A NOVEL METHOD FOR AREA FRAME STRATIFICATION BASED ON GEOSPATIAL CROP PLANTING FREQUENCY DATA LAYERS

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

ABSTRACT

This paper proposes a novel method for land cover area frame stratification based on corn planting frequency and percent cultivation. South Dakota U.S. geospatial crop frequency (2008-2013) and cultivation (2013) data layers created from NASS Cropland Data Layers are utilized to develop a novel area sampling frame (ASF) stratification design. Eight corn planting frequency strata are derived using a k-means clustering method based on mean corn planting frequency calculated at the NASS ASF primary sampling unit level. The corn planting frequency strata are then sub stratified based on percent cultivation, which, together, provide more crop specific information than the current NASS ASF based on percent cultivation alone. Using 2014 Farm Service Agency Common Land Unit Data as *in situ* validation, it is found that this novel ASF design predicts crop specific planting patterns well. These results indicate that the new stratification method has potential to improve ASF accuracy, efficiency and crop estimates.

Index Terms— Geospatial Crop Frequency Data Layers, area sampling frame, stratification, Cropland Data Layer, Cultivation Layer

1. INTRODUCTION

Area sampling frames (ASFs) are the foundation of the agricultural statistics program of the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) and many other statistical survey programs around the world. The stratification of land cover, used in the construction of the NASS Area Frames, is based on percent cultivated cropland. However, percent cultivation information does not provide predictive crop planting information, which is critical for estimation of specific crops.

This paper presents a novel land cover stratification using NASS geospatial crop planting frequency data layers [1], and the NASS cultivation layer [2] which are derived from historic Cropland Data Layers (CDLs) [3], to improve the accuracy, objectivity, efficiency of the NASS ASFs and ultimately crop survey estimates. The goal of this research is to determine 1) if the distribution of the primary sampling units (PSUs) in the new ASF design based on corn planting frequency and cultivation, is consistent with the PSU distribution in the current NASS ASF based on percent cultivation alone and 2) whether the corn frequency and cultivation stratification can predict patterns of future crop specific planting.

The 2008-2013 crop frequency data are used to derive objective crop-specific planting frequency statistics for NASS's ASF PSUs. The new stratification method based on crop frequency is then proposed. A k-means clustering method is used to group the ASF PSUs into eight primary strata based on corn frequency means with each PSU belonging to the strata with the nearest corn frequency mean [4], [5]. The corn planting frequency strata are then sub stratified based on percent cultivation resulting in 32 unique strata which provide more crop specific information.

The stratification results derived from the corn frequency and cultivation layers are evaluated using 2014 FSA CLU data for percent cultivation accuracy. The distribution of the PSUs in the new ASF design based on corn planting frequency and cultivation is compared with the PSU distribution in the current NASS ASF based on percent cultivation alone. Finally, the predictability of the corn frequency and cultivation stratification to the patterns of future crop specific planting is also verified by the 2014 FSA CLU data.

2. DATA

In this study, South Dakota (SD) U.S. was selected as the study area due to the range of crops and the availability of USDA FSA CLU data, which covers approximately 98% of the state. The geospatial datasets used include: 1) NASS SD ASF, 2) SD (2008-2013) Cropland Data Layers (CDLs), 3) SD (2008-2013) Corn Planting Frequency Layer, 4) SD 2013 Cultivation Layer and 4) USDA SD 2014 FSA CLU data. The NASS ASFs are made up of stratified land parcels, known as PSUs, which are digitized to permanent physical boundaries (roads, railroads, and rivers) on the ground [6]. The NASS CDLs are annually updated 30-56.0 meter geo-referenced, cropland classifications with accuracies ranging from 85% to 95% for major crops [3]. The archive of 1997 - 2014 CDL data is publically available from NASS's online geospatial application - CropScape [7]. The SD 2008 – 2013 crop frequency layers were derived from 2008-2013 CDL data. They capture planting frequency at the pixel level over time [1]. When intersected at the PSU level, the crop frequency data provide unique information on planting probabilities. In this research, corn is the focus as it is the primary state crop and rotated closely with soybeans. The 2008-2013 corn frequency and the 2013 cultivation layers were selected for stratification study. Further, cultivation data provide supplemental information to identify crop specific planting patterns. The FSA CLU data are GIS layers which provide agricultural information describing the nation's farms and support farm commodity, conservation programs and disaster response [8], [9].

3. METHODOLOGY

The proposed new method for area frame stratification is based on geospatial crop planting frequency and cultivation data layers. Corn is the primary crop grown in SD and is selected as the basis for the stratification design. The method to create and validate the crop frequency layers is described in Reference [1]. The method to create and validate the NASS cultivation layer is described in Reference [2]. The detailed steps to develop the corn planting frequency and cultivation stratification are as follows:

1) Load SD corn frequency layer.
2) Load a SD ASF PSU boundary layer and the 2013 cultivation layer individually.
3) Overlay an ASF PSU boundary on the corn frequency layer and on the cultivation layer.
4) Compute the corn frequency mean and percent cultivation of each ASF PSU within the PSU boundary.
5) Add new attributes with the values of the computed corn frequency mean and the percent cultivation to the SD ASF PSU boundary layer for all PSUs.
6) Run a k-means clustering algorithm based on the corn frequency mean variable (no spatial constraint and random seeds) to create eight strata. The number of strata selected for this research is the same number in the current NASS SD Area Frame based on percent cultivation.
7) Sub-stratify the eight corn strata, based on corn planting mean frequency, by percent cultivation using the NASS definitions of stratum 11 (> 75% cultivation), stratum 12 (> 50% - ≤ 75%), stratum 20 (> 15% - ≤ 50%), and stratum 40 (≥ 0.0 - ≤ 15%).

For this study, the new corn frequency and cultivation strata were compared to 2014 FSA CLU data to evaluate the new ASF stratification design. The 2014 FSA CLU data are a comprehensive source of *in situ* data reported in 2014. The crop frequency and cultivation layers are derived from 2008-2013 CDLs. The new stratification design was evaluated based on 1) mean corn acreage increasing from stratum 1 – 8 in 2014, 2) how accurately the new frame's PSU percent cultivation matches the percent cultivation reported in the 2014 FSA CLU data and 3) if the corn frequency and cultivation stratification accurately predicts crop specific planting patterns in 2014.

TABLE 1. SOUTH DAKOTA CORN PLANTING FREQUENCY STRATA AND 2014 FSA CLU REPORTED CORN ACREAGE

STRATA	*# PSUS*	*CORN FREQUENCY RANGE*	*MEAN CORN ACRES*	*MEAN CORN PERCENT*
1	*4634*	*0.00-0.2558*	*58.402*	*2.575*
2	*2443*	*0.256-0.6508*	*203.782*	*12.041*
3	*2497*	*0.651-1.0550*	*288.300*	*19.077*
4	*2500*	*1.0551-1.4593*	*400.515*	*26.975*
5	*2268*	*1.4596-1.8645*	*490.888*	*32.978*
6	*1909*	*1.8651-2.2901*	*638.074*	*39.241*
7	*1544*	*2.2902-2.8491*	*812.784*	*44.667*
8	*356*	*2.8506-4.8648*	*922.534*	*51.578*
1-8	*18151*	*0.000-4.8648*	*N/A*	*N/A*

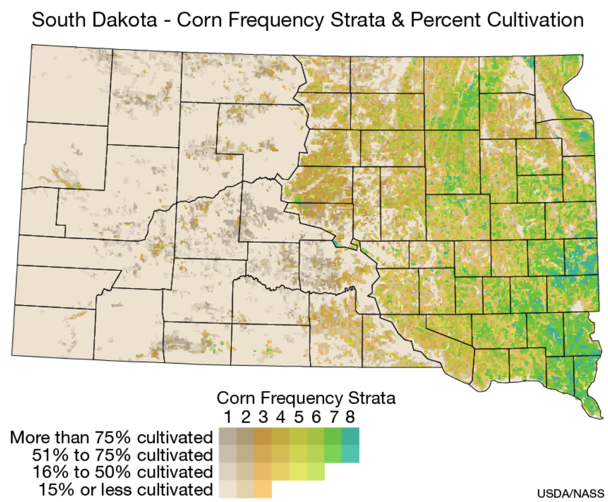

FIGURE 1: SOUTH DAKOTA CORN PLANTING FREQUENCY AND PERCENT CULTIVATION AREA SAMPLING FRAME

4. RESULTS AND DISCUSSION

The SD corn planting frequency and cultivation stratification utilized the SD 2008-2013 corn frequency and the SD 2013 cultivation data layers which were validated using 2008-2013 FSA CLU data. The overall accuracy of the SD corn frequency layer is 92.33% and of the 2013 SD cultivation layer is 96.07%. Table 1 summarizes the eight corn planting frequency strata, # PSUs, mean corn acres and percent corn reported in the 2014 FSA CLUs for each stratum. Results indicate that the corn frequency data created using the 2008-2013 CDLs are reliable predictors of corn planting with increasing quantities of corn acreage reported from stratum 1 (58.402 mean corn acres) to stratum 8 (922.534 mean corn acres) in 2014, while the number of PSUs decreases from 4634 to 356 as the corn planting frequency stratum goes up. This means that higher corn planting frequency indicates higher corn acreage. Moreover, the mean percent corn planting coverage per PSU increases from 2.575 to 51.578

acres as the corn planting frequency increases from stratum 1 to stratum 8.

TABLE 2. ILLUSTRATES THE EIGHT PRIMARY CORN FREQUENCY STRATA SUB STRATIFIED BASED ON PERCENT CULTIVATION (A-D) WITH ACCURACIES DETERMINED USING THE 2014 FSA CLU DATA

Strata	*Corn Range*	*# PSUs*	*% Cultivation*	*Mean % cultivation*	*% Cultivation Accuracy*	*Mean Corn Acres*	*Mean % Corn*	*Mean Soybean Acres*	*Mean % Soybean*	*Mean Wheat Acres*	*Mean % Wheat*
1	**0.00-0.2558**	**4634**			**81.72%**	**58.402**	**2.575**	**19.215**	**1.033**	**159.340**	**7.877**
1a		293	>75%	85.73		111.026	6.722	10.374	0.957	806.993	44.623
1b		379	>50 % - 75%	60.38		78.733	5.978	10.187	1.118	341.795	24.919
1c		1291	>15 % - 50%	27.07		79.436	3.836	21.673	1.419	210.356	9.533
1d		2671	≥ 0.0 - 15%	3.13		39.578	1.027	20.278	0.843	37.748	0.628
2	**0.256-0.6508**	**2443**			**76.79%**	**203.782**	**12.041**	**135.510**	**8.5208**	**190.753**	**11.236**
2a		214	>75%	85.35		228.739	14.364	48.0522	3.527	647.612	37.858
2b		376	>50 % - 75%	60.45		200.503	14.654	102.909	8.920	301.790	20.194
2c		1768	>15 % - 50%	29.72		205.263	11.465	155.78	9.162	119.998	6.592
2d		85	≥ 0.0 - 15%	11.23		124.657	6.614	78.198	5.998	21.061	1.197
3	**0.651-1.0550**	**2497**			**76.13%**	**288.300**	**19.077**	**247.204**	**16.556**	**190.753**	**12.217**
3a		325	>75%	86.67		321.878	19.284	209.854	12.433	637.358	36.386
3b		881	> 50% - 75%	59.74		314.189	21.898	280.884	19.290	220.855	15.296
3c		1290	> 15% - 50%	37.48		262.382	17.113	233.804	15.741	62.647	3.979
3d		1	≥ 0.0 - 15%	0.46		1.2025	0.193	0.00	0.00	450.682	72.171
4	**1.0551-1.4593**	**2500**			**66.84%**	**400.515**	**26.975**	**378.485**	**25.697**	**186.001**	**11.228**
4a		664	>75%	85.22		463.020	29.215	405.619	26.256	450.022	24.702
4b		1535	> 50% - 75%	61.16		397.023	26.881	394.624	26.777	101.961	7.078
4c		301	> 15% - 50%	43.79		280.436	22.514	236.328	18.952	32.153	2.672
4d		0	≥ 0.0 - 15%	N/A		N/A	N/A	N/A	N/A	N/A	N/A
5	**1.4596-1.8645**	**2268**			**65.43%**	**490.888**	**32.978**	**482.509**	**32.525**	**103.027**	**6.881**
5a		950	>75%	83.73		563.68	35.316	587.571	37.079	187.681	12.039
5b		1291	>50 % - 75%	64.49		440.613	31.341	411.637	29.543	42.734	3.217
5c		27	>15 % - 50%	45.82		333.532	28.987	174.598	14.859	7.366	0.567
5d		0	≥ 0.0 - 15%	N/A		N/A	N/A	N/A	N/A	N/A	N/A
6	**1.8651-2.2901**	**1909**			**59.61%**	**638.074**	**39.241**	**648.721**	**39.690**	**64.709**	**3.974**
6a		1524	>75%	84.08		681.226	40.414	707.220	41.701	75.169	4.519
6b		382	>50 % - 75%	67.88		469.300	34.609	420.242	31.949	23.487	1.827
6c		3	>15 % - 50%	46.88		206.875	32.659	24.014	3.870	0.00	0.00
6d		0	≥ 0.0 - 15%	N/A		N/A	N/A	N/A	N/A	N/A	N/A
7	**2.2902-2.8491**	**1544**			**66.65%**	**812.784**	**44.667**	**774.963**	**43.109**	**29.239**	**1.823**
7a		1475	>75%	86.82		831.616	45.094	795.546	43.705	29.323	1.761
7b		69	> 50% - 75%	69.32		410.222	35.530	334.951	30.377	27.438	3.152
7c		0	> 15% - 50%	N/A		N/A	N/A	N/A	N/A	N/A	N/A
7d		0	≥ 0.0 - 15%	N/A		N/A	N/A	N/A	N/A	N/A	N/A
8	**2.8506-4.8648**	**356**			**69.94%**	**922.534**	**51.578**	**673.690**	**39.174**	**11.791**	**0.827**
8a		347	>75%	89.13		933.077	51.788	682.331	39.211	12.089	0.849
8b		9	> 50% - 75%	73.15		516.047	43.456	340.523	37.757	0.013	0.277
8c		0	> 15% - 50%	N/A		N/A	N/A	N/A	N/A	N/A	N/A
8d		0	≥ 0.0 - 15%	N/A		N/A	N/A	N/A	N/A	N/A	N/A

As illustrated in Fig. 1, dark green colors identify strata with high levels of corn planting frequency and light colors identify strata with low corn planting frequency. The intensity of each color indicates the range of percent cultivation.

Table II illustrates the eight primary corn frequency strata with sub-strata based on percent cultivation (a-d) with accuracies determined using the 2014 FSA CLU data. Table II further identifies, by strata, the corn planting frequency range, # of PSUs, and PSUs' mean percent cultivation, percent cultivation accuracy (based on whether PSU percent cultivation matches the 2014 FSA CLU data summarized at the PSU level), mean corn acres, mean percent corn, mean soybean acres, mean percent soybean, mean wheat acres and mean percent wheat reported in the 2014 FSA CLU data.

The percent cultivation substrata of the new SD ASF design has an overall accuracy of 72.365% (validated using the 2014 FSA CLU data), which is consistent with the accuracy of the current NASS ASF of 74% based on percent cultivation, which is validated using June Agricultural Survey (JAS) data [6].

The new ASF design's PSU distribution is consistent with the PSU distribution of the current NASS ASF. As shown in Table 2, in stratum 1 which has the lowest corn intensity range, the majority of PSUs (2671) are 0.0% - 15% cultivation. In stratum 2 and 3 which also have low corn intensity, the majority of the PSUs (stratum 2 - 1768 and stratum 3 - 1290) are 15% - 50% cultivation. In stratum 4 and 5 which are the mid corn intensity strata, the majority of PSUs (stratum 4 - 1535 and stratum 5 - 1291) are 50% - 75% cultivation. Further, virtually all of the PSUs in the highest corn intensity strata including stratum 6 (1524), 7 (1475) and 8 (347) have >75% cultivation.

The sub-stratification of the eight primary corn frequency strata into 32 total strata provides additional crop specific information at the PSU level than the current NASS ASF based on percent cultivation. The novel ASF design creates strata where high concentrations of corn are likely planted (stratum 7 and 8) which were reported to have mean corn acres at the PSU level of 812.784 acres (stratum 7) and 922.534 acres (stratum 8) and particularly in the PSUs in strata with > 75% cultivation (strata 7a and 8a). The ASF design also identifies where wheat is grown in higher concentrations (strata 1a, 2a and 3a). Corn and soybeans are planted in equivalent quantities, in strata 4a, 4b, 5a, 5b, 6a, and 6b, which were anticipated. Furthermore, additional soybean and wheat planting land cover statistics calculated a at the PSU level and summed at the stratum level, as shown in Table 2, provide additional information regarding the corn and soybean rotation pattern and the distribution of wheat planting in these strata.

Future research will include conducting estimation based on the new ASF stratification design using the 2014 FSA CLU data as proxy survey data which can be compared with the 2014 JAS estimates.

5. CONCLUSION

This paper has presented a novel method for land cover area frame stratification based on corn planting frequency and percent cultivation. The South Dakota (SD) corn planting frequency layers were derived based on 2008-2013 NASS CDL data and the SD cultivation layer was derived using 2009-2013 CDL data. High accuracy of both data sets (Corn Frequency – 93.33%, Cultivation – 96.07%) warrants their use in area frame stratification. The corn planting frequency and cultivation stratification provides additional crop specific information when compared to the current NASS Area Sampling Frame design based on percent cultivation alone. The additional mean corn frequency data provides strong indications where corn is planted in large quantities (strata 7 and 8) and was confirmed using the 2014 FSA CLU data as in situ validation. The cultivation substrata further provide guidance regarding where to sample and expect to locate wheat acreage (strata 1a, 1b, 2a, 2b, 3a and 3b) the third largest crop in the state. In SD corn and soybeans are rotated. Consequently, Strata 4 – 6 identify the areas where equivalent quantities of corn and soybeans are grown. It was observed that the corn planting frequency and cultivation based stratification successfully predicted corn planting patterns as verified by the 2014 FSA CLU data. These results indicate that the new stratification method and the newly developed SD area frame using the geospatial corn planting frequency and cultivation data layers has great potential to improve ASF accuracy, efficiency and crop estimates.

6. REFERENCES

[1] C. Boryan.; Z. Yang.; P. Willis, "US Geospatial Crop Frequency Data Layers," Proc. of the 3rd Intel. Conf. on Agro-Geoinformatics, August 11-14, 2014, Beijing, China.

[2] C. Boryan., Z. Yang , and L. Di. "Deriving 2011 cultivated land cover data sets using usda national agricultural statistics service historic cropland data layers," Proc. of IEEE IGARSS, July 22-27, 2012, Munich, Germany

[3] 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.

[4] J.B. MacQueen, (1967). *Some Methods for classification and Analysis of Multivariate Observations*. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability **1**. University of California Press. pp. 281–297.

[5] E. Forgy, "Cluster analysis of multivariate data: efficiency versus interpretability of classifications". *Biometrics* **21**: 768–769. (1965).

[6] J. Cotter, C. Davies, J. Nealon, and R. Roberts, (2010) 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.

[7] W. Han, Z. Yang, L. Di, and R. Mueller, "CropScape: A web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support." Computer and Electronics in Agriculture, 84: 111–123, 2012.

[8] FSA, 2014. FSA CLU Information worksheet. http://www.fsa.usda.gov/Internet/FSA_File/clu__infosheet_2013.pdf (last accessed 21, May 2015).

[9] J. Heard, 2002. USDA establishes a common land unit. ESRI ArcUser Online. http://www.esri.com/news/arcuser/0402/usda.html (last accessed 20 May, 2015).