NASS Geospatial Applications from the Cropland Data Layer

Avery Sandborn, Rick Mueller, Claire Boryan, Dave Johnson, Zhengwei Yang, Lee Ebinger, Arthur Rosales, Patrick Willis, Robert Seffrin, Rachel Jennings, Matt Deaton, and Hubert Hamer*
United States Department of Agriculture/National Agricultural Statistics Service

Abstract:

For more than a decade, the United States Department of Agriculture's (USDA's) National Agricultural Statistics Service (NASS) has produced the Cropland Data Layer (CDL), a geospatial crop-specific land cover product covering the conterminous US. The CDL is increasingly being integrated into NASS's programs. An early application was the development of remotely sensed based crop acreage and yield estimates, which are independent of the survey estimates. Numerous derivative products have been created from the CDL. The Cultivated Layer and Crop Frequency Layer identify, respectively, where and how often a crop is planted. The June Area Survey (JAS) sample is drawn from the NASS area frame, and the stratification of the frame is now based on the CDL. Information for imputation of crop type and acreage for the JAS now comes from the CDL. Most recently, the CDL has been used as a primary input into disaster assessments. Now it is being considered as the foundation for integrating the diverse data sources available to NASS. This paper will discuss the creation and uses of the CDL and its derivative products and then focus on their potential future uses within the Agency.

Keywords:
CropScape; Cropland Data Layer; Cultivated Layer; June Area Survey; Disaster Assessments, Data Integration

1. Introduction:

The mission of the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is to provide timely, accurate, and useful statistics in service to U.S. agriculture. To help achieve NASS's mission, the Cropland Data Layer (CDL) utilizes remote sensing techniques to provide operational in-season acreage estimates to the NASS Agricultural Statistics Board and Regional Field Offices. The CDL is a 30-meter national raster, geo-referenced, crop-specific land cover classification product produced annually by NASS (Figure 1). The CDLs are published on the NASS CropScape web application (USDA/NASS Cropland Data Layer 2018). CropScape is designed to provide the public with open access to serve the CDL with interactive visualizations, data dissemination, geospatial queries, and online analytics (Han et al. 2012). Within NASS, CDL products have been used in a variety of research and operational applications, including masking crop extent for yield assessments (Johnson 2012), disaster assessments (Boryan et al. 2018), area frame stratification (Boryan and Yang 2017), improving estimates for the number of farms at the state and national-levels, and June Area Survey (JAS) imputation. Monitoring U.S. agriculture is important for food security and the CDL program provides a consistent geographical extent and spatial resolution over the past eleven years serving that purpose.

The original purpose of the CDL program was to generate acreage estimates of major commodities to reduce sampling error at the state, agricultural statistical district, and county-levels for internal NASS use by the Agricultural Statistics Board (Allen and Hanuschak 1988). The CDL is a supervised land-cover classification utilizing a decision tree machine learning approach using optical satellites while leveraging ground reference data collected from the USDA Farm Service Agency (FSA), as well as ancillary data from the U.S. Geological Survey (Boryan et al. 2011). Medium resolution satellites such as Landsat 8, Disaster Monitoring Constellation Deimos-1 and UK2, Resourcesat-2 LISS-III, and Sentinel-2 are used to collect imagery throughout the growing season. The CDL leverages ground reference data and multiple image collections across the growing season to capture the varying crop phenologies and derive a crop-specific land cover classification of planted area. CDL uses and applications external to NASS have been identified (Mueller and Harris 2013), and best practices and recommendations on studies with the CDL dataset have been developed (Lark et al. 2017). This paper focuses on internally driven applications that leverage the CDL product for improvement of agricultural statistics and geospatial data products.

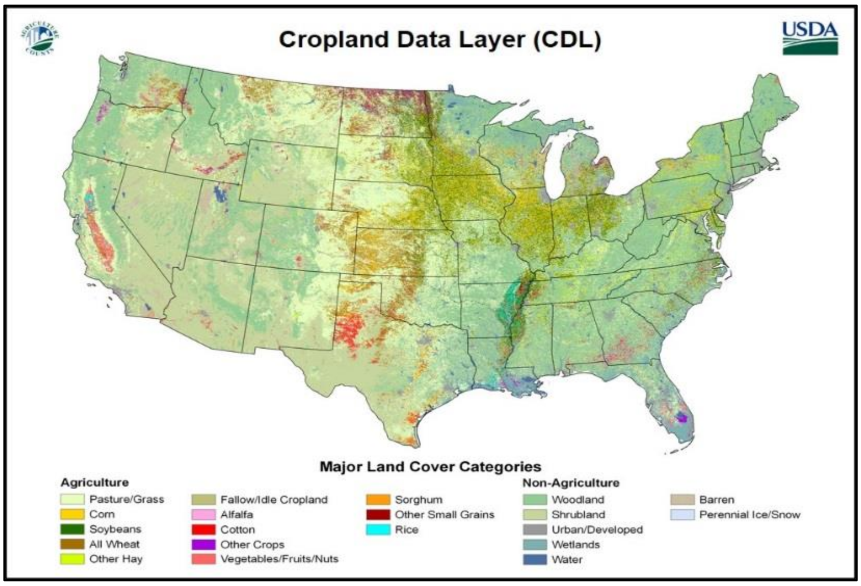

Figure 1: The Cropland Data Layer (CDL).

2. Current methods and products:

Frequency Layers

The U.S. national scale Crop Frequency Layers, which are derivative products of the CDL, identify planting frequency or intensity at the 30-meter pixel-level for corn, cotton, soybeans, and wheat. The four national-level crop frequency layers were built and validated with ground reference data from the FSA. These layers provide indicators for future crop planting, which is valuable for improving agricultural survey estimates, agricultural production planning, water resource management, natural resource allocation, and conservation (Boryan et al., 2014a). These layers are available for visualization, analysis, and download from CropScape.

Cultivated Layer

Another important derivative product of the CDL is the Cultivated Layer (Figure 2). The Cultivated Layer is a highly accurate characterization of cultivated land across the continental U.S. Unlike the original CDLs, which include more than 100 different crop categories, the Cultivated Layer includes only two categories: cultivation and non-cultivation. The cultivated land cover classes include tilled/planted crops and does not include hay, other hay, pasture, or rangeland. For operational purposes, five years of CDLs are combined to create the national-scale Cultivated Layer, which helps to reduce errors in identifying cultivated land pixels (Boryan et al., 2012). The Cultivated Layer is defined as any pixel identified as cultivated two out of the last five years or identified as cultivated in the most recent year. The 2018 Cultivated Layer is produced using 2014-2018 CDLs and validated using 2014-2018 FSA data.

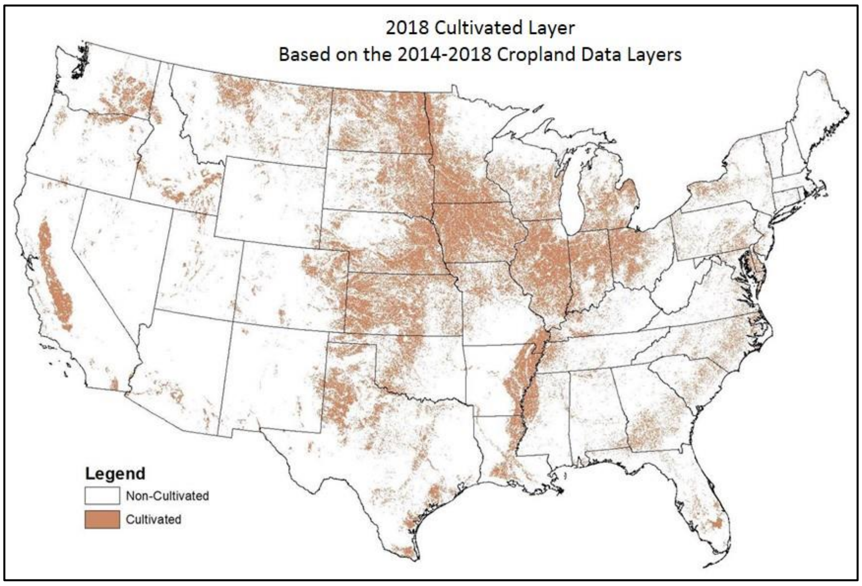

Figure 2: Cultivated Layer

Crop Yield Modelling

The CDL is also a foundational element for remotely-sensed yield estimation within NASS. Operational efforts to date have focused on corn and soybean yields within the Corn Belt at county and state-levels (Johnson, 2014). Research efforts are ongoing for other crops and areas of the U.S. (Johnson, 2016). Because crop progress, condition, and yield are dynamic, high revisit rate optical and thermal satellite imagery is required. Weekly observations are minimally sought and the best source for that type of information has been from the Moderate Resolution Imaging Spectroradiometer (MODIS), which has been in existence for nearly two decades. MODIS, while temporally ideal, is compromised spatially as it is only 250 meters in resolution, and is a challenge to identify with sufficient map precision field crop types and their boundaries. Thus, the 30 meter spatial resolution CDL fills this need.

Having highly accurate field-level crop type information ultimately allows one to isolate or mask the MODIS observations to only crop specific areas. This provides a clean signal of the vegetation profile throughout the growing season and improves yield model performance. Because the CDL is generated within season it can be leveraged as early as the August Crop Production report. More generally, the CDL can also be used to create year agnostic crop type or area masks (Johnson, 2012) by integrating the data over several seasons. This is useful for scenarios, such as needing to mask MODIS data that exists prior to the 2008 availability of national-level CDLs or when needing a crop predictive layer within season even before the CDL is available. This CDL-derived crop information is also useful for simplified yield modelling circumstances where managing year-specific crop maps is unwieldy.

Area Frame Stratification

A new automatic area frame stratification method (Boryan et al., 2014b; Boryan and Yang, 2017) was recently developed and implemented for NASS operations based on the Cultivated Layer. The NASS state-level area frames are stratified based on percent cultivated cropland within NASS Primary Sampling Units (Figure 3) and are used to select samples for NASS's annual JAS. For more than fifty years, the traditional area frame stratification method was conducted using visual interpretation of aerial photography or satellite data (Cotter et al., 2010). Research findings show that using the automated stratification method, based on the CDL, significantly improves area sampling frame stratification accuracies in intensively cropped areas (>75% cultivation) and overall stratification accuracies when compared to traditional stratification based on visual interpretation of aerial photography or satellite data, while reducing the cost of area frame construction (Boryan et al., 2014b).

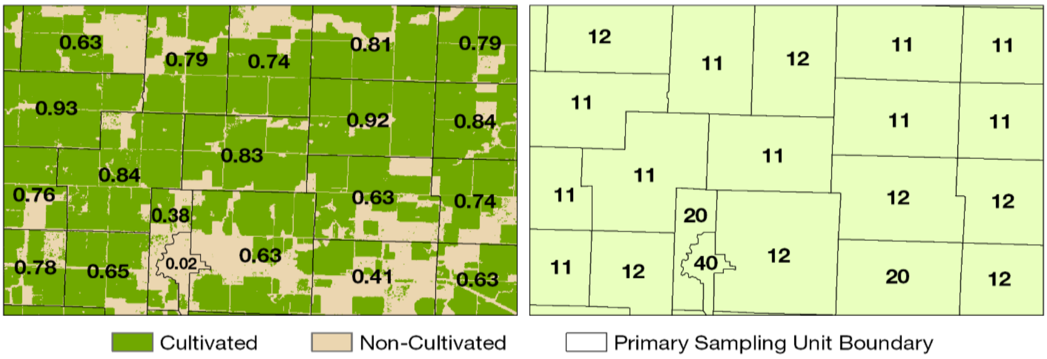

Figure 3: (left) CDL Cultivated Layer based Primary Sampling Unit with percent cultivation; (right) the same area sampling frame with CDL derived strata, where strata 11 represents greater than 75% cultivated, strata 12 represents 51-75% cultivated, strata 20 represents 15-50% cultivated, and 40 represents less than 15% cultivated.

Though the new fully-automated stratification method has improved stratification efficiency, objectivity, and accuracy in the intensively cropped areas, it achieves lower accuracies in low or non-agricultural areas. Consequently, a hybrid approach that integrates the automated stratification results with manual editing/review methods was implemented operationally. Since 2014, ten state-level area frames were built using the new integrated operational process. Boryan and Yang (2017) describe the area frame improvements, which include improved accuracy (30% improvement) and a reduction in labor costs, as well as other measurement criteria. As an example, the traditional Oklahoma area frame was constructed in 4,552 employee hours, while the hybrid frame required 1,980 employee hours.

<u>June Area Survey Imputation</u>

The JAS is the largest NASS annual survey in which approximately 9,000 square mile sample segments are visited by enumerators at the beginning of June to collect crop type and acreage information. Estimates of crop acreage and livestock inventories are based on these survey data (USDA/NASS, Understanding Statistics 2018). Although the JAS sampling frame provides complete coverage of all agriculture activity occurring on land pertaining to the target population, operator non-response and inaccessibility leads to the need to impute missing information. JAS imputation was traditionally based on historical records, which are not always available for the selected samples.

The CDL data provide a reliable alternative data source for imputation. Now NASS statisticians import the digital JAS segment boundary files directly into CropScape for analysis and use CDL data to 1) identify field-level planting history and crop rotation cycles, 2) confirm the area of irregularly shaped fields, and 3) review and resolve conflicts in reported data on segment-specific portions of the JAS questionnaire. Additionally, the statisticians can view other information layers in CropScape, such as major road networks and the Crop Frequency Layers for major commodities to guide the manual imputation process.

<u>Number of Farms</u>

There are regions within the U.S. where pre-screening of small farming operations is time-consuming, expensive, and subject to misclassification. A procedure was developed to use the CDL to obtain land cover statistics within the JAS segments to more accurately identify potential land use, prior to survey pre-screening activities. The specific goal was to improve the official estimate for number of farms at the state and national-level for both the JAS and quinquennial US Census of Agriculture. The most recent five years of the CDL are used to categorize the land in each JAS segment and estimate percentages of several predetermined categories, such as percent cultivation, impervious surface, corn, soybean and pasture calculated at the JAS segment level. These percentages are used as covariates when modelling the probability that a record represents a farm and ultimately adjusting the weights to obtain the estimate of the number of farms for the JAS and Census of Agriculture.

Natural Disasters

NASS utilizes the CDL and Cultivated Layer to monitor and assess affected cropland and livestock in the U.S. caused by hurricanes, regional flooding, and fire events. This capability is now possible due to a refined methodology utilizing freely available geospatial data products, which include the newly launched and freely available Synthetic Aperture Radar (SAR) Sentinel-1 data, optical satellite imagery, and supplemental geospatial hurricane and fire location data. During disaster response, the confidential in-season CDL data are used to provide timely crop acreage estimates of impacted land to NASS stakeholders. The non-confidential (previous year) CDL data are used to provide crop acreage estimates that can be released within USDA and to the public on the NASS Disaster Analysis website (USDA/NASS Disaster Analysis 2018).

Specifically for hurricanes and heavy-precipitation events, NASS utilizes freely available Copernicus Sentinel-1 SAR data to conduct operational flood mapping of agricultural land in near real-time (Copernicus, 2018). NASS produces multiple binary inundation raster products that are then overlaid with agricultural information from the CDL and Cultivated Layer. From this analysis, NASS is able to estimate the extent of cropland, pasture/hay, and specific crop types that are inundated. This operational flood monitoring process provides accurate results based on independent manually-derived ground reference data (above 95% producer's accuracy) (Boryan et al., 2018), and has been operationally used for Hurricanes Harvey and Irma in 2017 and Hurricanes Florence and Michael in 2018. During fire events, daily optical imagery, such as MODIS data, and active fire location geospatial data, such as Cal Fire (Cal Fire, 2019) and the USDA Forest Service Remote Sensing Applications Center (USDA/Forest Service Remote Sensing Applications Center, 2019), are combined with the CDL and Cultivated Layer to identify agricultural areas within an active fire perimeter. This methodology was implemented for agricultural areas impacted by the 2017 northern California wildfires 2017 and a 2018 Oregon substation fire.

3. Current and Future Efforts:

The CDL and its derivatives have been used to inform NASS processes as in the yield modelling and the JAS imputation. NASS is now moving to integrate its survey and administrative data with the remotely sensed information. Two examples will be discussed here. First, as discussed above, official estimates of the acreages planted to various crops are published each year. Currently, initial estimates are obtained from survey data, FSA data, and the CDL. These estimates are combined using expert opinion during the Agricultural Statistics Board process. However, the information from all of these sources can potentially be geospatially integrated to provide an improved estimate. Efforts are now underway to evaluate this approach for the state of Iowa with the goal of providing estimates for all states.

Currently, the responses to the Census of Agriculture questionnaire provide the foundation for all official Census estimates. For the 2022 and subsequent censuses, NASS is moving rapidly to make full use of the remotely sensed information and administrative data. In addition to using current administrative data to reduce respondent burden, remotely sensed data will be used to inform estimates. Using the CDL with the FSA administrative data as the foundation with the questionnaire being used to fill in gaps in information is being explored for possible implementation for the 2027 Census.

4. Discussion and Conclusion:

The numbers of uses and derivative products of the CDL have increased over its 11-year history. Most recently NASS has begun developing maps displaying the extent of the impact of natural disasters, such as floods and fires, as well as quantifying the extent of the potential damage by crop. The resulting products have been used, within NASS, within the US Department of Agriculture, and more broadly to inform relief efforts. Yet, other opportunities for using the CDL certainly exist, especially as NASS moves to integrate the CDL information with NASS survey and administrative data. By combining these different sources, the potential exists to both reduce respondent burden and to produce more precise official estimates. A longer-range goal is to explore the potential of having remotely sensed data as the primary foundation for analysis of Census of Agriculture data with the Census questionnaire and administrative data providing ground-truthing and supplemental data. NASS continues to look toward the future for remote sensing and geospatial technologies that can enhance agricultural statistics and be integrated into Agency operations.

References:

1. Allen, J. D. and Hanuschak, G. (1988). The remote Sensing Applications Program of the National Agricultural Statistics Service: 1980-1987. U.S. Department of Agriculture, NASS Staff Report No. SRB-88-08. https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/GIS_Reports/The Remote Sensing Applications Program of the National Agricultural Statistics Service 1980-1987.pdf.

2. Boryan C. and Yang, Z. (2017). Integration of the Cropland Data Layer Based Automatic Stratification Method into the Traditional Area Frame Construction Process. Survey Research Methods, 11 (3), 289-306. https://ojs.ub.uni-konstanz.de/srm/article/view/6725.

3. Boryan C., Yang, Z., and Di. L. (2012). 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, Munich, Germany. https://ieeexplore.ieee.org/document/6352699.

4. Boryan, C., Yang, Z., Di, L., and Hunt, K. (2014b). 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, 9 (11), 4317-4327. https://ieeexplore.ieee.org/document/6837419.

5. Boryan, C., Yang, Z., Mueller, R., and Craig, M. (2011). Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service Cropland Data Layer Program. Geocarto International, 26 (5), 341-358. https://www.tandfonline.com/doi/abs/10.1080/10106049.2011.562309.

6. Boryan, C., Yang, Z., Sandborn, A., Willis, P., and Haack, B. (2018). Operational Agricultural Flood Monitoring with Sentinel-1 Synthetic Aperture Radar. Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), 2018 IEEE International, Valencia, Spain, July 22 – 27, 2018. https://ieeexplore.ieee.org/document/8519458.

7. Boryan, C., Yang, Z., and Willis, P. (2014a). US geospatial crop frequency data layers. Proceedings of the Third International Conference on Agro-geoinformatics, August 11-14 2014, Beijing, China. https://ieeexplore.ieee.org/document/6910657.

8. Cal Fire, (2019). http://www.fire.ca.gov/general/firemaps. Ca.gov.

9. Copernicus, (2018). Copernicus Services Data Hub. https://cophub.copernicus.eu/. European Space Agency.

10. Cotter, J. Davies, C., Nealon, J., and Roberts, R. (2010). Area Frame Design for Agricultural Surveys in Agricultural Survey Methods (eds R. Benedetti, M. Bee, G. Espa and F. Piersimoni), John Wiley & Sons, TD, Chichester, UK. https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470665480.ch11.

11. Han, W., Yang, Z., Di, L., and Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Computers and Electronics in Agriculture, 84, 111–123. https://doi.org/10.1016/j.compag.2012.03.005.

12. Johnson, D.M. (2012). A 2010 map estimate of annually tilled cropland within the conterminous United States, Agricultural Systems, 114, 95-105. https://doi.org/10.1016/j.agsy.2012.08.004.

13. Johnson, D.M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sensing of Environment, 99, 341–356. https://doi.org/10.1016/j.rse.2013.10.027.

14. Johnson, D.M. (2016). A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products, International Journal of Applied Earth Observation and Geoinformation, 52, 65–81. https://doi.org/10.1016/j.jag.2016.05.010.

15. Lark, T.J., Mueller, R.M., Johnson, D.M., and Gibbs, H.K. (2017). Measuring land-use and land-cover change using the U.S. Department of Agriculture's Cropland Data Layer: Cautions and recommendations, International Journal of Applied Earth Observation and Geoformation, 62, 224-235. https://doi.org/10.1016/j.jag.2017.06.007

16. Mueller, R and Harris, J. (2013). Reported Uses of CropScape and the National Cropland Data Layer Program. ICAS VI, Rio de Janeiro, Brazil. https://repository.uantwerpen.be/docman/irua/78a9f0/130375.pdf.

17. USDA/Forest Service Remote Sensing Applications Center, (2019). https://fsapps.nwcg.gov/afm/gisdata.php. USDA-Forest Service, Salt Lake City, UT.

18. USDA NASS Disaster Analysis, (2018). https://www.nass.usda.gov/Research_and_Science/Disaster-Analysis/index.php. USDA-NASS, Washington, DC.

19. USDA NASS Cropland Data Layer, (2018). https://nassgeodata.gmu.edu/CropScape/. USDA-NASS, Washington, DC.

20. USDA NASS, Understanding Statistics, (2018). https://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/Foundation_of_Estimates/Area_Frame_Samples/. USDA-NASS, Washington, DC.