The 2009 Cropland Data Layer

by David M. Johnson and Richard Mueller

Introduction

The National Agricultural Statistics Service (NASS) recently released a crop-specific land cover classification product encompassing the entire conterminous United States (U.S.). Termed the Cropland Data Layer (CDL), the product depicts type and location for crops planted during the summer 2009 growing season. This effort arises from a confluence of factors including low cost and free mid-resolution satellite-based imagery, access to high quality ground truth, efficient and robust classification software, increased computer computational speed and data storage, and years of research and development. Spatial resolution of the raster-based CDL is 56 meters utilizing an Albers equal-area conic projection. Pixel-level accuracy for crop type categories average around 80 percent but are usually much higher for dominant commodity cover-types like corn, soybeans, wheat, cotton, and rice. General non-agricultural categories, while not the focus, are also mapped. The data are available for free and can be integrated and analyzed within most geographic information systems (GIS).

Background

NASS initiated its research into using remotely sensed data in the 1970s and early 1980s primarily through participation in the big multi-agency projects known as the Large Area Crop Inventory Experiment (LACIE) and Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing (AgRISTARS). One goal of these programs was to determine if crop acreage estimates could be derived from multi-spectral imagery and ground truth data. These early projects were successful at generating land cover classifications which were used to derive unbiased statistical estimates of crop area at state and county levels. Most importantly for NASS, the cropland classifications, when intersected with survey indications through regression analysis, provided a robust method to reduce statistical variance from farmer reported surveys. This regression estimation methodology is usually superior to simple "pixel counting" which is often biased. A comprehensive summary of remote sensing use for collecting agricultural statistics is provided by Carfagna and Gallego (2005).

The remote sensing acreage estimation research program at NASS has since evolved paving the way for the current, more operational, CDL products. It has grown, especially over the last few years, in geographic scope from the analysis of only a few central states within the U.S. to the entire country. For the core areas of Iowa, Illinois, Indiana, North Dakota, and Arkansas there is a decade's worth of annual CDLs produced by NASS. The other Corn Belt and Mississippi River Delta region states have a time series dating back to 2006. The remaining lower 48 states have been added within the last year or two. Timeliness has also improved with data released to the public each January, versus March, after the current season. Furthermore, overall quality and accuracy of the CDLs has steadily increased over time due to better ground truth and greater access to imagery.

continued on page 1202

continued from page 1201

Imagery Sources

The primary imagery source for deriving the 2009 CDL is the Indian Remote Sensing satellite IRS-P6, also known as Resourcesat-1. It was launched in 2003 by the Indian Space Research Organization. The sensor utilized from the platform is the Advanced Wide Field Sensor (AWiFS). The design of AWiFS is particularly suited to monitoring mechanized large area agricultural production because of its relatively wide 737km swath, spectral reflectance channels in the visible, near-infrared and mid-infrared bands, and minimum 5-day revisit rate. The nadir pixel resolution is 56 meters (about 0.3 hectares or 0.7 acres) yet detailed enough to identify crop type information from field sizes typically found within the U.S. NASS has utilized AWiFS data for several years and finds it comparable to 30 meter Landsat-5 Thematic Mapper (TM) data for agricultural monitoring (Johnson, 2008). The resulting trade-off is that AWiFS has a coarser spatial resolution but increased temporal frequency versus TM (5-day versus 16-day). Because agricultural cover types tend to be very dynamic, revisit rate is usually preferred by NASS over spatial resolution.

For the 2009 CDL, Landsat 5 TM data were also used to supplement the AWiFS. This was especially true in areas where AWiFS coverage was compromised because of a lack of AWiFS collects or regions with persistent haze or clouds. Many of the CDLs prior to 2006 relied on some, if not all, Landsat data as well. In rare cases, where little or no mid-resolution data existed, portions of the classification were supplemented with 250 meter data from the Terra satellite. In particular, Terra's Moderate Resolution Imaging Spectroradiometer (MODIS) 16-day composited Normalized Difference Vegetation Index (NDVI) imagery was used. Regardless of sensor, most of the imagery was acquired during the summer of the current growing season. Input imagery from the spring and prior fall was sometimes used as well when available and deemed useful. Although not technically imagery, other ancillary data sources were also relied upon to further improve the classification. Specifically, elevation data from the National Elevation Dataset (NED) (Gesch *et al.*, 2002) and the forest canopy and imperviousness layers from the National Land Cover Dataset (NLCD) (Homer *et al.*, 2007) were used.

Methodology

The CDLs were created using a supervised image classification methodology. In short, all of the collected geo-referenced satellite imagery and ancillary data were initially "stacked" regionally by state within a GIS. Next, samples were taken across the imagery stack from areas of "ground truth" which identified the pixel locations of specific crops. Those sample stacks were then "data mined" to determine what set of multi-spectral rules from the time-series of imagery best predicted what land cover category was found at the ground truth locations. Finally, once all of the classification rules were established, all pixels within the scene were placed into the class they best fit, thus building out the state-wide classification.

High quality ground truth data were the key to driving the classification process and leveraged through a NASS agreement with the Farm Service Agency (FSA). The field level information FSA provided included spatially detailed "Common Land Unit" (CLU) field boundary GIS polygon data and farmer reported "578" administrative data. These records were linked by NASS and thus created a large and timely database of agricultural

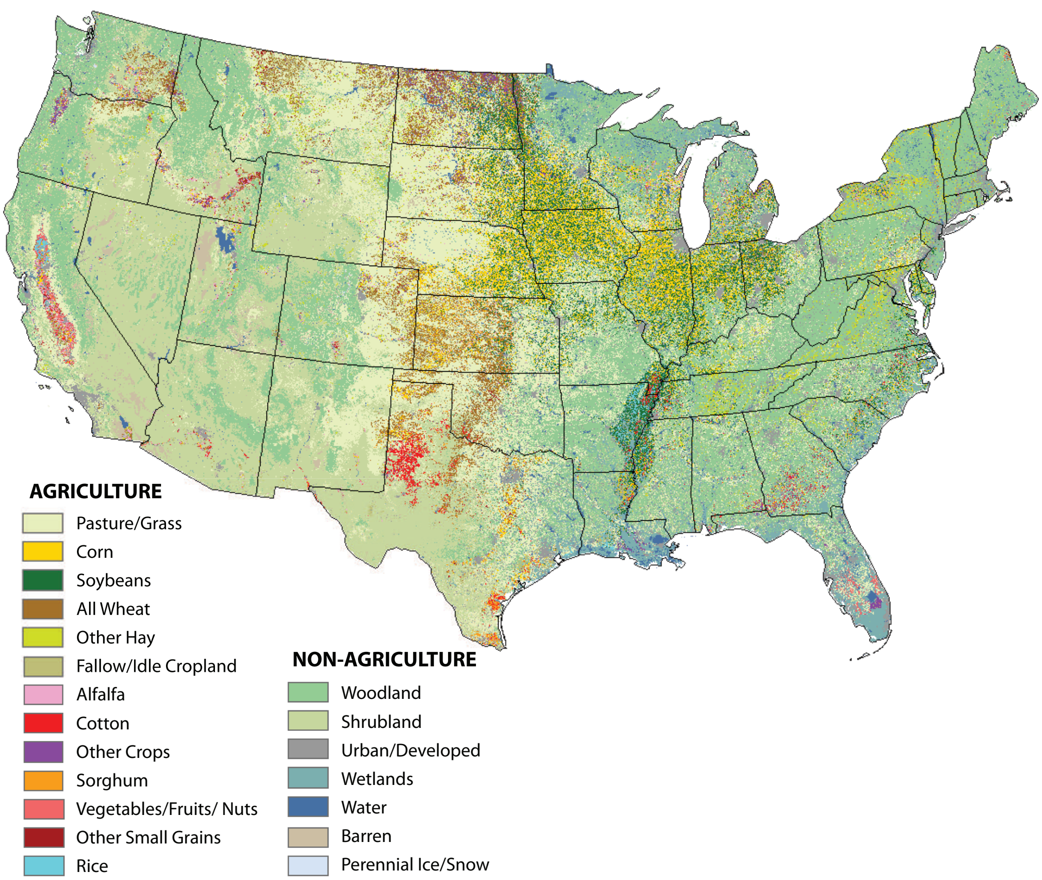

Figure 1. The Cropland Data Layer mapped at the national level and highlighting the distribution of major crops across the conterminous United States.

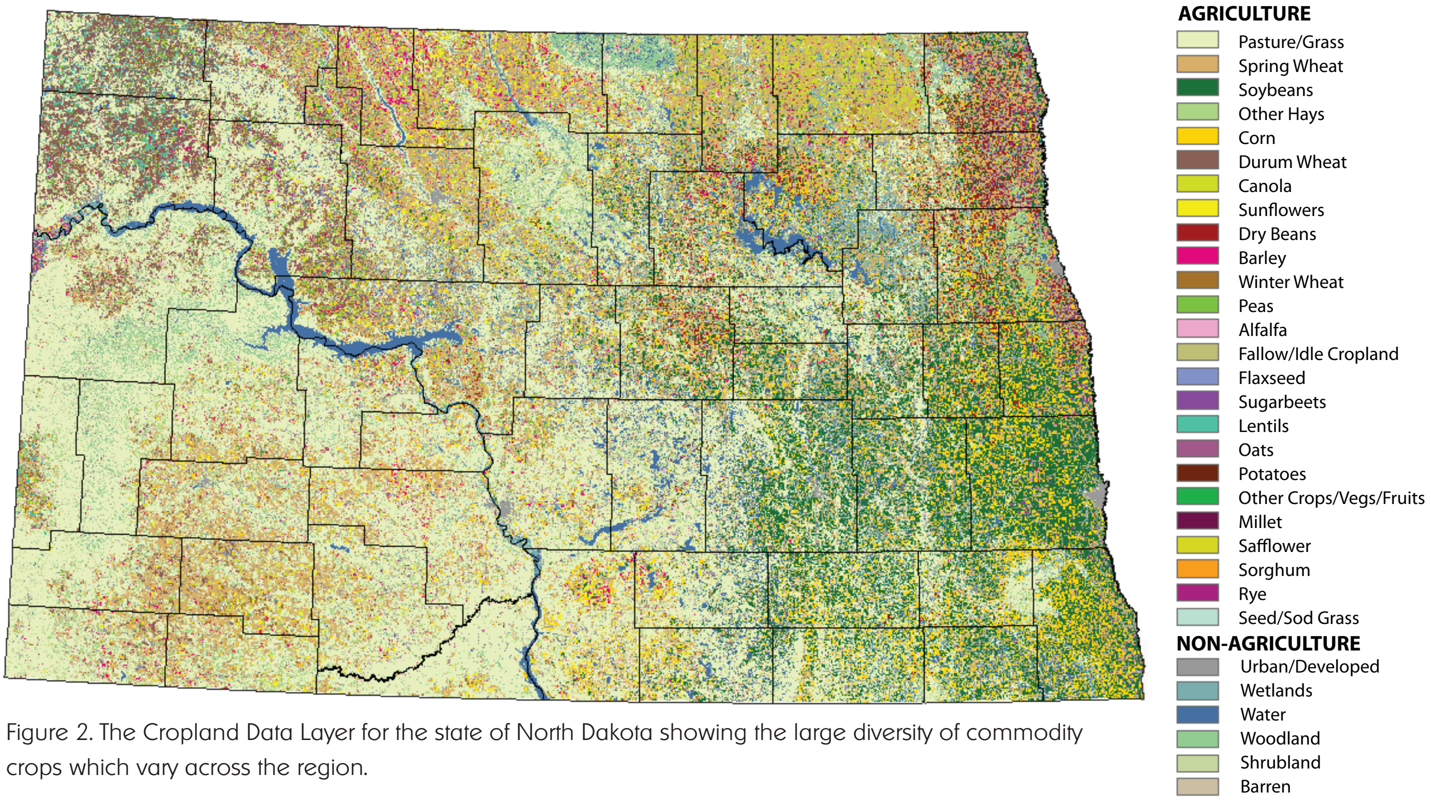

Figure 2. The Cropland Data Layer for the state of North Dakota showing the large diversity of commodity crops which vary across the region.

land use of many farm fields throughout the nation. The database was most comprehensive for the major summer grown commodity crops and provided a very good sample for training of the image classifier. Of note, neither of the FSA datasets are currently available for public access or distribution due to FSA confidentiality laws.

A second source of ground truth data were required to gain information about non-agricultural regions since the FSA data contained none. The land cover component of the 2001 NLCD was chosen as a proxy because of its national coverage and relatively up-to-date nature. Ultimately, the NLCD data were sampled at a similar proportion as the FSA data while ignoring the "cropland" category since it would have been redundant.

The crux of the production and high quality derivation of the classification tree rules, was performed by Rulequest Research See5.0. It was robust and efficient and therefore implemented versus other methodologies such as maximum likelihood or segmentation/object oriented approaches. ERDAS Imagine and Environmental Systems Research Institute (ESRI) ArcGIS were used for imagery and polygon management, respectively. All software and data storage was run on networked Microsoft Windows based desktop personal computers in a non-enterprise environment.

Specifications

The CDL product is based on a 56 meter cell-sized grid covering the conterminous U.S. The native map projection is Albers equal-area conic, same as used by the NLCD. The data are also available in the Universal Transverse Mercator (UTM) projection. File formats are GeoTiff (.tif) and Imagine (.img). The CDLs are distributed by state but they are "seamless" and can be merged for interstate analysis. There are dozens of crop categories within the classification. General non-agricultural categories, analogous to those in the NLCD, also exist. The data can be directly read, viewed, and analyzed by most commercial GIS and remote sensing software packages. Figures 1, 2, and 3 show examples of the CDL mapped at national, state, and county level scales, respectively.

State-level accuracy assessment information is available in metadata to give the users an idea of the confidence of each crop category's map accuracy. Pixel level accuracies for the agricultural cover types vary by crop type and region. Typically, they are best for the most intensive agricultural areas like those found in the U.S. "Corn Belt," and Mississippi River "Delta." The accuracies for dominant crops like corn, soybeans, wheat, rice, cotton typically exceed 90 percent. Less common, but still plentiful, crops like potatoes, sunflowers, Canola, and barley average around 80 percent. The overall crop accuracy for all states is estimated to be 78 percent. States most dominated by intensive cultivation have excellent accuracies in excess of 95 percent. The accuracy of non-cropped areas are not measured but assumed to be reasonable.

continued on page 1204

continued from page 1203

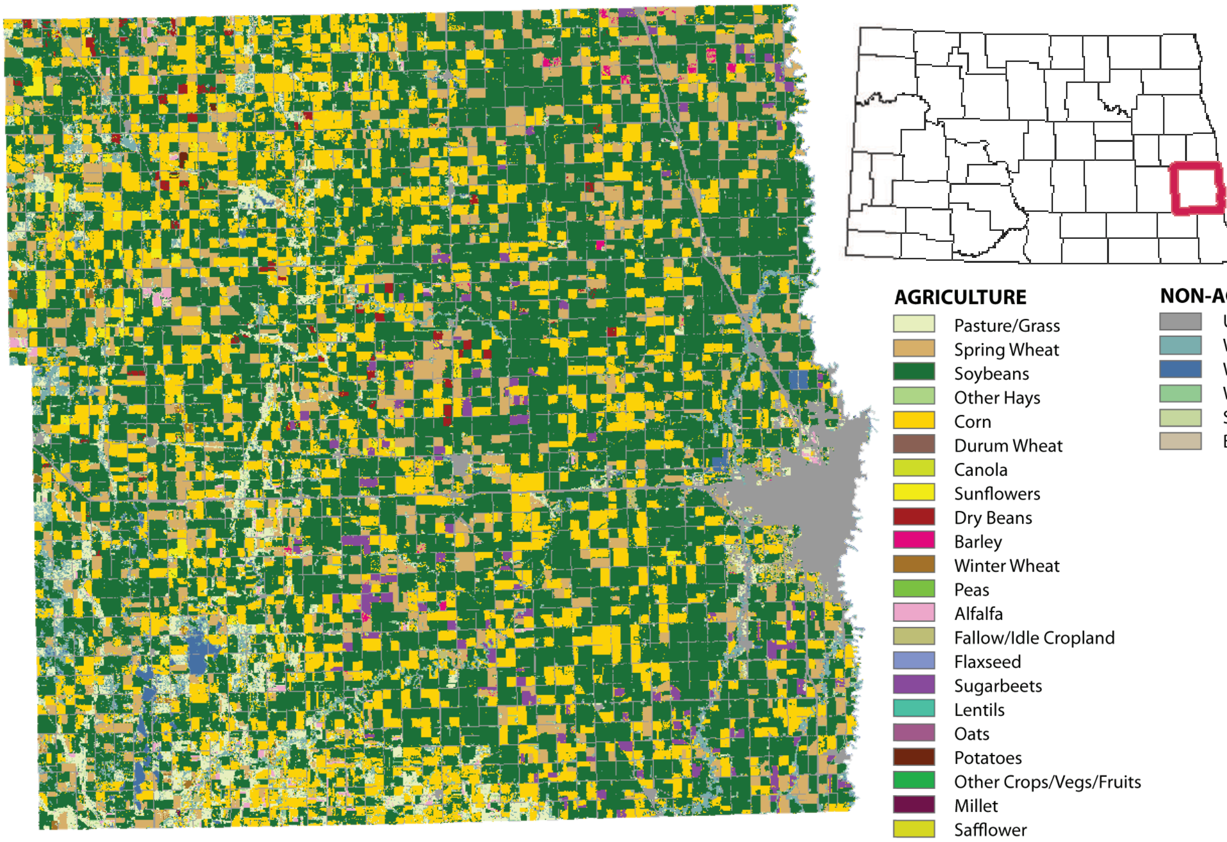

Figure 3. The Cropland Data Layer for Cass county North Dakota showing how field level details and cropping patterns become apparent at a local scale.

Utility

The ultimate mission of the NASS is, "to provide timely, accurate and useful statistics in service to U.S. agriculture". To achieve this goal, NASS conducts hundreds of surveys every year collecting information on virtually every aspect of agricultural activity. The CDL has now become another component to the agency's mission. Many of the dominant commodity estimates NASS derives from the CDL are used to guide or affirm in-season survey derived acreage indicators established by the NASS Agricultural Statistics Board and Field Offices. NASS also actively uses the CDL to further refine the sampling strategies for its annual flagship June Agricultural Survey. Additionally, the CDLs are a key component of NASS crop yield modeling research, and have been used by NASS to help isolate by commodity type crop areas that have experienced flood, hail, or excessive snow.

Outside of NASS the CDL products have been used by a wide variety of people desiring agriculture related land cover information for research, agri-business, or environmental applications. They include various governments, companies, not-for-profit organizations, academics, scientists, educators, students and crop producers themselves. Known uses of the CDL include ground truth for other classification efforts, time-series crop phenology analysis, field rotation modeling, crop area expansion monitoring, ethanol plant locational analysis, epidemiological studies, and small area land cover estimation.

Moving Forward

The 2009 CDL is the culmination of much research and development but is also believed to be a starting point for a new era of annually generated comprehensive and consistent agricultural land cover classifications. NASS plans to release a new CDL each winter with ongoing improvements to resolution and accuracies. The first likely near-term change is to push the product to a finer spatial resolution as this is known to improve the appearance and accuracy of the classification. A 30 meter resolution, which is somewhat of a "standard" among GIS datasets such as land cover and elevation, will be the first goal. Beyond that it is hoped to increase resolution further to a 20 meter product given there are a few sensors already currently collecting at that detail and combined are capable of providing frequent temporally national coverage. Other ongoing research and development improvements to the product include better discrimination between agricultural and non-agricultural herbaceous cover types, higher accuracies in specialty crops, and an earlier release schedule to the public.

The biggest obstacle to the future of the CDL program is whether or not there will be the continued access to low cost or free mid-resolution, multi-spectral, and multi-temporal imagery. Unfortunately, the availability of this type of data in the near future is tenuous. It is particularly true in the U.S. due to the aged and compromised Landsat missions with a replacement sensor not expected until 2013. However, NASS is open to and investigating other imagery sources, such as those from commercial or non-U.S. programs, and thus it is ultimately hoped future CDL analysis, and the value that NASS and others gain from it, will continue.

Acknowledgements

The 2009 CDL was a team effort by Claire Boryan, Mike Craig, Lee Ebinger, Mary Lindsey, Marty Saffel, Bob Seffrin, Fred Shore, Gail Wade, Patrick Willis, and Audra Zakzeski in addition to the authors. Additionally, funding was provided by the U.S. Environmental Protection Agency/Landscape Ecology Branch which helped complete 2009 CDL through the purchase of additional imagery and staffing time.

Data Access

The 2009 CDL can be downloaded directly from NASS at http://www.nass.usda.gov/research/Cropland/SARS1a.htm. Older products are available at the Geospatial Data Gateway http://datagateway.nrcs.usda.gov/.

References

Carfagna, E. and Gallego, F. J., 2005. Using Remote Sensing for Agricultural statistics, *International Statistical Review*, 73 (3): 389-404.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler, 2002. The National Elevation Dataset, *Photogrammetric Engineering & Remote Sensing*, 68 (1): 5-11.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham, 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 73 (4): 337-341.

Johnson, D. M., 2008. A comparison of coincident Landsat-5 TM and Resourcesat-1 AWiFS imagery for classifying croplands, *Photogrammetric Engineering and Remote Sensing*, 74 (11): 1413-1423.

Authors

David M. Johnson and Richard Mueller
United States Department of Agriculture/National Agricultural Statistics Service
3251 Old Lee Highway, Suite 305
Fairfax, VA 22030
dave_johnson@nass.usda.gov, rick_mueller@nass.usda.gov