layer-based automated stratification method, recently developed by Boryan and Yang, into the area frame construction operational process as the primary stratification method for building new state area frames as illustrated in flow chart Fig. 4. The implementation integrates automated stratification (Boryan et al., 2014) with the traditional method (Cotter et al., 2010). This new operational process utilizes results of both traditional and automatic stratification methods and further refines frames with manual review and editing procedures.

In this paper, new 2014–2015 area frames for Oklahoma, South Dakota, Arizona, New Mexico, Georgia, Alabama, and North Carolina were constructed using the new operational process. It was found that the new hybrid process achieved 12 to 51 percentage point accuracy improvements, based on June Area Survey reported data, in the seven test states. An average of 15 percentage point accuracy improvement can be directly attributed to the addition of the automated stratification method and an additional average of 15 percentage point accuracy improvement can be attributed to the manual editing and review procedures as shown in Fig. 6. The mean percent cultivated land of all strata for all states, based on the NASS 2014 Cultivated Layer, became consistent with the strata definitions after revision. In addition, stratum homogeneity was significantly improved using the automated stratification method with reduced standard deviations recorded for 20 of 23 strata across the seven states. Six out of nine crop estimate coefficients of variation were lower (bold in Table 4) using the new hybrid area frames, which provides further support that the new stratification methodology results in June Area Survey crop estimates with improved precision. Further, primary sampling unit size reductions, for all cultivated strata (11, 12, 13 and 20) in all seven states using manual review and editing techniques, resulted in additional accuracy gains (15 percentage point) and overall significantly improved (30 percentage point) NASS area sampling frames

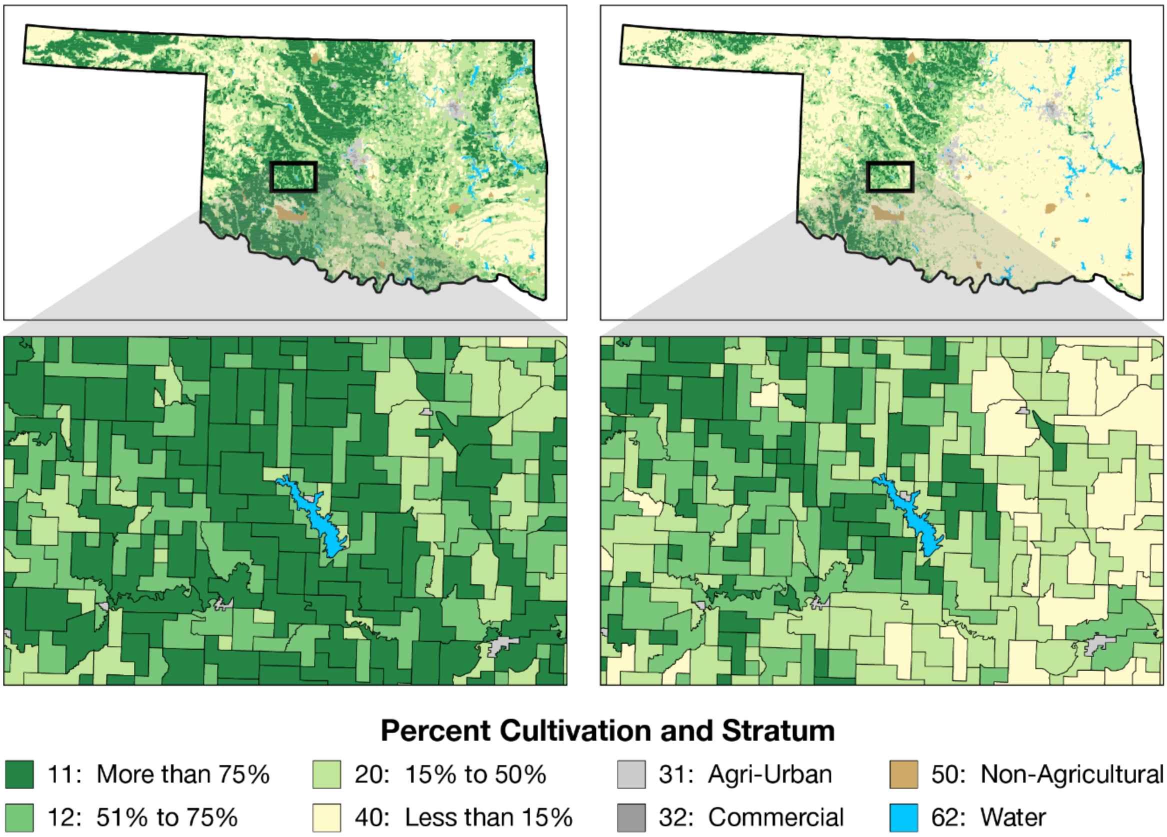

Figure 7. Oklahoma traditional area frame (top left) and hybrid area frame (top right). Zooms of the same location in both the traditional area frame (bottom left) and the hybrid area frame (bottom right) illustrate the greater amount of detail in the hybrid area frame

constructed at reduced cost.

The improvement of the NASS area frames is an ongoing process. Overall, the new process that integrates the new automated stratification method, based on the Cropland Data Layer, with traditional stratification is robust and has resulted in significant improvements in area frame accuracy, efficiency, objectivity based on June Area Survey reported data (Table 3) and a reduction in the cost of operational area frame construction. The hybrid approach, presented in this paper, is being extended by NASS to all states and provides the opportunity to revise area frames more frequently, if needed. This new NASS operational area frame construction method, based on available geospatial data, is easily transferrable to other agencies who conduct area frame based surveys. The operational process is straightforward and is anticipated to result in improved crop estimates and significantly reduced labor cost for area frame construction. Revising the area frames for states with minimal agriculture poses a challenge because the Cropland Data Layers are generally less accurate in these states. Consequently, the manual editing and review process is relied upon more heavily in these states. Finally, research continues into the assessment of a multi-crop specific area frame, (Boryan, Yang, Willis, & Di, 2017), evaluating the impact of Cropland Data Layer pixel level buffering on stratification and the use of higher resolution Cropland Data Layers for automated area frame stratification.

Acknowledgment

The authors wish to thank Ms. Lee Ebinger of the National Agricultural Statistics Service for creating the figures for this paper and Mr. Terry Broz and Ms. Avery Sandborn of the National Agricultural Statistics Service for reviewing and providing suggestions that significantly improved this paper.

References

Arroway, P., Abreu, D. A., Lamas, A. C., Lopiano, K. K., & Young, L. Y. (2010). An alternate approach to assessing misclassification in JAS. In *Proceedings of the joint statistical meetings*.

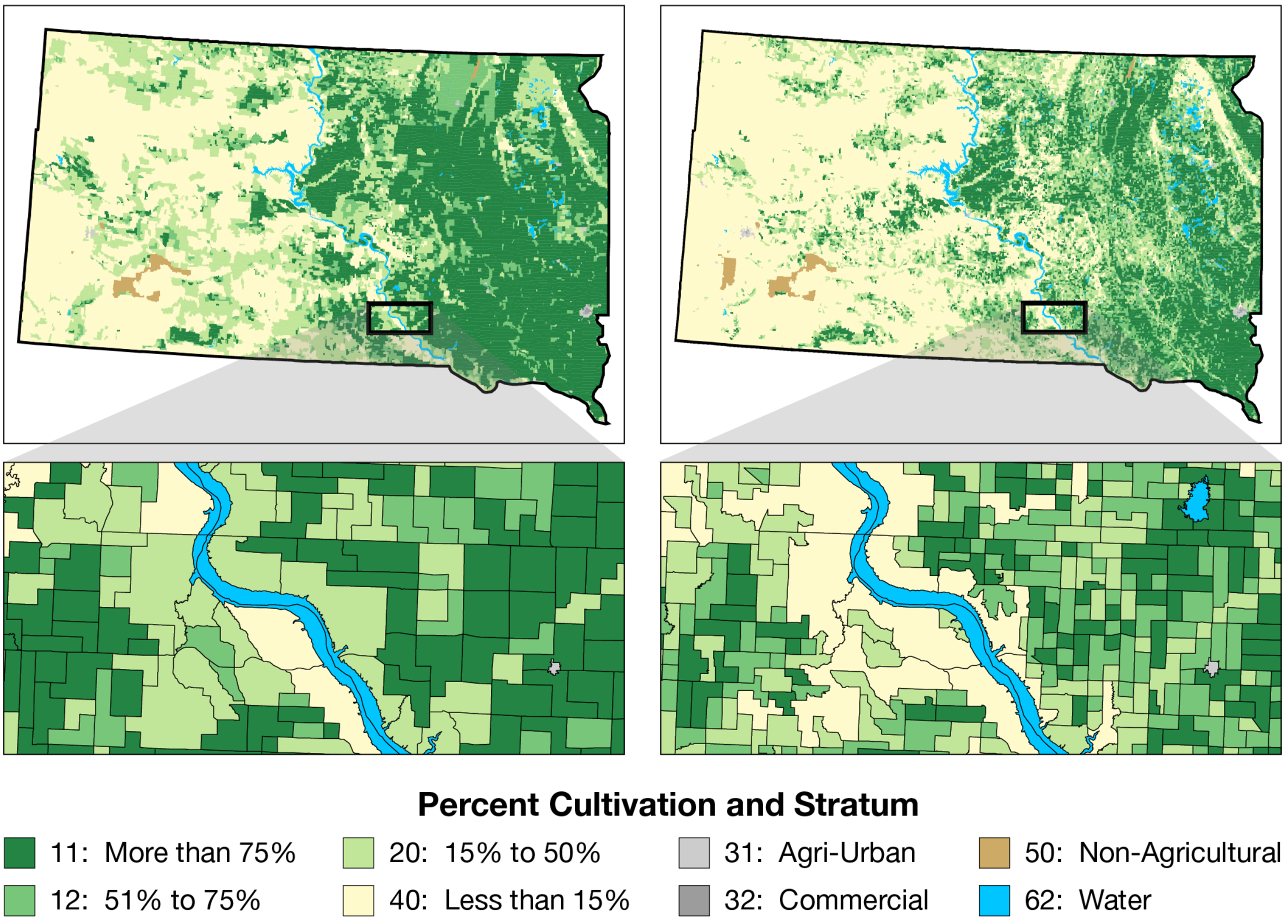

Figure 8. South Dakota traditional area frame (top left) and hybrid area frame (top right). Zooms of the same location in both the traditional area frame (bottom left) and the hybrid area frame (bottom right) illustrate the smaller primary sampling unit sizes and larger amount of detail in the hybrid area frame

Benedetti, R., Piersimoni, F., & Postiglione, P. (2015). *Sampling spatial units for agricultural surveys*. Berlin, Germany: Springer.

Bing Maps. (2017). Retrieved from https://www.bing.com/maps

Boryan, C. G., Yang, Z., & Di, L. (2012). Deriving 2011 cultivated land cover data sets using USDA national agricultural statistics service historic cropland data layers. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS): Munich, Germany, July 22-27, 2012* (pp. 6297–6300). IEEE.

Boryan, C. G., Yang, Z., Di, L., & Hunt, K. (2014). A new automatic stratification method for US agricultural area sampling frame construction based on the cropland data layer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(11), 4317–4327. doi:10.1109/JSTARS.2014.2322584

Boryan, C. G., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, *26*(5), 341–358.

Boryan, C. G., Yang, Z., Willis, P., & Di, L. (2017). Developing crop specific area frame stratifications based on geospatial crop frequency and cultivation data layers. *Journal of Integrative Agriculture*, *16*(2), 312–323. doi:10.1016/S2095-3119(16)61396-5-5

Carfagna, E. & Gallego, F. J. (2005). Using remote sensing for agricultural statistics. *International statistical review*, *73*(3), 389–404.

Cotter, J. J., Davies, C., Nealon, J., & R., R. (2010). Area frame design for agricultural surveys. In R. Benedetti, F. Piersimoni, M. Bee, & G. Espa (Eds.), *Agricultural survey methods* (pp. 169–192). Chichester, United Kingdom: John Wiley & Sons. doi:10.1002/9780470665480.ch11

Cotter, J. J. & Tomczak, C. (1994). An image analysis system to develop area sampling frames for agricultural surveys. *Photogrammetric engineering and remote sensing*, *60*(3), 299–306.

Davies, C., Hunt, K., & Willen, E. (2014). *Area frame section: 2014 state area frame evaluation and ranking for future frame updates*. USDA National Agricultural Statistics Service Internal Report.

Dunham, P., Weyermann, D., & Azuma, D. (2002). A comparison of stratification effectiveness between the national land cover data set and photointerpretation in western Oregon. In R. E. McRoberts, G. A. Reams, P. C. Van Deusen, & J. W. Moser (Eds.), *Proceedings of the 3d annual forest inventory and analysis symposium: Traverse City, Michigan, October 17-19, 2001 (General Technical Report NC-230)* (pp. 63–68). St. Paul, MN, U.S.: North Central Research Station, Forest Service, U.S. Department of Agriculture.

ESA. (2017a). Copernicus open access hub. Retrieved from https://scihub.copernicus.eu/

ESA. (2017b). Sentinel 1 toolbox. Retrieved from https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1

FAO. (1998). Multiple frame agricultural surveys. (Chap. Agricultural survey programmes based on area frame or dual frame (area and list) sample designs, 10, p. 274). AO Statistical Development Series. Rome, Italy: FAO.

Faulkenberry, G. D. & Garoui, A. (1991). Estimating a population total using an area frame. *Journal of the American Statistical Association*, *86*(414), 445–449.

Ford, B. L., Nealon, J., Tortora, R. D., et al. (1986). Area frame estimators in agricultural surveys: sampling versus nonsampling errors. *Agricultural Economics Research*, *38*(2), 1–9.

Gallego, F. J., Carfagna, E., & Peedell, S. (1999). The use of CORINE Land Cover to improve area frame survey estimates. *Research in Official Statistics*, *2*(2), 99–122.

Gallego, F. J. & Delincé, J. (2010). The European land use and cover area-frame statistical survey. In R. Benedetti, F. Piersimoni, M. Bee, & G. Espa (Eds.), *Agricultural survey methods* (pp. 151–168). Chichester, United Kingdom: John Wiley & Sons. doi:10.1002/9780470665480.ch11

Gallego, F. J., Delincé, J., & Carfagna, E. (1994). Two stage area frame sampling on square segments for farm surveys. *Survey Methodology*, *20*(2), 107–115.

Google Earth Street View Data. (2017). Retrieved from https://www.google.com/maps/streetview/

Han, W., Yang, Z., Di, L., & Mueller, R. (2012). CropScape: a Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture*, *84*, 111–123.

Hansen, M. H. & Wendt, D. G. (2000). Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In R. E. McRoberts, G. A. Reams, P. C. Van Deusen, & J. W. Moser (Eds.), *Proceedings of the first annual forest inventory and analysis symposium (General Technical Report NC-213)* (pp. 20–27). St. Paul, MN, U.S.: North Central Research Station, Forest Service, U.S. Department of Agriculture.

Hanuschak, G. & Morrissey, K. M. (1977). *Pilot study of the potential contributions of landsat data in the construction of area sampling frames*. Washington, D.C., U.S.: Statistical Reporting Service, US Department of Agriculture.

Heald, J. (2002). ESRI ArcUser Online. april-june 2002. USDA establishes a common land unit. Retrieved from http://www.esri.com/news/arcuser/0402/usda.html

Liknes, G. C., Nelson, M. D., Gormanson, D. D., & M. Hansen, M. (2009). The utility of the cropland data layer for forest inventory and analysis. In R. E. McRoberts, G. A. Reams, P. C. Van Deusen, & J. W. Moser (Eds.), *Proceedings of the 8th annual forest inventory and analysis symposium: Monterey, CA, October 16-19, 2006 (General Technical Report WO-79)* (pp. 259–264). Washington, D.C., U.S.: Forest Service, U.S. Department of Agriculture.

Liknes, G. C., Nelson, M. D., & McRoberts, R. E. (2004). Evaluating classified MODIS satellite imagery as a stratification tool. In *Proceedings of he joint meeting of the 15th annual conference of the international environmetrics society and the 6th international symposium on spatial accuracy assessment in natural resources and environmental sciences: Portland, Maine, 28 June – 1 July, 2004.*

Martinez, L. I. (2013). Improving the use of GPS, GIS and Remote Sensing for setting up a master sampling frame. Global Strategy: Improving Agricultural & Rural Statistics, Technical Report Series GO-06-2015. July.

McRoberts, R. E., Wendt, C. G., Nelson, M. D., & Hansen, M. H. (2001). Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment*, *81*(1), 36–44.

McRoberts, R. E. & Wendt, D. G. (2002). Implementing a land cover stratification on-the-fly. In R. E. McRoberts, G. A. Reams, P. C. Van Deusen, & J. W. Moser (Eds.), *Proceedings of the 3d annual forest inventory and analysis symposium: Traverse City, Michigan, October 17-19, 2001 (General Technical Report NC-230)* (pp. 137–145). St. Paul, MN, U.S.:

North Central Research Station, Forest Service, U.S. Department of Agriculture.

Nusser, S. M. & House, C. C. (2009). Sampling, data collection, and estimation in agricultural surveys. In D. Pfeffermann & C. R. Rao (Eds.), *Handbook of statistics 29a: sample surveys: design, methods and applications* (Vol. 29, pp. 471–486). Amsterdam, The Netherlands: Elsevier.

Perry, C. R. (2000). Improving the efficiency of the arkansas area frame using categorized satellite imagery. NASS Technical Report, U.S. Department of Agriculture.

Perry, C. R. & Gentle, J. (2000). Optimal stratification of area frames. In *Proceedings of the second international conference on establishment surveys survey methods for businesses, farms, and institutions: Buffalo, New York. June 17-21, 2000*.

Pradhan, S. (2001). Crop area estimation using GIS, remote sensing and area frame sampling. *International Journal of Applied Earth Observation and Geoinformation*, *3*(1), 86–92.

QGIS. (2017). A free and open source geographic information system. Retrieved from http://qgis.org/en/site/

Tsiligirides, T. A. (1998). Remote sensing as a tool for agricultural statistics: a case study of area frame sampling methodology in Hellas. *Computers and electronics in agriculture*, *20*(1), 45–77.

U.S. Census Bureau. (2017). 2010 census - census block maps. Retrieved from https://www.census.gov/geo/maps-data/maps/block/2010/

USDA FSA. (2017a). FSA CLU information worksheet. Retrieved from http://www.fsa.usda.gov/Internet/FSA_File/clu__infosheet_2013.pdf

USDA FSA. (2017b). June area survey. Retrieved from https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/June_Area/

USDA FSA. (2017c). MCropScape and Cropland Data Layer. Retrieved from http://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php

USDA FSA. (2017d). Quickstats. Retrieved from https://quickstats.nass.usda.gov/

USDA FSA. (2017e). USDA, National Agricultural Statistics Service, 2016 Cultivated Layer. Retrieved from https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/2016_cultivated_layer_metadata.htm

USGS. (2017a). Landsat 8 (L8) data users handbook. Retrieved from http://landsat.usgs.gov/l8handbook.php

USGS. (2017b). Landsat data missions: landsat data access. Retrieved from https://landsat.usgs.gov/landsat-data-access

USGS. (2017c). The national map. Retrieved from http://nationalmap.gov/ustopo/about.html

Vogel, F. A. (1995). The evolution and development of agricultural statistics at the united states department of agriculture. *Journal of Official Statistics*, *11*(2), 161.

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, N. (2001). Completion of the 1990s national land cover data set for the conterminous United States from Landsat thematic mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, *67*(6), 650–662.

Workneh, F., Tylka, G. L., Yang, X. B., Faghihi, J., & Ferris, J. M. (1999). Regional assessment of soybean brown stem rot, phytophthora sojae, and heterodera glycines using area-frame sampling: prevalence and effects of tillage. *Phytopathology*, *89*(3), 204–211.