Journal of Integrative Agriculture 2017, 16(2): 312–323

Available online at www.sciencedirect.com

ScienceDirect

RESEARCH ARTICLE

Developing crop specific area frame stratifications based on geospatial crop frequency and cultivation data layers

Claire G. Boryan[1], Zhengwei Yang[1], Patrick Willis[1], Liping Di[2]

[1] *Research and Development Division, USDA National Agricultural Statistics Service, VA 22030, USA*
[2] *Center for Spatial Information Science and Systems, George Mason University, VA 22030, USA*

Abstract

Area Sampling Frames (ASFs) are the basis of many statistical programs around the world. To improve the accuracy, objectivity and efficiency of crop survey estimates, an automated stratification method based on geospatial crop planting frequency and cultivation data is proposed. This paper investigates using 2008–2013 geospatial corn, soybean and wheat planting frequency data layers to create three corresponding single crop specific and one multi-crop specific South Dakota (SD) U.S. ASF stratifications. Corn, soybeans and wheat are three major crops in South Dakota. The crop specific ASF stratifications are developed based on crop frequency statistics derived at the primary sampling unit (PSU) level based on the Crop Frequency Data Layers. The SD corn, soybean and wheat mean planting frequency strata of the single crop stratifications are substratified by percent cultivation based on the 2013 Cultivation Layer. The three newly derived ASF stratifications provide more crop specific information when compared to the current National Agricultural Statistics Service (NASS) ASF based on percent cultivation alone. Further, a multi-crop stratification is developed based on the individual corn, soybean and wheat planting frequency data layers. It is observed that all four crop frequency based ASF stratifications consistently predict corn, soybean and wheat planting patterns well as verified by the 2014 Farm Service Agency (FSA) Common Land Unit (CLU) and 578 administrative data. This demonstrates that the new stratifications based on crop planting frequency and cultivation are crop type independent and applicable to all major crops. Further, these results indicate that the new crop specific ASF stratifications have great potential to improve ASF accuracy, efficiency and crop estimates.

Keywords: cropland data layer, crop planting frequency data layers, automated stratification, crop specific stratification, multi-crop stratification

1. Introduction

It is well known that the purpose of stratification is to increase the efficiency of estimators by grouping sampling units into relatively homogeneous strata. Area Sampling Frames (ASFs) have been used by the National Agricultural Statistics Service (NASS) since 1954 as a primary tool for conducting surveys to gather information on crop acreage and other agricultural information. They are considered "the backbone to the agricultural statistics program of the National Agricultural

Received 22 October, 2015 Accepted 9 May, 2016
Correspondence Claire G. Boryan, Tel: +1-202-690-0449, E-mail: Claire.boryan@nass.usda.gov

doi: 10.1016/S2095-3119(16)61396-5

Statistics Service" (Ford *et al.* 1986; Vogel 1995; Nusser and House 2009; Arroway *et al.* 2010; Cotter *et al.* 2010). NASS's primary area frame based survey is the June Area Survey (JAS) in which approximately 11 000 one square mile sample segments are visited by enumerators each year at the beginning of the growing season to collect crop type and acreage information. Estimates of crop acreage and livestock inventories are based on the data collected during the JAS. The NASS stratification is based on a categorization of the land into agricultural intensity groups based on percent cultivation. This traditional stratification of land cover has been conducted using visual interpretation of aerial or satellite data, or topographic maps since the 1950s. This traditional state area frame stratification method has been widely used around the world (Cotter and Tomczac 1994; Hanuschak and Morrisey 1978; Villalabos and Wallace 1998; Cotter *et al.* 2010).

There are different varieties of the area frame stratification method and its applications. The European Union (EU) Monitoring Agricultural Resources (MARS) remote sensing project used the NASS traditional ASF construction method to conduct the Land Use/Land Cover Area Frame Survey (LUCAS) in 1988. They proposed using a regular grid of square segments instead of physical boundaries (roads, railroads and rivers) for segment delineation to reduce cost (Gallego *et al.* 1994). The LUCAS sample units are points defined as 3 m diameter circles (Gallego and Delincè 2010). The original objective of the LUCAS was to provide annual European crop estimates. These objectives were expanded to include the collection of land use and environmental monitoring data (Gallego and Delincè 2010). The Italian Agricultural Survey (AGRIT) program uses an area frame created from a regular grid of points with a resolution of 500 m. Point sampling is used for data collection for the multi-purpose survey to derive combined estimates of crop acreage, yields and land use (Carfagna and Gallego 2005; Benedetti *et al.* 2015). 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. In 2005, the survey was redesigned to provide consistency with the TER-UTI. The survey is now known as the TER-UTI LUCAS (Benedetti *et al.* 2015). A combination of area sampling frames and list sampling frames are used for agricultural surveys in Brazil, Canada, Honduras, the U.S. and Albania, while an ASF alone is used in Argentina, Morocco, Nicaragua, Pakistan, Columbia, Costa Rica, Dominican Republic, El Salvador and Guatemala, France, Italy and Spain (Villalabos and Wallace 1998). In addition, ASFs were used in a variety of agricultural research studies such as: 1) evaluation of the prevalence of brown stem rot in the north central U.S. (Workneh *et al.* 1999), 2) improvement of agricultural ground survey estimates as part of the MARS project launched by the European Union in 1989 (Tsiligirides 1998), 3) comparing the utility of frequently used estimators based on an ASF utilized in agricultural surveys (Faulkenberry and Garoui 1991) and 4) development of crop area estimation at a regional level in the Islamic Republic of Iran (Pradhan 2001).

The majority of area sampling frames are developed using visual interpretation of aerial photography or satellite data for stratification. Geospatial data are not used in a fully automated manner in any of these applications. Boryan *et al.* (2015) recently developed a new automated stratification method that utilized available NASS geospatial data Cropland Data Layers (CDLs) (Boryan *et al.* 2011), rather than using visual interpretation of satellite imagery and aerial photography (traditional method) to define percent cultivation of land areas. The new method resulted in ASFs with significantly improved accuracy, in strata with greater than 15% cultivation, as well as, improved objectivity and efficiency at reduced cost (Boryan *et al.* 2014a). This automated stratification method was integrated into NASS Area Frame construction operations and evolved into a hybrid process which combines automated geospatial stratification with manual editing and review (Boryan *et al.* 2014a; Boryan and Yang 2014).

Information on the geospatial distribution of future crop specific planting is critical for improving agricultural survey estimates (increasing the efficiency), and other agricultural business related decision making, such as agricultural production planning and agricultural product commodity inventory control. The NASS's traditional area sampling frame stratification and the CDL based automated stratification method (Boryan *et al.* 2014a), are based on spatial percent cultivation data, which have very limited predictability of future crop planting. To improve geospatial predicative information regarding future crop planting, Boryan *et al.* (2015) developed US national 30 m resolution crop specific planting frequency data layers using multi-year CDL data (Boryan *et al.* 2014b). These data sets provide crop specific planting frequency information for major crops such as corn, soybean, wheat and cotton. The Crop Planting Frequency Data Layers and Cultivation Layers are highly accurate (USDA, NASS 2016). Using Crop Planting Frequency Data Layers for crop specific area frame stratification is therefore proposed in this paper to improve the predictive capability of the developed area frame so that ultimately the obtained estimates will be more accurate and with smaller coefficients of variation.

In this paper, three single crop (corn, soybean and wheat) specific stratifications are developed individually using the specific Crop Frequency Data Layers of interest. The corn, soybean or wheat specific planting frequency strata are sub-stratified based on percent cultivation resulting in 32 unique strata (per frame). The substratification was conducted to determine if it was possible to identify where "other" crops (other than targeted crops), would be planted. The single crop specific stratifications are not intended for direct use to stratify the NASS ASFs for use in the JAS, in which multiple crops are estimated. However, by conducting substratification of the crop planting frequency based stratifications, with the cultivation data, the single crop specific stratifications are potentially applicable to multi-crop surveys and estimation. In addition, a multi-crop specific stratification is also developed to illustrate how these single Crop Frequency Layers can be used together to create one unified multi-crop specific stratification. This paper presents the following results: 1) an evaluation of the percent cultivation accuracy of the three single crop ASF stratifications in comparison to the current NASS South Dakota ASF, 2) a determination whether the primary sampling unit (PSU) distributions in the three single crop stratifications are consistent with the current NASS ASF based on percent cultivation alone and, most importantly, 3) a determination whether the crop specific stratifications based on 2008–2013 data can predict patterns of future crop specific planting at the PSU level as verified by 2014 Farm Service Agency (FSA) Common Land Unit (CLU) Data (Heard 2002; FSA 2014).

2. Materials and methods

2.1. Study site and experimental data

South Dakota (SD) U.S. is selected as the study area for this investigation because 1) it is a major corn, soybean and wheat producing state and serves as a good example of the range of agriculture grown in the U.S. and 2) the USDA FSA has digitized the SD field boundaries and attributes them annually with crop and non-crop information, known as FSA CLU data, for approximately 98% of SD's land cover. The FSA CLU data are utilized for validation of the crop specific stratifications. The geospatial data used in this research include: the NASS SD 2008–2013 CDLs; the NASS SD 2013 Cultivation Layer, the SD Corn, Soybean and Wheat-Crop Planting Frequency Data Layers; the NASS SD Area Sampling Frame; and 2014 SD Farm Service Agency Common Land Unit data. The Cropland Data Layers are annually updated, georeferenced crop specific land cover classifications with a 30–56 m resolution. The CDLs are 85–95% accurate for major crops in large agricultural states (Boryan *et al.* 2011). The NASS 2013 Cultivation Layer is created using 2009–2013 CDL data and identifies land as cultivated or non-cultivated at a 30-m resolution (Boryan *et al.* 2012). The NASS 2013 national scale Cultivated Layer has a producer accuracy of 97.3% (non-cultivation) and 98.5% (cultivation) and a user accuracy of 97.8% (non-cultivation) and 97.5% (cultivation) (Boryan *et al.* 2012). The 2008–2013 SD Crop Planting Frequency Data Layers identify the number of years that a 30-m pixel is classified as planted to a specific crop over a defined time period in the CDL input data. The accuracies of the national scale Crop Planting Frequency Data Layers are 91.00, 90.13, 87.67 and 85.96% for corn, cotton, soybeans and wheat, respectively (Boryan *et al.* 2014b). Fig. 1-A illustrates the U.S. NASS 2013 Cultivated Layer. The black box identifies the state of SD. The dark pixels identify cultivation and the light pixels identify non-cultivation. Fig. 1-B illustrates the SD 2008–2013 Corn Planting Frequency Data Layer with a zoom that indicates the levels of corn planting frequency from 2008–2013. In the Crop Planting Frequency Data Layers, the darker shades identify pixels that are planted more years to the specific crop. The lighter shades indicate lower planting frequency from 2008–2013. The illustration on Fig. 1-C is the SD 2008–2013 Soybean Planting Frequency Data Layer. Illustration (Fig. 1-D) is the SD 2008–2013 Wheat Planting Frequency Data Layer. In SD, corn and soybeans are commonly rotated in the same areas and wheat is planted more often in central and western SD.

In SD, a relatively small amount of acreage is planted to multiple crops in the same year, due to the short growing season. For example, approximately 11 054 acres of double crop winter wheat and corn, 406.0 acres of double crop corn and soybeans, and 3 658 acres of double crop winter wheat and soybeans were grown in 2014 out of a total of 43 257 079 million acres of farmland in SD (Census of Agriculture 2012). As shown in Table 1, crops planted more than once during a growing season are defined in one single category such as CDL code 225 and 226. For this assessment, 30 m pixels (0.22 acre) that are identified as planted to two crops in the same year are included in both Crop Planting Frequency Data Layers. Table 1 illustrates the CDL categories that are recoded to create the Corn, Soybean and Wheat Planting Frequency Data Layers (Boryan *et al.* 2014b).

The state level ASFs, as illustrated in Fig. 2-A, in shape file format, are constructed with PSUs defined by physical features on the ground (roads, railroads and rivers) and stratified based on percent cultivation (all crop types included). The new crop specific frequency data layer based area frames are constructed based on the existing PSUs. As illustrated in Fig. 2-B, the FSA CLU data are standardized geographic information system (GIS) data layers which are updated annually with farmer reports that provide crop type and acreage information of the nation's farms and fields that were established to support farm commodity, conservation and disaster response (Heard 2002; FSA 2014). The NASS

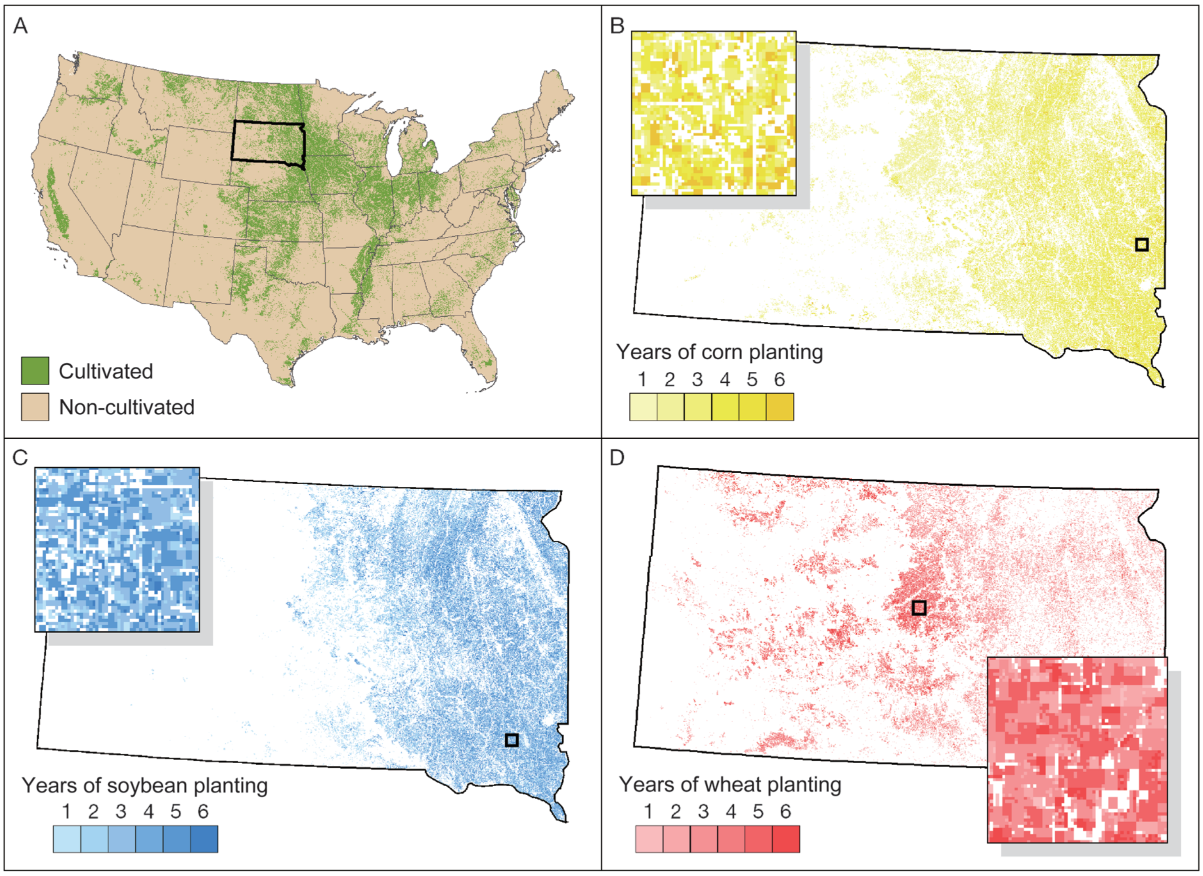

Fig. 1 A, United States 2013 Cultivated Layer. Black box identifies South Dakota. B, South Dakota Corn Planting Frequency Data Layer (2008–2013). C, South Dakota Soybean Planting Frequency Data Layer (2008–2013). D, South Dakota Wheat Planting Frequency Data Layer. The small black boxes in B–D identify the location of the zoomed areas in the Corn, Soybean and Wheat Planting Frequency Data Layers. The darker shades in B–D identify pixels with higher planting frequency and lighter shades identify pixels with lower planting frequency.

CDLs, the Crop Planting Frequency and Cultivated Layer, with metadata, are publically available on the NASS web GIS application CropScape (Han *et al.* 2012).

2.2. Methodology

In this study, 2008–2013 Crop Frequency Data Layers are used to derive single and multiple crop-specific stratifications based on ASF PSUs rather than using Cultivation Data Layers. Crop Frequency Data Layers are used to automatically calculate crop specific planting frequency means at the PSU level for stratification because the Crop Planting Frequency Data Layers provide improved predictive information of future crop specific planting. In this study, the K-means clustering method was selected to group the ASF primary sampling units into eight primary strata based on corn, soybean or wheat planting frequency means (depending on crop targeted) where PSUs are grouped into strata with the nearest crop specific planting frequency mean. The K-means clustering method was selected because it is a robust algorithm, which has been used for many decades and is well documented (Forgy 1965; MacQueen 1967; Hartigan and Wong 1979). Further, the K-means clustering algorithm is readily available in ESRI's ArcGIS software which is used within NASS for Area Frame construction. Crop specific planting frequencies were clustered to define strata that are more homogeneous within strata and heterogeneous among strata (based on the crop specific planting frequency mean values calculated at the PSU level). The detailed procedures to construct the single and multi-crop stratification using the Crop Frequency Data Layers are provided in the following subsections.

2.3. Single crop specific stratifications

The proposed crop specific stratifications are based on the individual Crop Frequency Data Layers. The Cultivation Layer is used for sub-stratification only. Three single crop

Table 1 Cropland Data Layer (CDL) categories recoded to build crop planting frequency data layers

CDL code	Crop description
Corn	
1	Corn
225	Double crop winter wheat/Corn
226	Double crop oats/Corn
237	Double crop barley/Corn
241	Double crop corn/Soybeans
Cotton	
2	Cotton
232	Double crop lettuce/Cotton
238	Double crop winter wheat/Cotton
239	Double crop soybeans/Cotton
Soybean	
5	Soybeans
26	Double crop winter wheat/Soybeans
239	Double crop soybeans/Cotton
240	Double crop soybeans/Oats
241	Double crop corn/Soybeans
254	Double crop barley/Soybeans
Wheat	
22	Durum wheat
23	Spring wheat
24	Winter wheat
26	Double crop winter wheat/Soybeans
225	Double crop winter wheat/Corn
230	Double crop lettuce/Durum wheat
234	Double crop durum wheat/Sorghum
236	Double crop winter wheat/Sorghum
238	Double crop winter wheat/Cotton

specific stratifications are derived independently with the same procedure including 1) a corn frequency stratification, 2) a soybean frequency stratification and 3) a wheat frequency stratification. The detailed steps of the procedure to conduct single crop stratification based on crop specific planting frequency and cultivation, are given as follows:

1) Load a SD Crop Specific (corn, soybeans or wheat) Planting Frequency Layer.

2) Load a SD ASF PSU boundary layer and the 2013 Cultivation Layer individually.

3) Overlay ASF PSU boundary on the Crop Planting Frequency Layer and on the 2013 Cultivation layer.

4) Compute the crop (corn, soybeans or wheat) planting frequency mean and percent cultivation of each ASF PSU within the PSU boundary.

5) Add values of the computed PSU level crop specific frequency mean and the percent cultivation as new attributes to the SD ASF PSU boundary layer for all PSUs.

6) Conduct a K-means clustering based on the crop specific frequency mean variable (no spatial constraint) to create eight primary strata.

7) Sub-stratify the eight primary crop specific strata, based on crop planting mean frequency ranges, by percent cultivation into four substrata including: 1) >75% cultivation, 2) >50–≤75%), 3) >15–≤50%, 4) ≥0.0–≤15%.

For this assessment, all SD ASF PSUs are grouped into eight strata because this is the same number of strata in the current NASS SD Area Frame. Eight strata were determined to provide a sufficient number of homogeneous groups to improve efficiency while maintaining a sufficiently large sample size within strata. Utilizing the same number of strata, as the current SD ASF, will facilitate the comparison of the new crop frequency stratification method, with traditional stratification, based on percent cultivation, in the future. Finally, the corn, soybean or wheat planting frequency strata are substratified, based on percent cultivation, resulting in 32 unique strata (per frame). The substratification based on cultivation was conducted to identify where crops, other than those targeted with the specific Crop Frequency Data Layer, were located. The single crop specific stratifications are potentially applicable to multi-crop surveys and estimation due to the substratification based on percent cultivation.

2.4. Multi-crop specific stratification

In reality, single crop surveys are not common. Most agricultural surveys target multiple crops in a single survey. The single Crop Frequency Layers are naturally for single crop stratification. However, they can be easily used for multi-crop specific stratification. The procedure to create the multiple crop specific stratification based on the crop planting frequency and cultivation are similar to that of single crop stratification. There are three changes in the multi-crop stratification procedure: 1) in the first step, all targeted Crops' (in this study, SD corn, soybean and wheat) Frequency Layers are loaded (no Cultivation Layer); 2) in step 6, the K-means clustering is based on multivariable corn, soybean and wheat planting frequency means which are calculated at the PSU level (no spatial constraint) to create eight primary strata; and 3) there is no sub-stratification step. Every created stratum has three variables: corn, soybean and wheat frequencies. The proportions of crops are different within stratum and among strata. This information will help multivariate sample allocation (Bethel 1989).

2.5. Crop specific stratification assessment

For this research, three single crop specific stratifications and one multi-crop specific stratification were created. The 2014 FSA CLU data were used to assess their predictive capabilities. The 2014 FSA CLU/ 578 data are a comprehensive source of *in situ* data reported in 2014. The three single crop specific stratifications are evaluated based on: 1) mean corn, soybean or wheat acreage increasing from strata 1–8 in 2014, 2) how accurately the new frame's PSU

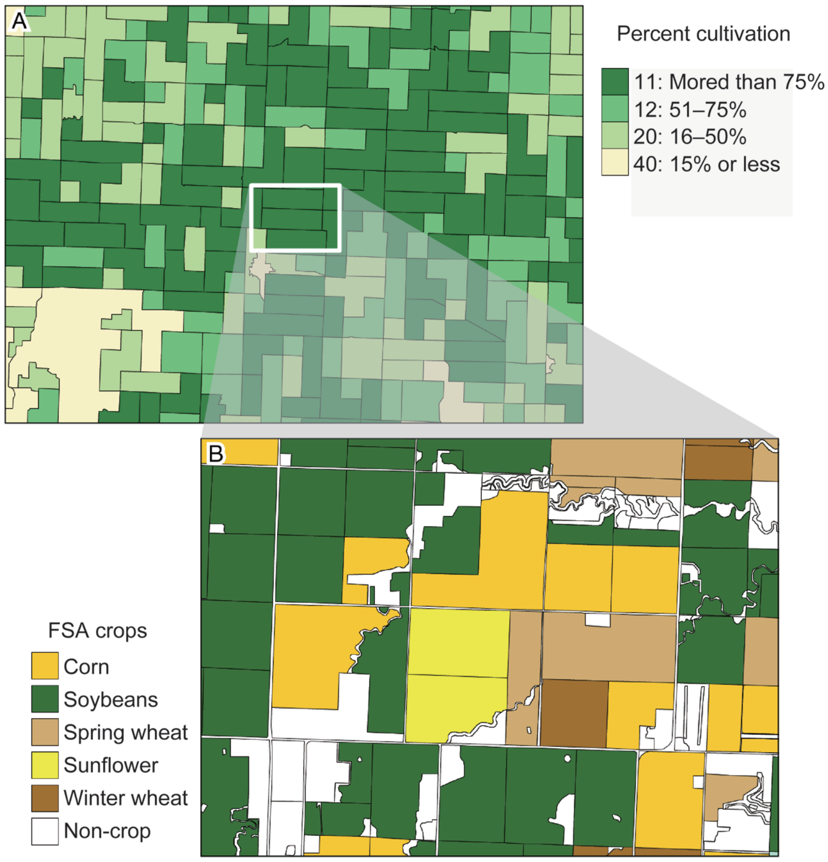

Fig. 2 National Agricultural Statistics Service's (NASS) South Dakota Area Frame Primary Sampling Units (PSUs) stratified by percent cultivation are illustrated in A and South Dakota Farm Service Agency (FSA) Common Land Unit (CLU) ground truth data are illustrated in B with field boundaries and associated crop and non-crop attributes. The white box identifies a zoom area which includes approximately three PSUs to illustrate the FSA Common Land Unit crop fields included in the zoom area.

percent cultivation matches the percent cultivation reported in the 2014 FSA CLU data, 3) if the PSU distribution of the three stratifications are similar to the PSU distribution of the current NASS ASFs and 4) if the corn, soybean and wheat frequency and cultivation stratifications accurately predict crop specific planting patterns in 2014. The multi-crop (corn, soybean and wheat) frequency stratification is evaluated primarily based on how reliably it predicts crop specific planting patterns as well.

3. Results and discussion

3.1. Single crop specific stratifications

In the three single crop specific stratifications, strata are defined by crop specific planting frequency ranges as illustrated in Tables 2–4. For example, in the corn specific stratification as shown in Table 2, stratum 8, the highest frequency range, includes PSUs with a mean corn frequency between 2.85–4.86 years out of 6 (2008–2013) years. Table 2 summarizes the SD eight corn planting frequency strata, PSU distribution, and corn planting frequency ranges that define each stratum, as well as, mean corn acres and percent corn cover reported in the 2014 FSA CLU data by stratum. Results indicate that mean corn planting frequency calculated at the PSU level is a reliable predictor of corn planting with increasing quantities of mean corn acreage reported from strata 1 to 8 in 2014. Further, the number of PSUs, per stratum, decreases from strata 1 to 8 as the corn planting frequency stratum increases. This means that higher corn planting frequency is a strong indication of higher corn acreage. Moreover, the percent corn planting coverage increases from strata 1 to 8 as corn planting frequency increases. These results show the strong predictive power of the newly developed corn specific area frame stratification.

Fig. 3 illustrates the spatial distribution of the strata derived from planting frequency and percent cultivation for three crops in SD U.S. Fig. 3-A illustrates the corn strata distribution. The dark green colors identify PSUs, grouped into strata with high corn planting frequency (strata 7 and 8) located predominantly in south eastern SD. The light green, yellow and orange illustrate the mid corn planting frequency

Table 2 South Dakota corn planting frequency strata based on corn frequency ranges, frame primary sampling unit (PSU) distribution, and 2014 Farm Service Agency reported mean corn acres/percent corn coverage by stratum

Strata	No. of PSUs	Corn frequency range	Mean corn acres	Mean corn (%)
1	4634	0.000–0.2558	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

(strata 3–5) located in central SD and the brown colors identify low levels (strata 1–2) of corn planting frequency. Color saturation is used for the three crop specific stratifications to illustrate stratum percent cultivation. The dark shades identify PSUs with the highest percent cultivation and the light shades represent low levels of cultivation. Specifically, the darkest shade identifies PSUs which are greater than 75% cultivation, while the lightest shade identifies PSUs with 15% or less cultivation.

Table 3 summarizes the soybean planting frequency strata, PSU distribution, the soybean planting frequency ranges by stratum, 2014 FSA CLU reported mean soybean acres and mean percent soybean coverage by stratum. As shown in Table 3, the soybean planting frequency statistics are also a reliable predictor of soybean planting with increasing quantities of mean soybean acreage reported from strata 1 to 8. Mean percent soybean coverage by stratum also increases from strata 1 to 8. Fig. 3-B illustrates the distribution of SD soybean planting and percent cultivation. Dark blue colors identify strata with high soybean planting frequency (strata 7 and 8). The spatial distribution of soybean planting in SD

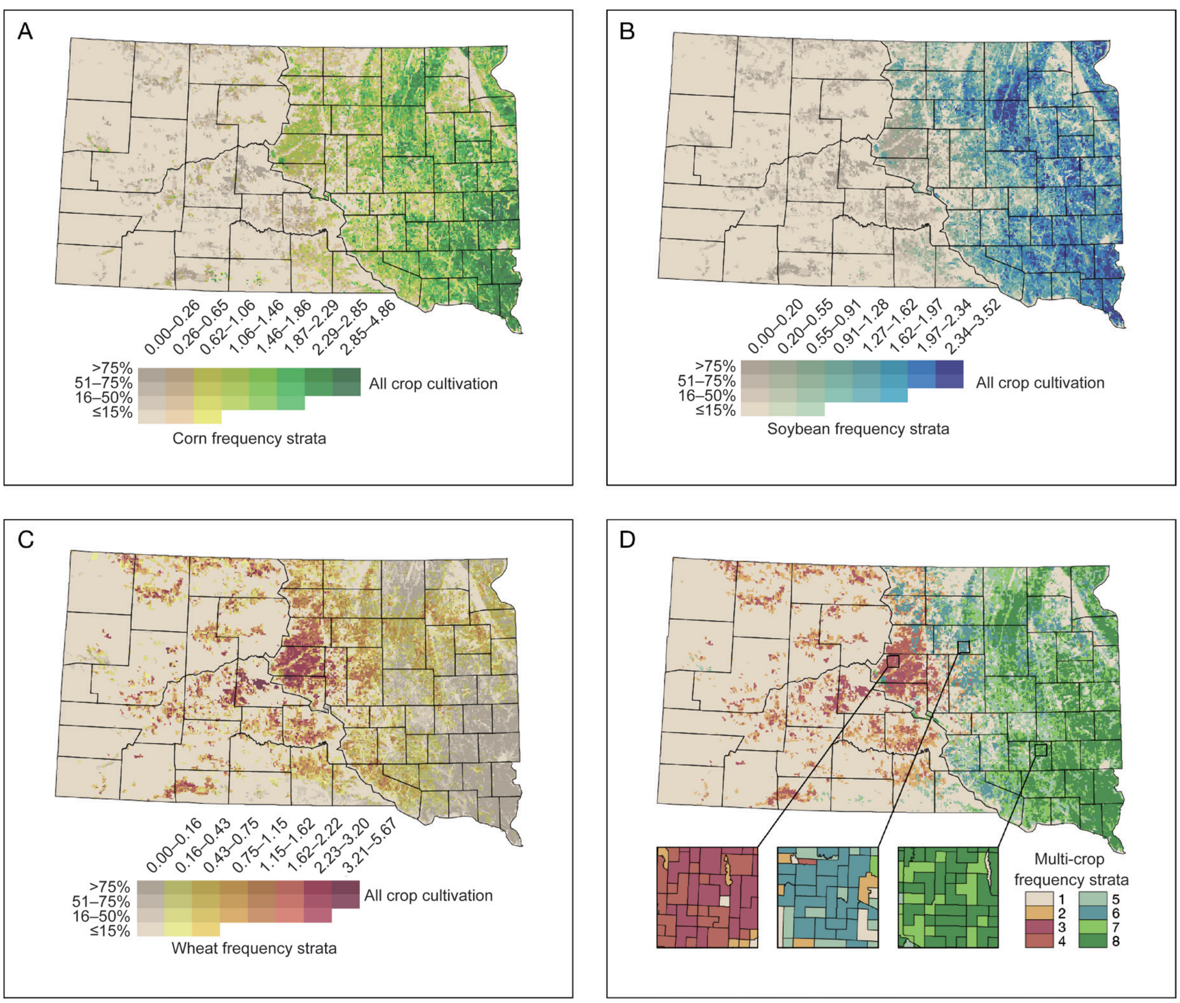

Fig. 3 South Dakota Crop Specific Area Frame Stratifications. A, corn frequency and percent cultivation. B, soybean frequency and percent cultivation. C, wheat frequency and percent cultivation. D, multi-crop corn, soybean and wheat frequency stratification.

Table 3 South Dakota soybean planting frequency strata based on soybean frequency ranges, frame primary sampling unit (PSU) distribution, and 2014 Farm Service Agency Common Land Unit reported mean soybean acres/percent soybean coverage by stratum

Strata	No. of PSUs	Soybean frequency range	Mean soybean acres	Mean soybean (%)
1	6364	0.0000–0.1989	26.7632	1.4916
2	1853	0.1995–0.5479	215.7756	13.4733
3	1835	0.5481–0.9111	287.5780	19.7053
4	1803	0.9119–1.2745	393.9005	26.9931
5	1873	1.2749–1.6228	471.4280	31.8156
6	1803	1.6233–1.9715	599.8244	37.4098
7	1676	1.9721–2.3448	732.8084	42.0424
8	944	2.3456–3.5209	796.5825	47.0770

Table 4 South Dakota wheat planting frequency strata based on wheat frequency ranges, frame primary sampling unit distribution, and 2014 Farm Service Agency Common Land Unit reported mean wheat acres/percent wheat coverage by stratum

Strata	No. of PSUs	Wheat frequency range	Mean wheat acres	Mean wheat (%)
1	8112	0.0000–0.1626	23.255	0.968
2	3640	0.1629–0.4262	86.111	4.892
3	2467	0.4268–0.7499	174.050	9.910
4	1521	0.7503–1.1463	251.572	15.682
5	1006	1.1469–1.6249	346.729	23.217
6	749	1.6254–2.2204	501.761	31.395
7	550	2.2256–3.1929	752.500	40.113
8	106	3.2129–5.6657	1002.771	56.010

closely mirrors corn's spatial distribution but soybeans are not planted as intensely in Central SD. The light blue and green color ranges indicate mid and low levels of soybean planting frequency. In Western SD, very little soybeans are planted as evidenced by Fig. 3-B.

Table 4 summarizes similar results for the eight wheat frequency strata. Similar to the Corn and Soybean Planting Frequency Data Layers, the Wheat Planting Frequency Data Layer is a reliable predictor of wheat planting in 2014 with reported acreage increasing from strata 1 to 8. The percent of land planted to wheat in 2014 increases predictably from strata 1 to 8, as the wheat planting frequency ranges increase. Fig. 3-C illustrates the spatial distribution of the SD wheat crop and percent cultivation. Wheat is planted with high frequency in Central SD (strata 7 and 8) and in Western SD and at a low frequency in the far West and East of SD (strata 1 and 2). As shown in Fig. 3, the derived single crop specific stratifications provide more detailed geospatial distribution information of crop specific planting frequency and cultivation for sample selection when compared with an ASF stratification based on percent cultivation alone.

More detailed analysis of the soybean frequency and cultivation stratification results are summarized in Table 5 as an example. The detailed results of the corn frequency/cultivation (Boryan *et al.* 2015) and wheat frequency/cultivation stratifications are very similar to the soybean results. Therefore, the detailed analysis results of the corn and wheat stratification will not be included in this paper. Table 5 illustrates the eight primary soybean frequency strata with substrata based on percent cultivation (a–d) with accuracies determined using the 2014 FSA CLU data. Table 5 further identifies by strata the soybean planting frequency range, number of 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 values for corn acres, percent corn, soybean acres, percent soybean, wheat acres and 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 the JAS data (Cotter *et al.* 2010). It should be noted that accuracy of cultivation is usually higher than the accuracy of an individual crop classification due to the error cancellation of the cultivation calculation.

The new crop specific stratification's PSU distribution is consistent with the PSU distribution of the current NASS ASF. As shown in Table 5, in stratum 1, which has the lowest soybean frequency range, the majority of PSUs are 0.0–15% cultivation. In strata 2 and 3, which also have low soybean frequency, the majority of the PSUs are 15–50% cultivation. In strata 4 and 5 which are the mid soybean intensity strata, the majority of PSUs are 50–75% cultivation. Further, virtually all PSUs in the highest soybean intensity strata are greater than 75% cultivation.

The soybean frequency/percent cultivation stratification indicates that soybean and corn are rotated closely, but they are not always grown together. As soybean planting increases by strata, corn planting also increases in parallel. In stratum 1, soybean planting is low and corn planting is higher but in strata 3–8 the acreage and percentages are comparable. Further, wheat planting is almost a compliment of soybean planting in SD with the highest levels of wheat planting in strata 1 and 2 of the soybean frequency/percent cultivation stratification and the lowest levels of wheat in stratum 8.

The substratification of eight primary soybean frequency strata into 32 total strata provides more PSU level crop specific information than the current NASS ASF based on percent cultivation alone. The novel ASF stratification creates strata where high concentrations of soybeans are likely planted (strata 7 and 8) and particularly in the PSUs in strata with greater than 75% cultivation (strata 7a and 8a). The crop specific stratification also identifies where wheat is

Table 5 The primary eight soybean frequency strata with substrata based on percent cultivation (a–d)

Strata	Soybean frequency range	No. of PSUs	Cultivation (%)[1)]	Cultivation accuracy (%)	Mean soybean acres	Mean soybean (%)[1)]	Mean corn acres	Mean corn (%)[1)]	Mean wheat acres	Mean wheat (%)[1)]
1	0.0000–0.1989	6364		78.6	26.76	1.49	112.28	6.01	214.63	11.61
1a		829	>75		33.27	2.04	225.20	13.59	704.83	40.07
1b		876	>50–75		23.54	1.78	167.11	11.80	327.52	23.06
1c		1971	>15–50		34.62	2.03	136.91	6.93	199.42	9.47
1d		2688	≥0.0–15		20.04	0.83	41.52	1.11	37.81	0.66
2	0.1995–0.5479	1853		78.7	215.78	13.47	267.07	16.48	178.25	10.27
2a		205	>75		240.12	15.10	458.09	26.33	688.80	35.32
2b		331	>50–75		263.27	19.93	364.62	26.53	251.03	17.37
2c		1249	>15–50		205.44	11.87	220.65	12.86	83.72	4.78
2d		68	≥0.0–15		100.99	6.68	69.01	4.36	21.19	1.11
3	0.5481–0.9111	1835		77.9	287.58	19.71	303.62	20.49	118.34	7.50
3a		227	>75		387.23	25.99	459.30	29.95	385.89	21.94
3b		559	>50–75		331.94	23.59	354.69	24.84	132.20	9.17
3c		1048	>15–50		242.56	16.23	242.94	16.11	53.10	3.48
3d		1	≥0.0–15		0.00	0.00	20.92	28.28	0.00	0.00
4	0.9119–1.2745	1803		65.1	393.90	26.99	389.71	26.27	102.08	6.59
4a		318	>75		528.24	33.25	556.85	34.76	205.74	11.73
4b		1121	>50–75		400.44	27.15	390.63	25.96	88.69	6.00
4c		364	>15–50		256.41	21.04	240.88	19.83	52.76	3.85
4d		0	≥0.0–15		0.00	0.00	0.00	0.00	0.00	0.00
5	1.2749–1.6228	1873		64.2	471.43	31.82	476.03	31.56	83.09	5.64
5a		585	>75		603.70	37.52	615.29	37.23	134.07	8.42
5b		1245	>50–75		419.46	29.56	419.20	29.13	60.99	4.41
5c		43	>15–50		176.70	19.58	226.87	24.83	29.30	3.44
5d		0	≥0.0–15		0.00	0.00	0.00	0.00	0.00	0.00
6	1.6233–1.9715	1803		62.2	599.82	37.41	591.75	36.62	78.45	5.20
6a		1130	>75		690.28	40.31	678.82	38.78	94.20	5.93
6b		668	>50–75		449.84	32.60	448.11	33.14	51.36	3.88
6c		5	>15–50		195.35	24.40	101.26	12.15	137.26	16.59
6d		0	≥ 0.0–15		0.00	0.00	0.00	0.00	0.00	0.00
7	1.9721–2.3448	1676		60.8	732.81	42.04	726.33	41.85	63.82	3.94
7a		1565	>75		756.07	42.55	749.36	42.28	64.49	3.90
7b		111	>50–75		404.83	34.92	401.51	35.78	54.38	4.63
7c		0	>15–50		0.00	0.00	0.00	0.00	0.00	0.00
7d		0	≥0.0–15		0.00	0.00	0.00	0.00	0.00	0.00
8	2.3456–3.5209	944		78.4	796.53	47.08	804.43	47.07	60.92	3.99
8a		933	>75		801.64	47.18	810.21	47.22	61.06	3.97
8b		11	>50–75		367.34	38.79	314.56	34.22	49.28	5.09
8c		0	>15–50		0.00	0.00	0.00	0.00	0.00	0.00
8d		0	≥0.0–15		0.00	0.00	0.00	0.00	0.00	0.00

[1)] Percentages were calculated based on total PSU and cover.

grown in higher concentrations (strata 1a and 2a). Soybeans and corn are planted in equivalent quantities, in strata 4a, 5a, 6a. Furthermore, additional corn and wheat planting land cover statistics calculated at the PSU level and summed at the stratum level, as shown in Table 5, provide additional information regarding the corn and soybean rotation pattern and the distribution of wheat planting in these strata.

The single crop planting frequency stratifications are developed for individual crop surveys that are of great interest to commodity trading groups or researchers evaluating individual crops. However, the substratification of the crop planting frequency strata, by percent cultivation, can provide spatial distribution information of other crops. Therefore, the crop specific stratification results can also be used for multivariate sampling for other crops with appropriate sample allocation. A set of weights can be assigned for different strata. Of course, the appropriate sample allocation schema has to be developed.

3.2. Multi-crop specific stratifications

A multi crop specific (corn, soybeans, and wheat) stratification is also created by running the multivariate K-means clustering algorithm based on the corn, soybean and wheat frequency mean values together calculated at the PSU level. Table 6 illustrates the resulting stratum definitions (including no. of PSUs per stratum) for the eight strata in the multi-crop stratification, which include a combination of characteristics including soybean planting frequency means, corn planting frequency means and wheat planting frequency means calculated at the PSU level. Further, soybean, corn and wheat acreage, reported in the SD FSA CLU data, are calculated at the stratum level to evaluate the predictability of the multiple crop specific stratification.

Fig. 3-D illustrates the spatial distribution of the multi-crop stratification. Table 6 includes the eight multi-crop strata definitions, which include the stratification variables and mean stratum values for each variable, as well as mean corn, soybeans and wheat acres reported in the FSA CLU data in 2014 for each stratum. General observations of the multi-crop stratification are that stratum 1 (light tan) in Fig. 3-D would be expected, based on crop planting frequency mean values, listed in Table 6, to include the least amount of crop acreage. Stratum 2 illustrated in the tan color in Fig. 3 would be expected to include moderately low levels of crop acreage for all three crops (Table 6). Stratum 3 (dark red/brown) would have the largest quantities of wheat planting based on the high wheat frequency mean values. Stratum 4 (lighter red/brown) would be anticipated to include large amounts of wheat planting but not as much as stratum 3. From stratum 5 to 8 (blues to dark green) progressively larger amounts of corn and soybeans are anticipated to be reported in these strata based on the corresponding higher corn and soybean planting frequency ranges in Table 6. Stratum 8 (darkest green) would be expected to have the largest amounts of corn and soybean acreage based on the high soybean and corn planting frequency mean values. The FSA CLU data, reported in 2014 and listed in Table 6, confirm these general predications.

Fig. 4-A illustrates the current NASS SD traditional ASF based on percent cultivation. Fig. 4-B illustrates the new multi-crop stratification based on the Crop Planting Frequency Layers. The spatial distribution of crop planting is consistent between these two stratifications. However, the multi-crop specific stratification provides more targeted crop specific distribution information with the homogeneous strata

Table 6 The eight multi-crop specific strata definitions, mean crop frequency values, and mean soybean, corn and wheat acres reported in 2014

	Variables	Mean	Mean 2014 soybean acres	Mean 2014 corn acres	Mean 2014 wheat acres
Stratum 1 (PSU no. 4254)	Soybean frequency	0.0649	2.3255		
	Corn frequency	0.1244		78.3554	
	Wheat frequency	0.0954			57.8367
Stratum 2 (PSU no. 1675)	Soybean frequency	0.1239	78.8352		
	Corn frequency	0.3902		174.9277	
	Wheat frequency	0.8391			286.2222
Stratum 3 (PSU no. 500)	Soybean frequency	0.0882	60.7055		
	Corn frequency	0.4988		235.8626	
	Wheat frequency	2.9082			831.4245
Stratum 4 (PSU no. 1224)	Soybean frequency	0.1766	100.7665		
	Corn frequency	0.7117		241.3915	
	Wheat frequency	1.8089			498.2871
Stratum 5 (PSU no. 2807)	Soybean frequency	0.7160	271.4267		
	Corn frequency	0.8967		285.8428	
	Wheat frequency	0.2474			69.1390
Stratum 6 (PSU no. 1458)	Soybean frequency	1.4895	512.0829		
	Corn frequency	1.3613		447.7145	
	Wheat frequency	1.0274			224.2992
Stratum 7 (PSU no. 3008)	Soybean frequency	1.4444	450.6398		
	Corn frequency	1.6290		460.4094	
	Wheat frequency	0.2083			50.7613
Stratum 8 (PSU no. 3225)	Soybean frequency	2.1224	749.7595		
	Corn frequency	2.3779		782.8406	
	Wheat frequency	0.1549			41.5645
Strata 1–8 (PSU no. 18151)	Soybean frequency	0.8878	316.9351		
	Corn frequency	1.0674		352.8414	
	Wheat frequency	0.4847			140.9791

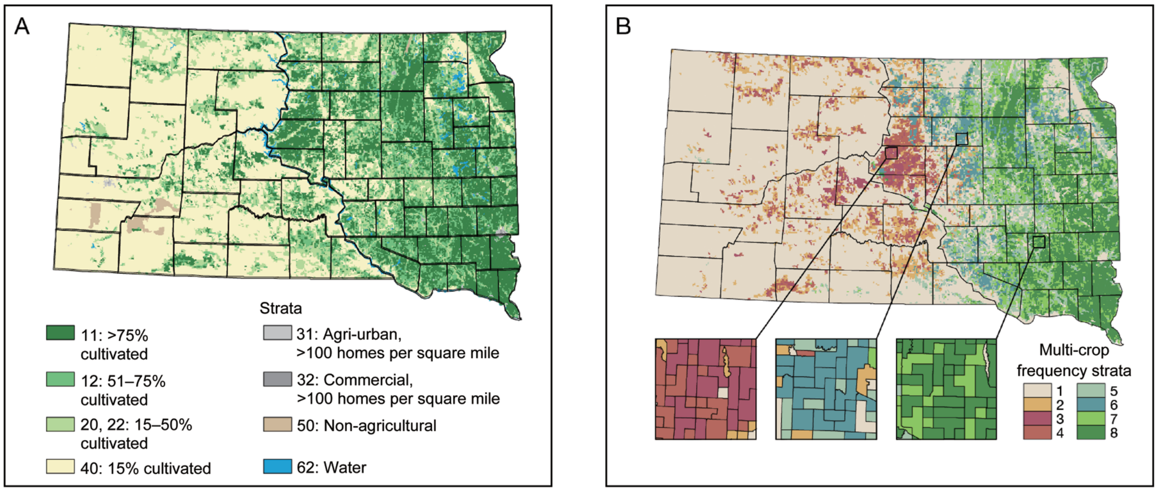

Fig. 4 NASS South Dakota area frame stratified based on percent cultivation alone is illustrated in A with the highest percent cultivation illustrated in dark green. B illustrates the South Dakota multi-crop stratification created as part of this research using Corn, Soybean and Wheat Planting Frequency Layers.

based on crop specific planting frequency statistics while the traditional ASF has no crop specific planting information. As shown in Fig. 4-B, the corn, soybeans and wheat planting areas are clearly identified within strata.

Crop specific ASF stratifications, such as those presented in this paper, can be useful in targeting individual crops for sampling and estimation. Future research will include continued assessments of the multi-crop stratification that can be used to select samples and conduct estimation using the 2014 FSA CLU and 578 administrative data as proxy survey data which can be compared with the 2014 JAS estimates.

4. Conclusion

This paper proposed using a novel method for land cover ASF stratification based on crop specific planting frequency and percent cultivation to build four new crop specific ASF stratification designs. In this paper, new SD corn, soybean and wheat stratifications were derived based on crop planting frequency and cultivation. Three single crop specific stratifications and a multi-crop specific stratification produced strata where reliably predicted quantities of target crops were reported grown in 2014. The four novel ASF stratifications provided additional crop specific information when compared to traditional stratification based on percent cultivation alone. The additional mean corn, soybean and wheat frequency data provided strong indications about where these crops are planted in large quantities, which was confirmed using the 2014 FSA CLU and administrative data 578 as *in situ* validation. In SD, soybeans and corn are grown in similar areas and rotated. Consequently, strata 3–8 in the soybean frequency/percent cultivation stratification identifies the areas where equivalent quantities of soybeans and corn are grown. It is observed that all four ASF stratifications predict corn, soybean and wheat planting patterns well as verified by the 2014 FSA and 578 administrative data. This demonstrates that the new stratifications based on geospatial crop planting frequency (2008–2013) and cultivation are crop type independent and applicable to all major crops in addition to the specified crops. These results indicate that new SD crop specific stratifications based on the Corn, Soybean and Wheat Planting Frequency (2008–2013) and Cultivation Data Layers have great potential to improve ASF accuracy, efficiency and crop estimates.

Acknowledgements

The authors would like to thank Ms. Lee Ebinger from the United States Department of Agriculture for creating the figures for this paper and Dr. Shirley Sampson from the United States Department of Agriculture for editorial review.

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 Section on Survey Research Methods JSM 2010*. American Statistical Association, Alexandria, VA.

Benedetti R, Postiglione P, Piersimoni F. 2015. *Sampling Spatial Units for Agricultural Surveys. Advances in Spatial Science*. Springer, Berlin.

Bethel J. 1989. Sample allocation in multivariate surveys. *Survey Methodology*, **15**, 47–57.

Boryan C, Yang Z. 2014. Operational implementation of a new automatic stratification method using geospatial cropland data layers in the NASS area frame section. In: *Proceedings of IGARSS 2014*, *IGARSS 2014 & 35th Canadian Symposium on Remote Sensing*. Quebec City, Canada.

Boryan C, Yang Z, Di L. 2012. Deriving 2011 cultivated land cover data sets using USDA national agricultural statistics service historic Cropland Data Layers, In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*. July 22–27, 2012. Munich, Germany.

Boryan C, Yang Z, Di L, Hunt K. 2014a. 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*, **7**, 4317–4327.

Boryan C, 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*, **5**, 341–358.

Boryan C G, Yang Z, Willis P. 2014b. US geospatial crop frequency data layers. In: *Proceedings of the Third International Conference on Agro-Geoinformatics (Agro-Geoinformatics 2014)*, August 11–14, 2014. IEEE, Beijing, China.

Boryan C G, Yang Z, Willis P. 2015. A novel method for area frame stratification based on geospatial crop planting frequency data layers. In: *Proceedings of IEEE*, *International Geoscience and Remote Sensing Symposium*. July 26–31, 2015. IEEE, Milan, Italy.

Carfagna E, Gallego F J. 2005. Using remote sensing for agricultural statistics. *International Statistical Review*, **73**, 389–404.

Census of Agriculture. 2012. [2016-1-23]. http://www.agcensus.usda.gov/

Cotter J, Davies C, Nealon J, Roberts, R. 2010. Area frame design for agricultural surveys. In: Benedetti R, Bee M, Espa G, Piersimoni F, eds., *Agricultural Survey Methods*. John Wiley & Sons, Chichester, UK.

Cotter J, Tomczac C M. 1994. An image analysis system to develop area sampling frames for agricultural surveys, *Photogrammetric Engineering and Remote Sensing*, **60**, 299–306.

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

Ford B I, Nealon J, Tortora R D. 1986. Area frame estimators in agricultural surveys — Sampling versus nonsampling errors. *Agricultural Economics Research*, **38**, 1–10.

Forgy E. 1965. Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, **21**, 768–769.

FSA (Farm Service Agency). 2014. Farm service agency (FSA) common land unit (CLU) information worksheet. [2015-9-10]. http://www.fsa.usda.gov/Internet/FSA_File/clu__infosheet_2013.pdf

Gallego F J, Feunette I, Cargagna E. 1994. Two stage area frame sampling on square segments for farm surveys. *Survey Methodology*, **20**, 107–115.

Gallego J, Delincè J. 2010. The European land use and cover area-frame statistical survey. In: Benedetti R, Bee M, Espa G, Piersimoni F, eds., *Agricultural Survey Methods*, John Wiley & Sons, Chichester, UK. pp. 151–168.

Han H, 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. *Computer and Electronics in Agriculture*, **84**, 111–123.

Hanuschak G A, Morrissey K M. 1978. *Pilot Study of the Potential Contributions for Landsat Data in the Construction of Area Sampling Frames, Statistical Reporting Service Staff Report 1977*. Washington, D.C.

Hartigan J A, Manchek A W. 1979. Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society* (Series C: Applied Statistics), **28**, 100–108.

Heard J. 2002. USDA establishes a common land unit. [2015-9-24]. http://www.esri.com/news/arcuser/0402/usda.html

MacQueen J B. 1967. Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1*. University of California Press, USA. pp. 281–297.

Nusser S M, House C C. 2009. Sampling, data collection, and estimation in agricultural surveys. In: Pfeffermann D, Rao C R, eds., *Handbook of Statistics 29A, Sample Surveys*: *Design*, *Methods and Applications*. Elsevier, The Netherlands. pp. 471–486.

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

Tsiligirides T. 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**, 45–77.

USDA, NASS. 2016. Cropland data layer, cultivation layer and crop frequency data layer "Metadata". [2016-1-13]. http://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php

Villalobos A G, Wallace M A. 1998. *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. p. 274.

Vogel F A. 1995. The evolution and development of agricultural statistics at the United States Department of Agriculture. *Journal of Official Statistics*, **11**, 161–180.

Workneh F, Tylka G, Yang X, Faghihi J, Ferris J. 1999. Regional assessment of soybean brown stem rot, phytophthora sojae, and heterodera glycines using area-frame sampling: Prevalence and effects of tillage. *Phytopathological*, **89**, 204–211.

(Managing editor SUN Lu-juan)