POST STRATIFICATION ASSESSMENT OF THE NASS AUTOMATED STRATIFICATION METHOD BASED ON THE CROPLAND DATA LAYER

Claire G. Boryan, Zhengwei Yang and Robert Seffrin
USDA National Agricultural Statistics Service
3251 Old Lee Highway, Room 301, Fairfax, VA 22030, U.S.A.
Email: Claire.Boryan@nass.usda.gov

ABSTRACT

Area Sampling Frames (ASFs) are the foundation of the agricultural statistics program of USDA National Agricultural Statistics Service (NASS). A geospatial Cropland Data Layer (CDL) based automated stratification (AS) method was recently implemented to achieve higher accuracies than traditional stratification (TS), based on visual interpretation, in cultivated areas. This paper extends the AS assessment to the post stratification estimates. South Dakota (SD) US 2013 post stratification estimates, based on AS, are compared with the SD 2013 June Agricultural Survey estimates based on TS. Post stratification estimates obtained using AS are comparable, to the TS estimates, based on estimate percent differences. Considering the significant improvement in accuracy using AS in cultivated strata in five test states, improved accuracy in the highly cultivated stratum and improved stratum homogeneity in this study, it is concluded that the CDL based AS method generates ASFs that are more objective, efficient, accurate, and homogeneous and reduces labor costs.

***Index Terms*— Area sampling frame, automated stratification, Cropland Data Layer, post stratification estimation assessment**

1. INTRODUCTION

Area Sampling Frames (ASFs) are the foundation of the agricultural statistics program of the USDA National Agricultural Statistics Service (NASS) and many other statistical survey programs around the world [1], [2], [3]. Research was recently conducted to develop and assess an automated method to stratify the NASS ASFs by calculating percent cultivation at the Primary Sampling Unit (PSU) level based on geospatial NASS Cropland Data Layers (CDL) [4]. The NASS CDLs are 30-56.0 meter raster-formatted, geo-referenced, cropland cover classifications derived from satellite data [5]. Experimental results indicate that the CDL automated stratification (AS) method achieved higher accuracies in intensively cropped areas while the traditional method (visual interpretation) achieved higher accuracies in low agricultural areas. Accuracy differences were statistically significant at a 95% confidence level [4].

The purpose of this paper is to extend the assessment of the new CDL AS method to the post stratification crop estimates. The 2013 South Dakota (SD), U.S. post stratification crop estimates, based on the CDL AS method are derived and the results are compared with the NASS SD 2013 June Agricultural Survey (JAS) estimates, which are based on the traditional stratification (TS) method. South Dakota 2013 Farm Service Agency (FSA) Common Land Unit (CLU) data are utilized as a novel source of proxy survey data. Preliminary results provide further evidence, beyond ASF accuracy [4], of the utility of the CDL AS method for NASS ASF stratification. In addition, an accuracy comparison between TS and AS is conducted based on the percentage of 2013 JAS segments which report a percent cultivation that matches the stratum definition of the PSU from which they are selected. Traditional stratification and AS stratum percent cultivation standard deviations (STD) are also calculated from the 2012 SD cultivated layer to provide a measurement of stratum homogeneity

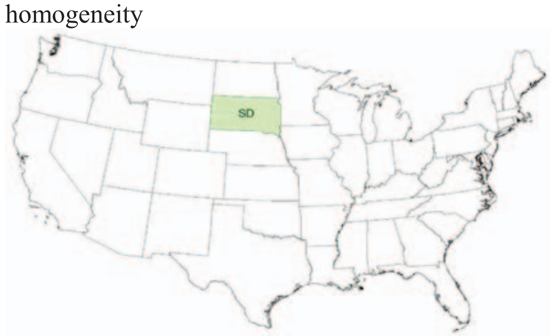

Fig. 1. South Dakota U.S. - Study Area for Post Stratification Estimation Assessment

2. DATA AND STUDY SCOPE

In this study, South Dakota, U.S. was selected as the study area due to the range of crops grown and the availability of USDA FSA CLU data, which covers approximately 98% of the state's cropland (Fig. 1). The geospatial datasets used in this study include: the NASS 2013 SD Area Sampling Frame [1], the NASS 2013 Cropland Data Layer [5], the NASS SD 2012 Cultivated Layer [6], NASS SD 2013 JAS segment data [1] and USDA SD 2013 FSA CLU data [7].

The NASS ASFs are made up of stratified parcels of land, known as PSUs, which are digitized to physical boundaries (roads, railroads, and rivers) on the ground. The NASS ASF stratification is based on percent cultivation of the land cover within PSUs. Table 1 illustrates NASS South Dakota land-use stratification codes and definitions. Once stratum definitions are assigned, all land is subdivided into PSUs which are designed to reduce labor cost in random sampling by eliminating the need to delineate the entire segment population. Selected PSUs are further subdivided into segments or sample units, and a segment is randomly selected from each selected PSU for enumeration [1].

TABLE 1. Land-Use Stratification Codes and Definitions Represented in the NASS South Dakota Area Sampling Frame

Land-Use Strata Codes	Codes Strata Definitions
11	General Cropland, greater than 75% cultivated
12	General Cropland, 51-75% 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

3. METHODOLOGY

Post stratification estimates are derived based on an AS 2013 SD ASF. The 2012 SD cultivated layer [6] is used as the input data to derive percent cultivation at the ASF PSU level. The method to conduct the CDL AS is described in Reference [4]. Post stratification estimates are compared to NASS 2013 SD JAS estimates, derived from the NASS TS and to the NASS SD 2013 "official" published estimates. Two sets of post stratification estimates are derived based on the SD 2013 ASF automated stratification. Figure. 2 illustrates the 396 selected PSUs (red polygons) with intersected FSA CLU data (black polygons). The selected FSA CLUs polygons used to create the segments are highlighted in yellow. The same number of segments (396) are used in the TS vs. AS comparison. The steps to conduct post stratification estimation are detailed as follows:

1) Run sample allocation program based on automated stratification, to determine the number of stratum segments sampled

2) Randomly select ASF PSUs based on allocation results

3) Overlay the SD 2013 ASF PSU boundary on the SD 2013 FSA CLUs

4) Identify selected PSUs, and for each PSU individually, manually select intersecting FSA CLU polygons to create FSA CLU segments based on stratum specific target segment sizes.

5) Link FSA CLU segments to FSA 578 administrative data

6) Generate first set of crop estimates, based on automated stratification, using data from FSA CLU/578 segments

7) Populate and summarize FSA CLU polygon segments with SD 2013 CDL data

8) Generate second set of estimates, based on automated stratification, using data from FSA CLU/CDL data

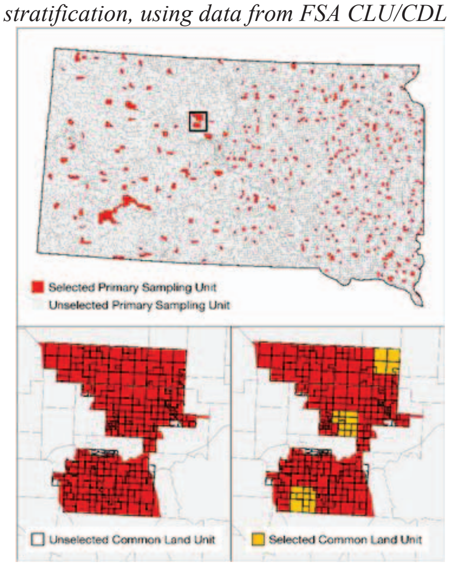

Fig 2. SD 2013 Post Stratification Estimation – Selected PSU and FSA CLU segments

4. RESULTS AND DISCUSSION

A comparison between TS and AS is performed based on the percent of 2013 JAS segments that report a percent cultivation that matches the stratum definition of the PSU from which they are selected. This "accuracy" measure is based on the assumption that all PSU segments belong to the same stratum, which is the basis for calculating stratum populations Therefore, stratum homogeneity is critical for ASF performance. Table 2 illustrates stratum "accuracy" results. Similar to the AS accuracy results in five states [4] the AS is more accurate in stratum 11 with a p-value close to zero indicating that the differences in accuracy between the TS and AS are statistically significant. However, the remaining strata accuracy differences, are not statistically significant. TS and AS stratum percent cultivation STDs are calculated from the 2012 SD cultivated data layer. The STDs, are lower for all strata in the AS, which indicates that strata defined using the AS are more homogeneous than the strata defined with TS, which is an important result.

Table 2. SD 2013 ASF – Comparison of Traditional vs Automated Stratification Accuracies, Mean Cultivation Standard Deviations and p-values.

2013 SOUTH DAKOTA ASF STRATIFICATION METHOD COMPARISON							
	Traditional Stratification			**CDL Stratification**			***p*-Value**
Stratum #- Percent Cultivation	***Accuracy (p_1)***	***Mean % Cultivation***	***Standard Deviations STD****	***Accuracy (p_2)***	***Mean % Cultivation***	***Standard Deviations-STD****	***Ha: $p_1 \neq p_2$***
11 - >75%	60.00%	67.5	17.9	82.05%	83.5	5.5	0.00045
12 -51% -75%	31.67%	44.6	18.8	40.00%	63.3	7.2	0.26974
20 -15% - 50%	52.86%	20.1	15.6	46.08%	33.0	10.2	0.38226
40 -< 15%	91.07%	2.7	5.6	82.56%	4.1	4.4	0.15409
Total	58.84%			59.09%			0.94241

4.1. POST STRATIFICATION ESTIMATION ASSESSMENT

The ultimate test of the new AS method is a comparison of TS vs. AS post stratification estimates. The TS estimates are the JAS estimates, which are referred to as the TS JAS estimates in this paper. For comparison, two sets of post stratification estimates are derived using the AS: 1) estimates from FSA CLU segments populated with FSA 578 "crop" data which are AS FSA CLU & 578; and 2) estimates from FSA CLU segments populated with NASS 2013 SD CDL data which are AS CLU & CDL. Estimates for SD 2013 planted *All Wheat, Winter Wheat, Spring Wheat, Corn, and Soybeans* at the state and strata levels are summarized. However, the TS JAS estimates are considered confidential and not released by NASS. Consequently, Table 3 includes the NASS official state published estimates (1) the estimate percent differences between the TS JAS and the AS CLU & 578 estimates (2) and between the TS JAS and the AS CLU & CDL estimates (3). Table 3 also includes, in columns (4-6), the estimate coefficients of variation (CVs).

The same SD ASF parcel data are used for TS and AS. The resulting ASFs created from TS and AS are used for corresponding sample allocation and selection for estimation. Sample allocation is conducted based on a new autostratified ASF and FSA CLUs are used as segments to derive the AS estimates. Consequently, different PSUs are selected and new segments are created using the FSA CLU polygons which are linked to FSA 578 (crop) data to serve as proxy survey data. The two AS estimates are based on the same segments but the FSA CLUs are summarized with 2013 CDL data for AS CLU & CDL estimates.

As illustrated in Table 3, the difference between the TS JAS estimates and the AS CLU & 578 estimates for *All Wheat* are relatively small at the state level while stratum differences are larger, particularly for the cultivated strata. The stratum level differences are larger than at the state level because different PSUs are included in different strata, based on the different methods and different segments are utilized for data collection.

Based on post stratification results (Table 3) state level TS and AS *Corn* and *Soybean* estimates are relatively close. Further the TS and the AS CLU & 578 estimates are relatively close for *All Wheat*. There are no *All Wheat* AS CLU & CDL estimates because there is no *All Wheat* CDL category. The *Spring Wheat* and *Winter Wheat* estimates are very different between the TS JAS and AS CLU & 578 estimates. However, the AS CLU & 578 estimates for *All Wheat*, and both AS estimates for *Spring Wheat* and *Winter Wheat* are all much closer to the NASS published estimates than the TS JAS estimates. The AS FSA CLU & 578 estimates have lower CVs than the TS JAS for 13 out of 20 strata estimates. The AS FSA CLU & CDL estimates have lower CVs than the TS JAS for 10 of 16 strata estimates. The *All Wheat*, *Corn, Spring Wheat* and *Winter Wheat* AS FSA CLU & 578 estimates have lower state CVs indicating improved precision with the AS estimates.

These results imply that FSA CLU data are useful for a quick, low cost estimation assessment. However, FSA CLU data are not a perfect source for estimation assessment since it was observed that FSA data have about a 5.0 % rate for non-matching errors for *Corn* and *Soybeans*.

There are large differences between the TS and both AS strata estimates for *Winter Wheat* and *Spring Wheat* (Table 3). However, the two AS estimates for *Winter Wheat* and *Spring Wheat* are much closer to the NASS published estimates than the TS JAS estimates. The differences between *Corn* and *Soybean* state estimates derived from the TS and both AS stratifications are smaller. However, relatively large differences in the strata estimates are observed between TS JAS and both AS estimates. In addition, the AS CLU & 578 state estimates for *All Wheat, Corn, Spring Wheat* and *Winter Wheat* have reduced CVs when compared with the TS JAS estimates for these crops.

5. CONCLUSION

Stratification accuracy and post stratification estimation comparisons of the NASS TS and a new CDL AS method are conducted in this paper. Results indicate that stratification accuracies of the intensely cultivated strata are statistically significantly higher using AS than TS. All remaining strata and state accuracy differences are not statistically significant. Stratum percent cultivation STDs, based on the AS are all lower than the TS STDs indicating that strata defined using AS are more homogeneous.

Overall, post stratification estimate results obtained from the AS are comparable to TS, but not improved based on percent differences. In considering the statistically significant improvement in ASF accuracy, reported using the AS in all strata with greater than 15% cultivation in five test states [4]; the higher accuracy for the intensely cultivated strata in this study; and improved stratum homogeneity; the new CDL AS method is concluded to provide a better solution for generating ASFs; that are more objective, efficient, accurate, homogenous; and reduces labor costs.

Table 3. South Dakota 2013 Traditional vs. Automated Stratification Crop Estimate Difference (in Percentage) Comparison – *Estimates are in Planted Acres and not published at the stratum level.

Crop	Stratum	NASS Official Estimate (1)*	Difference TS JAS and AS CLU & FSA 578 (2)	Difference TS JAS and AS CLU & CDL (3)	TS JAS CV (4)	AS CLU & FSA 578 CV (5)	AS CLU & CDL CV (6)
All Wheat	11	N/A	63.41%	N/A	11.2	14.6	N/A
	12	N/A	173.25%	N/A	23.7	17.0	N/A
	20	N/A	8.69%	N/A	22.5	20.2	N/A
	40	N/A	14.45%	N/A	69.6	43.9	N/A
	State	***2,490,000***	**16.85%**	**N/A**	**13.1**	**11.1**	**N/A**
Corn	11	N/A	44.18%	40.60%	5.4	4.4	3.9
	12	N/A	210.18%	258.23%	12.9	8.6	7.5
	20	N/A	42.26%	60.22%	16.9	12.3	11.9
	40	N/A	23.55%	39.44%	60.3	41.3	45.2
	State	***6,200,000***	**6.74%**	**2.85%**	**4.9**	**4.3**	**3.9**
Soybeans	11	N/A	51.06%	46.25%	5.6	5.6	4.9
	12	N/A	125.19%	144.51%	14.6	11.5	9.8
	20	N/A	78.93%	75.62%	18.8	24.0	22.7
	40	N/A	N.A	N/A	0.00	51.9	52.0
	State	***4,600,000***	**19.03%**	**12.79%**	**5.0**	**6.0**	**5.2**
Spring Wheat	11	N/A	68.58%	12.79%	14.2	21.1	19.8
	12	N/A	95.56%	66.37%	28.4	26.6	24.9
	20	N/A	42.57%	97.60%	31.6	32.1	27.2
	40	N/A	66.70%	31.51%	92.3	42.7	47.4
	State	***1,190,000***	**47.10%**	**74.75%**	**17.6**	**14.3**	**13.3**
Winter Wheat	11	N/A	51.39%	67.22%	20.8	20.9	23.3
	12	N/A	346.30%	210.08%	35.4	22.1	23.0
	20	N/A	88.01%	39.52%	27.8	27.4	25.5
	40	N/A	107.37%	69.37%	88.8	58.6	69.1
	State	***1,300,000***	**46.34%**	**7.24%**	**17.9**	**16.8**	**19.0**

6. REFERENCES

[1] J. Cotter, C. Davies, J. Nealon, and R. Roberts, *Area Frame Design for Agricultural Surveys, in Agricultural Survey Methods* (eds R. Benedetti, M. Bee, G. Espa and F. Piersimoni), John Wiley & Sons, Ltd, Chichester, UK. 2010. DOI: 10.1002/9780470665480.ch11,

[2] J. J. Cotter and C. M. Tomczac, "An image analysis system to develop area sampling frames for agricultural surveys," *Photogrammetric Engineering and Remote Sensing*, 60 (3): 299-306, 1994. Washington, DC. USDA.< http://handle.nal.usda.gov/10113/21452 Last accessed 23/12/2015.

[3] FAO, *"Multiple Frame Agricultural Surveys. Volume 2:* "Agricultural survey programmes based on area frame or dual frame (area and list) sample designs." FAO Statistical Development Series no. 10. Rome: FO. 274, p. 1998

[4] C. Boryan, Z. Yang, L. Di, and K. Hunt, "A New Automatic Stratification Method for U.S. Agricultural Area Sampling Frame Construction Based on the Cropland Data Layer," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* pp. 1939-1404, May 2014, DOI: 10.1109/JSTARS.2014.2322584.

[5] C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring US Agriculture: The US Department of Agriculture, National Agricultural Statistics Service Cropland Data Layer Program," *Geocarto* International 26, (5): 341-358, 2011.

[6] C. Boryan., Z. Yang, and L. Di. "Deriving 2011 Cultivated Landcover Data sets using USDA National Agricultural Statistics Service historic Cropland Data Layers," *Proc. of IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany. July 22-27, 2012.

[7] FSA, Farm Service Agency Common Land Unit Information worksheet.<http://www.fsa.usda.gov/Internet/FSA_File/clu__infosheet_2013.pdf> (last accessed 11, Nov 2015), 2014.