US GEOSPATIAL CROP FREQUENCY DATA LAYERS

Claire G. Boryan, Zhengwei Yang and Patrick Willis
Research and Development Division
USDA National Agricultural Statistics Service
Fairfax, USA
Claire.Boryan@nass.usda.gov

***Abstract*—Information on future crop specific planting is valuable for improving agricultural survey estimates. This information is critical for agricultural production planning, agricultural product commodity inventory control, natural resource allocation and conservation, etc. However, future crop planting details are generally unavailable. This paper proposes to use crop specific planting frequency data as indicators to indirectly provide information regarding future crop planting. A methodology to derive crop planting frequency data layers based on 2008-2013 Cropland Data Layers has been presented in this paper. Crop frequency layers for corn, soybeans, wheat and cotton were successfully built at the national level and for two states including Indiana and Mississippi, USA. Multi-year (2008-2013) Farm Service Agency (FSA) Common Land Unit (CLU) data were utilized to assess the accuracy of the derived crop frequency data layers. The accuracies of the national scale crop frequency data layers are 91.00%, 90.13%, 87.67% and 85.96% for corn, cotton, soybean and wheat respectively.**

Keywords—Cropland Data Layer; crop frequency map; automated stratification; crop specific stratification; multi-year land cover data;

I. INTRODUCTION

In agricultural survey sampling, information on future crop specific planting is valuable for improving survey estimates. This information is critical for agricultural production planning, agricultural product commodity inventory control, natural resource allocation and conservation, etc. However, future crop planting details are generally unavailable. This paper proposes to use crop specific planting frequency data as indicators to indirectly provide information regarding future crop planting. In the past, cropping "intensity", which was another name for crop frequency, and related cropping patterns were studied using satellite data. Cropping patterns were examined using MODIS 250m time series Normalized Difference Vegetation Index (NDVI) data over the Central Great Plains [1]. Biradar and Xiao identified and mapped cropping intensity in India using 500 m MODIS eight-day composite data. A phenology algorithm was developed based on NDVI, the Enhanced Vegetation Index (EVI) and the Land Surface Water Index (LSWI) to create crop intensity geospatial data sets [2]. Jain *et al*., compared different methods to identify and map cropping intensity of small holder farms in India [3]. Brown *et al*. used MODIS 250m EVI data to detect single and double cropping practices in Amazonia [4]. Boryan *et al*. used the USDA National Agricultural Statistics Service (NASS) Cropland Data Layers (CDLs), satellite data derivative products, to evaluate single crop planting "intensities" and crop rotation patterns from 2003 – 2007, focusing on continuous corn cropping systems, in Nebraska, Iowa and Illinois [5]. However, none of these works addressed national scale crop planting frequency for generic use. This paper describes a methodology for using 2008-2013 CDLs to derive crop planting frequency data layers. Crop planting frequency or crop frequency, in this case, is defined as the number of years that a pixel was identified to a specific crop over a given period. In this paper, corn, soybeans, wheat and cotton frequency layers for the states of Indiana and Mississippi and for the conterminous U.S. were built using NASS 2008-2013 CDL products as the layer inputs. The crop specific frequency layers have wide spread use both within NASS and to the public. Within USDA NASS, they can be used for area frame construction to create crop specific strata and to impute for missing survey data. In addition, they can be used for environmental and healthcare research, and agriculture production planning.

II. DATA AND METHODOLOGY

A. Data and Study Scope

In this paper, Indiana and Mississippi, U.S. were selected as study areas due to their crop varieties. However, the final products are the crop frequency data layers covering the conterminous U.S. Corn, soybeans, wheat and cotton frequency layers were built for the two states and for the conterminous U.S. The crop frequency data layers were derived from NASS 2008-2013 CDL products. The Cropland Data Layers (CDLs) are crop specific land cover classifications, which have been produced by the USDA NASS for all 48 conterminous states since 2008 [6]. The NASS CDLs are made available to the public on the NASS web GIS application CropScape [7]. CDLs are used within NASS to produce major crop acreage estimates in agriculturally intensive states using a linear model which regresses CDL pixel data to NASS June Area Survey (JAS) segment summary data [8]. Single year and multiyear CDL data were used in NASS area sampling frame (ASF) construction to automatically and objectively stratify land

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:53:12 UTC from IEEE Xplore. Restrictions apply.

cover in the U.S. based on percent cultivation. The automated CDL based stratification significantly improved stratification accuracies statistically at a 95 percent confidence level in intensively cropped areas as compared with the visual interpretation based traditional stratification method [9] [10], [11]. The proposed automated stratification methodology has been implemented into USDA NASS Area Frame operational procedures. Improved results were achieved in the construction of the state level area frames. Therefore, the multiyear CDL data were used to derive the crop frequency data layers with the goal of achieving further stratification and survey estimate improvement.

The crop frequency data layer's validation was conducted using Farm Service Agency (FSA) Common Land Unit (CLU data) from which the validation data were sampled. The sampling population included original 2008-2013 CDLs' training and validation data derived from the CLU data. This crop frequency layer validation data set included only pixels identified as a specific crop type over all six years (2008-2013). This CLU validation subset data for crop frequency were recoded in the same manner as the recoded CDL data.

TABLE I. TABLE CROPSPECIFIC LAYERS

Corn	
CDL Code	**Crop Description**
1	Corn
225	Dbl Crop WinWht/Corn
226	Dbl Crop Oats/Corn
237	Dbl Crop Barley/Corn
241	Dbl Crop Corn/Soybeans
Cotton	
2	Cotton
232	Dbl Crop Lettuce/Cotton
238	Dbl Crop WinWht/Cotton
239	Dbl Crop Soybeans/Cotton
Soybean	
5	Soybeans
26	Dbl Crop WinWht/Soybeans
239	Dbl Crop Soybeans/Cotton
240	Dbl Crop Soybeans/Oats
241	Dbl Crop Corn/Soybeans
254	Dbl Crop Barley/Soybean
Wheat	
22	Durum Wheat
23	Spring Wheat
24	Winter Wheat
26	Dbl Crop WinWht/Soybeans
225	Dbl Crop WinWht/Corn
230	Dbl Crop Lettuce/Durum Wht
234	Dbl Crop Durum Wht/Sorghum
236	Dbl Crop WinWht/Sorghum
238	Dbl Crop WinWht/Cotton

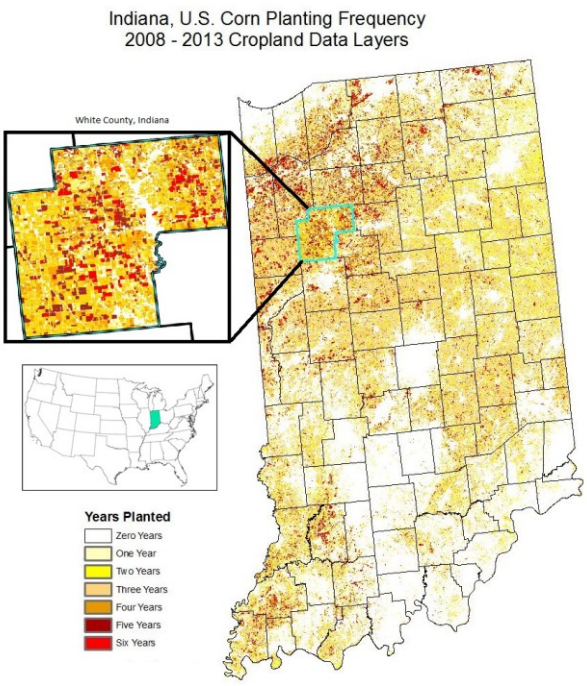

Fig. 1. Indiana corn planting frequency data layer

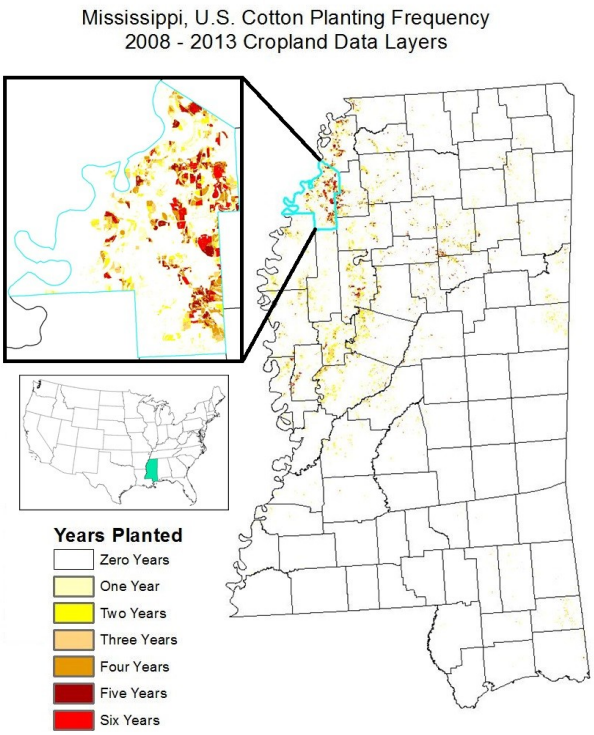

Fig. 2. Mississippi cotton planting frequency data layer

B. *Methodology*

CDL data products are published in 30m and 56m resolutions and are readily available for use. The method to produce the crop frequency data layers is straightforward. First, to make CDL data, produced in different years, consistent, all 56m resolution CDL data were resampled to 30m using nearest neighbor interpolation. Second, all of the original CDLs and validation data were recoded into the targeted crop categories. In this study, four crop categories: corn, soybean, wheat, and cotton were targeted. Therefore, six

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:53:12 UTC from IEEE Xplore. Restrictions apply.

years of CDLs were utilized to create the corn, soybean, wheat, and cotton layers by recoding the original CDLs. The recoding rules for the targeted crop categories are illustrated in Table 1 which shows how the CDL categories (with definitions) are grouped and recoded to create the crop specific layers. In recoding, all pixels identified as non-targeted crop categories in the original CDLs were recoded as background. After the six years of individual "crop specific" layers (one for each year) were generated, they were then used to count the number of times that each targeted crop type (only one at a time) was identified in the six years for every pixel to create single 2008-2013 corn, soybean, wheat and cotton frequency layers as outputs. The resulting output files have "0", "1". "2", "3", "4", "5", and "6" at the pixel level to indicate the number of years that the pixel was categorized to a specific crop from 2008-2013. Finally, the 2008-2013 FSA CLU data subset including only pixels selected over all six years, were utilized to assess the accuracy of the crop frequency layers. All crop categories were recoded into a set of new codes, as shown in Table 1, different from the original CDLs.

III. Results and Discussion

Four crop categories corn, soybean, wheat and cotton frequency data layers for the Indiana, Mississippi and at the national scale were derived based on the above mentioned methodology. As a demonstration, the Indiana corn planting frequency data layer and the Mississippi cotton planting frequency data layer were illustrated in Fig. 1 and Fig. 2. All crop frequency data layers were validated using 2008-2013 FSA CLU data as described in Section II. Accuracy assessment results for all crop specific layers, at all frequency levels (1, 2, 3, 4, 5, 6) and pixel distributions are summarized in Table 2. Both producer and user accuracies, as well as overall (1-6) accuracy, for all layers were tabulated. The percentage of pixels for all frequency levels are also listed with their corresponding accuracies.

The crop frequency layer pixel distributions, planting frequencies and how this distribution relates to crop rotation patterns and the layer accuracies for the states of Indiana, Mississippi and at the U.S. national level are discussed in the following subsections.

A. Indiana Crop Frequency Data Layer (2008-2013)

In Indiana, corn, soybean and winter wheat are major crops. As observed from Table 2, the most common (64.91%) corn planting frequency was three times planted to corn. This frequency had the highest producer (94.50%) and user (94.41%) accuracies which contribute to the high overall (90.66%) accuracy for the Indiana corn frequency layer (2008-2013). Correspondingly, it was observed that the highest percent (67.92%) of planting pixels in Indiana was also three times planted to soybeans. This frequency corresponded to the highest accuracy with 94.27% and 94.35% producer and user accuracies, and an overall accuracy of (90.46%) in Indiana

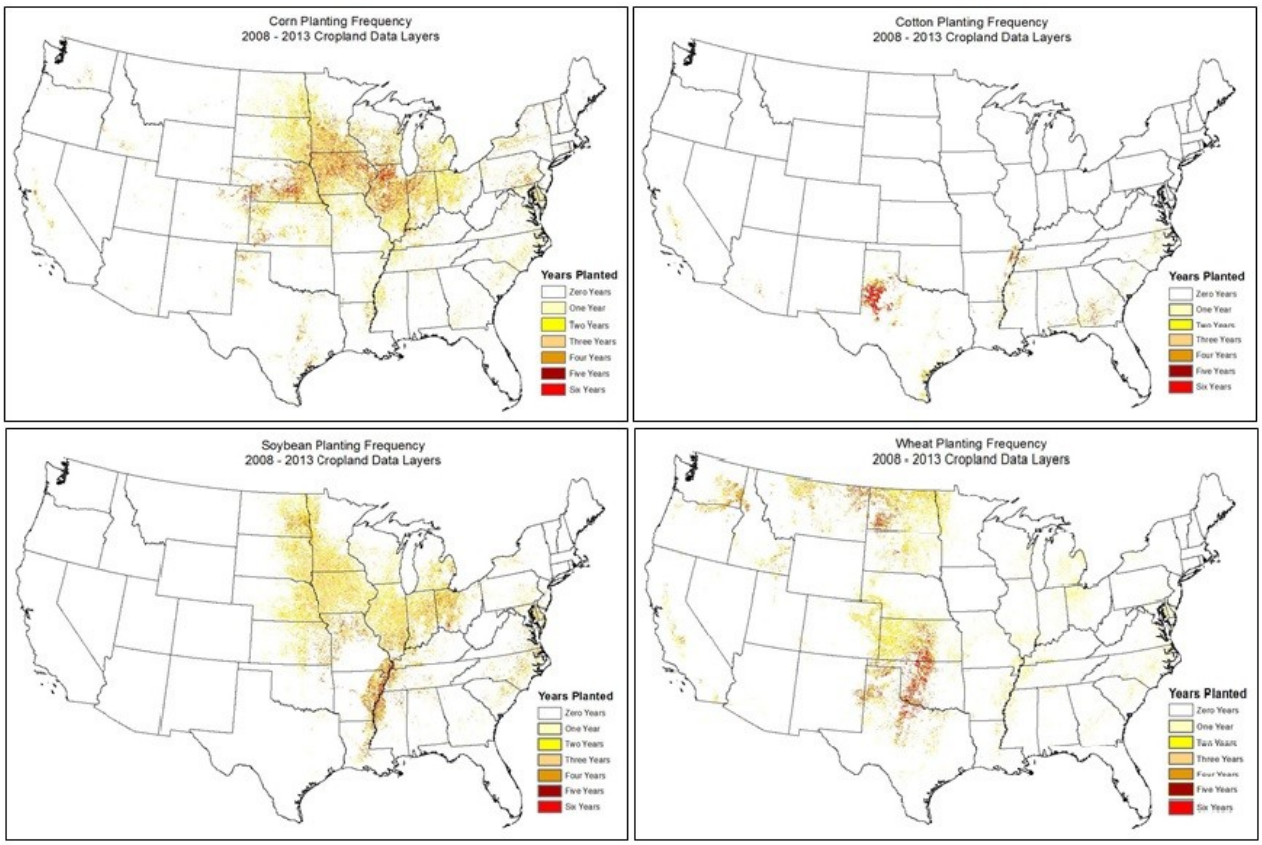

Fig. 3. National corn planting frequency data layers

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:53:12 UTC from IEEE Xplore. Restrictions apply.

TABLE II. INDIANA, MISSISSIPPI AND NATIONAL CROP FREQUENCY DATA LAYER ACCURACY RESULTS

INDIANA CROP FREQUENCY (2008-2013)																			
	Frequency - (1)			***Frequency- (2)***			***Frequency- (3)***			***Frequency- (4)***			***Frequency- (5)***			***Frequency- (6)***			
Crop	*Prod*	*User*	*% Pixels*	*Prod*	*User*	*% Pixels*	*Prod*	*User*	*% Pixels*	*Prod*	*User*	*% Pixels*	*Prod*	*User*	*% Pixels*	*Prod*	*User*	*% Pixels*	***Overall (1-6)***
Corn	76.97	79.44	***2.69***	76.48	80.60	***13.14***	94.50	94.41	***64.91***	87.89	83.66	***11.21***	89.15	86.39	***4.38***	93.98	94.19	***3.67***	90.66
Soy-beans	89.65	90.90	***5.48***	79.82	84.50	***16.09***	94.27	94.35	***67.92***	84.17	75.31	***8.35***	76.70	70.54	***1.49***	75.77	86.35	***0.67***	90.46
Winter Wheat	98.38	97.46	***67.30***	91.83	93.12	***24.50***	91.77	94.99	***8.20***	47.08	78.29	***0.00***	n/a	n/a	***0.00***	n/a	n/a	***0.00***	96.20
MISSISSIPPI CROP FREQUENCY																			
Corn	88.44	93.52	***29.26***	82.23	82.44	***29.0***	87.84	80.30	***27.08***	73.92	79.75	***10.50***	81.83	86.16	***3.04***	94.47	91.32	***1.12***	84.82
Soy-beans	85.06	90.59	***8.29***	74.63	76.83	***12.98***	81.28	76.30	***23.63***	76.85	76.81	***22.24***	79.73	74.08	***16.76***	80.65	91.78	***16.10***	79.39
Winter Wheat	93.60	94.99	***49.13***	86.53	81.21	***28.06***	81.79	85.43	***16.95***	83.37	88.74	***4.88***	86.03	92.83	***0.86***	65.00	95.79	***0.12***	89.02
Cotton	92.78	90.83	***33.80***	79.91	80.56	***24.53***	73.22	78.41	***19.45***	73.22	74.83	***10.91***	64.63	69.00	***5.40***	72.95	94.68	***5.91***	83.11
NATIONAL CROP FREQUENCY																			
Corn	92.42	91.23	***9.54***	81.09	84.21	***14.74***	93.67	85.91	***49.39***	90.25	85.91	***12.90***	89.13	87.98	***6.92***	94.54	94.04	***6.51***	91.00
Soy-bean	91.94	92.48	***11.73***	83.65	86.38	***20.49***	93.60	93.90	***55.71***	84.01	76.54	***7.83***	79.43	74.51	***2.50***	85.76	87.64	***1.74***	90.13
Winter Wheat	92.52	91.49	***17.0***	85.15	85.46	***21.57***	90.34	91.10	***35.08***	81.09	79.80	***10.08***	73.96	70.54	***5.49***	89.55	91.66	***10.78***	87.67
Cotton	93.82	90.32	***24.48***	82.42	82.42	***20.23***	81.05	84.75	***18.95***	79.87	77.59	***10.33***	78.88	76.44	***8.31***	91.22	94.97	***17.70***	85.97

*Prod –producer accuracy, User – user accuracy, % pixels – percentage of pixels in the layer attributed to a specific frequency.

TABLE III. NATIONAL CROP FREQUENCY DATA LAYER PRODUCER ACCURACY AND PERCENT OF PIXELS MISCLASSIFIED ONLY ONE TIME (ERROR)

	Frequency- (1)		Frequency- (2)		Frequency - (3)		Frequency- (4)		Frequency- (5)		Frequency- (6)	
Crop	Prod	Error	Prod	Error	Prod	Error	Prod	Error	Prod	Error	Prod	Error
Corn	92.42	7.02	81.09	18.39	93.67	5.99	90.25	9.06%	89.13	10.09	94.54	4.48
Soybeans	91.94	7.33	83.65	15.80	93.60	6.02	84.01	14.69	79.43	18.42	85.76	11.68
Winter Wheat	92.52	6.29	85.15	14.03	90.34	8.72	81.09	16.45	73.96	23.68	89.55	8.08
Cotton	93.82	5.62	82.46	16.94	81.05	16.98	79.87	17.36	78.88	19.19	91.22	7.42

soybean frequency layer. These frequency results directly related to the common corn/soybean rotation pattern in Indiana corn production. As shown in Table 2, repeated planting to winter wheat in Indiana is not common. The most common frequency for wheat plantings in Indiana is one time planted with 67.30% pixels planted to winter wheat. As observed for the period 2008-2013, pixels with two and three times planted to winter wheat are only 24.5% and 8.2% respectively. There are no pixels identified as being planted four, five or six times in the Indiana wheat frequency layer. The high producer (98.38%) and user (97.46%) accuracies for a single time planted contribute to the high overall (96.20%) accuracy for the Indiana wheat frequency layer. Fig. 1 illustrated the Indiana corn planting frequency distribution. As shown in Fig. 1, corn was planted across most of Indiana though it was planted most frequently in northwestern and southwestern Indiana.

B. Mississippi Crop Frequency Data Layer (2008-2013)

In Mississippi, the major crops include corn, soybean, winter wheat and cotton. The major agricultural area is in the Mississippi delta area (northwestern Mississippi). As shown in Fig. 2, the majority of cotton planting is also in this area. The cotton planting areas are scattered. Cotton is planted with a high degree of repetition in only a few counties in northwestern Mississippi. The crop rotation patterns in Mississippi differ from Indiana with only 27.08% of the corn and 23.63% of the soybeans in the corresponding Mississippi frequency layers identified as being planted three times to corn and soybeans. Cotton and wheat were identified as being planted most commonly a single time with 33.80% of the cotton crop and 49.13% of the wheat crop planted only one time over the six year time period.

C. National Crop Frequency Data Layer (2008-2013)

The Indiana and Mississippi crop frequency data layers demonstrated obvious local agricultural planting characteristics, which depended on their geographic conditions. One can expect significant differences among different states' crop frequency data layers as conjectured from these two state level crop frequency data layers. To meet general public and NASS' internal operational and research needs, national crop frequency data layer for four major crops were generated as shown in Fig 3. As a corn /soybean rotation is the most common in the majority of the large corn and soybean producing states in the Midwest U.S., the most common corn planting frequency (49.39%) and soybean planting (55.71%) was three times planted as shown in Table

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:53:12 UTC from IEEE Xplore. Restrictions apply.

2. Thirty five percent of all pixels in the national wheat crop frequency layer were identified as planted three times to wheat with corresponding producer (90.34%) and user (91.10%) accuracies. The overall accuracy for the national wheat frequency layer was 87.67%. Planting patterns for the U.S. cotton crop differ from other crops and vary regionally. Cotton was planted most often (24.48%) in only one of the six years. However, cotton is grown more continuously (17.70%) in Texas and in the delta region of Arkansas, where the crop is heavily irrigated. Other regionally distinct patterns exist which can be identified in the national crop frequency layers. These regional patterns include the common corn/soy rotation in the Midwest of the U.S., continuous planting to corn along the Platte River in Nebraska and in northern Illinois, continuous planting to varieties of wheat in south central Kansas, central Oklahoma, north central Texas, south west North Dakota and north west Montana.

D. ErrorDistribution

Table 3 illustrates the producer accuracies for the national crop frequency layers for all crops and frequency levels. To fully understand the quality of the crop specific data layers it is useful to look at how often the CDL pixels were misclassified. It was observed from the classification error matrices generated from validation that most misclassification errors occurred only one time to the crop of interest. For example, a pixel planted to corn three times most likely was mistakenly identified as being planted to corn either two or four times. It's much less likely to be identified as planted to corn either once or more than four times. This fact indicates the robustness of the classification. Table 3 illustrates the producer accuracy for each crop and frequency level with the percentage of pixels that were misclassified one time. For example, the national corn frequency layer had a producer accuracy for the level one frequency of (92.42%) with (7.02%) of the pixels misclassified one time. The percent of pixels misclassified more than one time was only 0.56%.

IV. CONCLUSION

This paper proposes to use crop specific planting frequency data as indicators to indirectly provide information regarding future crop planting. A methodology to derive crop specific planting frequency layers based on 2008-2013 CDL data has been presented in this paper. Crop frequency layers for four crops, including corn, soybeans, wheat and cotton, were successfully built at the national scale and for Indiana and Mississippi, USA. Multi-year (2008-2013) FSA CLU data were utilized to assess the accuracy of the corn, soybean, wheat and cotton frequency layers. The accuracies of the national scale crop frequency data layers are 91.00%, 90.13%, 87.67% and 85.96% for corn, cotton, soybean and wheat respectively. The accuracy assessment results indicate that the national crop frequency data layers are robust with most misclassifications occurring in only one year of the targeted crop specific layers.

Overall, the national crop frequency data layers provide a useful and robust data source for major crop production and inventory analysis and will have widespread use to NASS internal users and to the general public. One immediate use of the crop frequency data layers within NASS will be for Area Frame construction. Currently, the NASS Area Frame is stratified based on percentage of cultivation. These validated national crop frequency data sets, based on six years of CDL data, provide unique information on planting probabilities at the primary sampling unit (PSU) level. The national crop specific frequency data layers provide objective crop specific information on the possibility of future crop planting, which will facilitate area frame sub-stratification and potentially improve survey estimates. The use of these data sets to make improvements in Area Frame construction will be an area of continued research. The crop frequency data sets can also be used within NASS for imputation research and by the general public for environmental assessments, healthcare, bioenergy production research and food production decision support.

REFERENCES

[1] B. Wardlow and S. Egbert. "Discriminating cropping patterns for the U.S. central Great Plains region using time-series MODIS 250m NDVI data—preliminary results," In Integrated Remote Sensing at the Global, Regional and Local Scale, 10–15 November 2002, Denver, CO.

[2] C. Biradar and X. Xiao, "Quantifying the area and spatial distribution of double- and triple-cropping croplands in India with multi-temporal MODIS imagery in 2005," *International Journal of Remote Sensing*, 2011, 32:2, 367-386.

[3] M. Jain, P. Mondal, R. DeFries, C. Small and G. Galford, "Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors, " *Remote Sensing of Environment*, 2013, (134): 210-223.

[4] J. Brown, J. Kastens, B. Wardlow, W. Jepson, A. Coutinho, A. Venturieri, J. Lomas and J. Price, "Using MODIS to detect cropping frequency variation in mechanized agriculture in Amazonia," Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 2007, 21-26 abril 2007, INPE, p. 99-101.

[5] C. Boryan, M. Craig, and P. Willis, "An Evaluation of Single Crop Planting Intensity and Crop Rotation Patterns in Nebraska, Iowa and Illinois 2004–2008," In the ASPRS Conference, 2009, .Baltimore, MD, March (pp. 11-13).

[6] 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*, 2011, 26, (5): 341-358.

[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*, 2012, Vol. 84, pp. 111–123.

[8] C. Day, "A compilation of PEDITOR estimation formulas". NASS/RDD Research Paper RDD-02-03, USDA, NASS, Washington, D.C. January, 2002.

[9] 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.

[10] 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 International Geoscience and Remote Sensing Symposium*, July 22-27, 2012, Munich, Germany.

[11] C. Boryan, Z. Yang, "Deriving crop specific covariate data sets from multi-year NASS geospatial cropland data layers", *Proc. of the IEEE International Geoscience and Remote Sensing Symposium*, IGARSS 13, Melbourne, Australia, July 2013.

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:53:12 UTC from IEEE Xplore. Restrictions apply.