Comparing Raster-based and Survey-based Cropland Change in Illinois, 1999 to 2015

Alicia Waller | Department of Geology and Geography | Eastern Illinois University

The Problem

When examining problems related to land use/land cover (LULC) change, researchers rely on specific data sources to conduct analyses relevant to their research. Two main sources of land cover data are satellite-derived raster data from the USDA's Cropland Data Layer (CDL) and the survey-based acreage data from USDA's NASS June Survey. Both sources are useful in providing detailed information on LULC, however, there is a degree of error associated with each one that should be acknowledged before using the data in studies (Larsen and others 2015).

The USDA's Cropland Data Layer is a derived from satellite imagery, meaning the LULC type is determined by the unique spectral reflectance values of objects detected by the satellite. The problem with this type of data acquisition is that similar-looking crop types or plant species can sometimes be mistaken for each other, causing error in the classification of LULC data. Xie and others (2008) discussed this error in their paper about using remote sensing imagery to classify and map vegetation. They speculated that similar vegetation types on the ground may end up having different spectral values or vice versa. This error can lead to misclassification of crops like corn and sorghum or varieties of small grains because of their spectral similarities.

The survey-based data from the USDA's National Agricultural Statistics Service has its own degree of error. Because the dataset originates from numbers provided on surveys, it is subject to over- or under-estimation based on sample size and the accuracy of the returned surveys. Both datasets carry with them a certain amount of error due to the uncertainties related to spectral data classification (CDL) or surveys (NASS June Survey). The purpose of this research was to analyze the comparability of the CDL and June Survey by studying the consistency of reported acreage and changes over time.

Data Sources

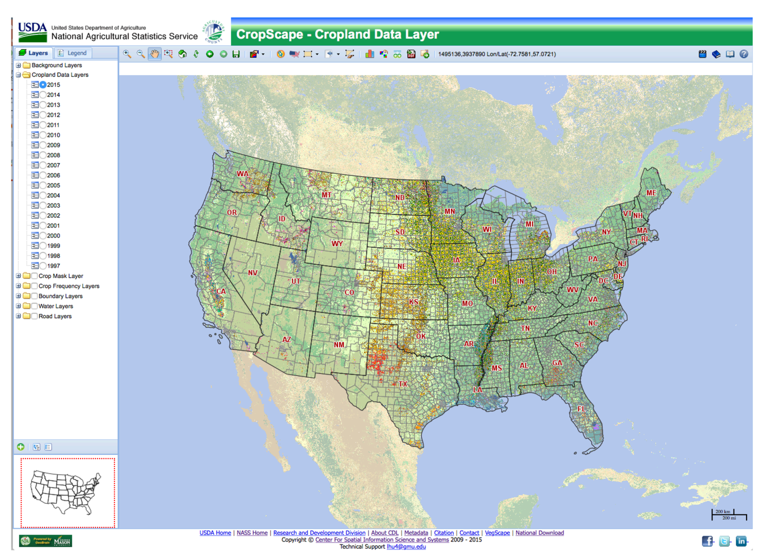


Figure 1: Cropland Data Layer web interface, http://nassgeodata.gmu.edu/CropScape/

The CDL (Figure 1) is a raster-based data product created and released annually by the USDA's National Agricultural Statistical Service. In the 1970s and early 1980s NASS had a goal to use multi-spectral imagery to estimate acreages of major commodity crops including corn, soybeans, and wheat (Craig 2010).

Inputs into the current CDL data come from five sources: the AWiFS (56-meter resolution) sensor mounted on a satellite flown by India; 2) NASA Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM+) imagery at 30-meter resolution via Landsat 5, 7, and 8 satellites; 3) Terra's MODIS (250-meter resolution) satellite; 4) the USDA Farm Service Agency's (FSA) Common Land Unit (CLU) data used for ground truth; and 5) National Land Cover Database (NLCD) data that are used to classify image-elements deemed "non-agricultural." For a more detailed treatment on hardware and data specifications and methods see Boryan and others (2011), Johnson and Mueller (2010), and Johnson (2013).

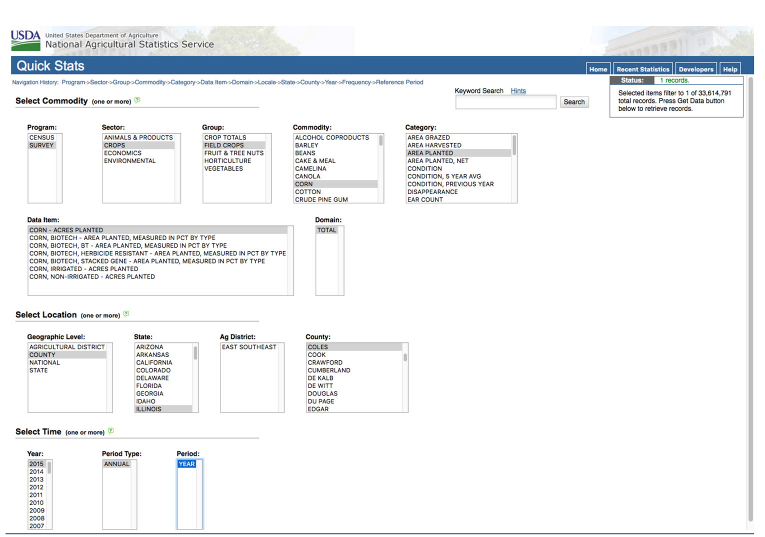


Figure 2: NASS June Survey (QuickStats) web interface. http://quickstats.nass.usda.gov

Each June NASS conducts one of its largest annual surveys (Figure 2). Data gathered through the June Area survey are the foundation on which the NASS survey program functions and they provide statistical measures for quality assurance involving other NASS products (USDA, NASS 2015). The framework for the survey consists of nearly 11,000 one square-mile parcels of land from which land use data are obtained. Farmers who operate individual land units within each square mile surveyed are interviewed. NASS reports that most years approximately 85,000 individual tracts of land from within the 11,000 sample areas are identified. These produce some 35,000 interviews/surveys conducted with individuals responsible for farming or maintaining that land (USDA, NASS 2015).

CDL vs. June Survey

There are numerous sources of LULC data that scientists use to determine/map extent and change. Laingen (2015) compared cropland data from four sources, both survey- and raster-based, and found cropland totals and change varied by as much as 5-million acres (Figure 3). While such differences are explainable based on what data are being used, those using the data typically do not report 1) the methods by which the data were obtained or 2) the data's shortcomings. This can lead to dubious conclusions.

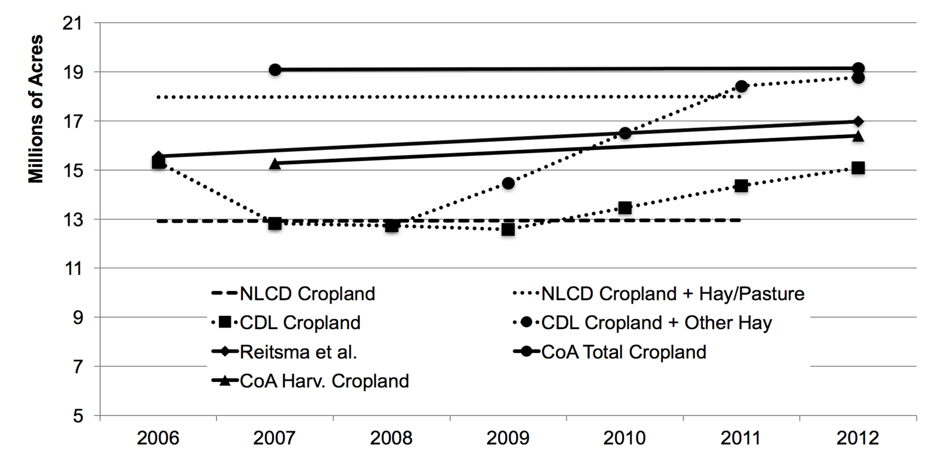


Figure 3: U.S. cropland extent and change derived from four sources (Laingen 2015).

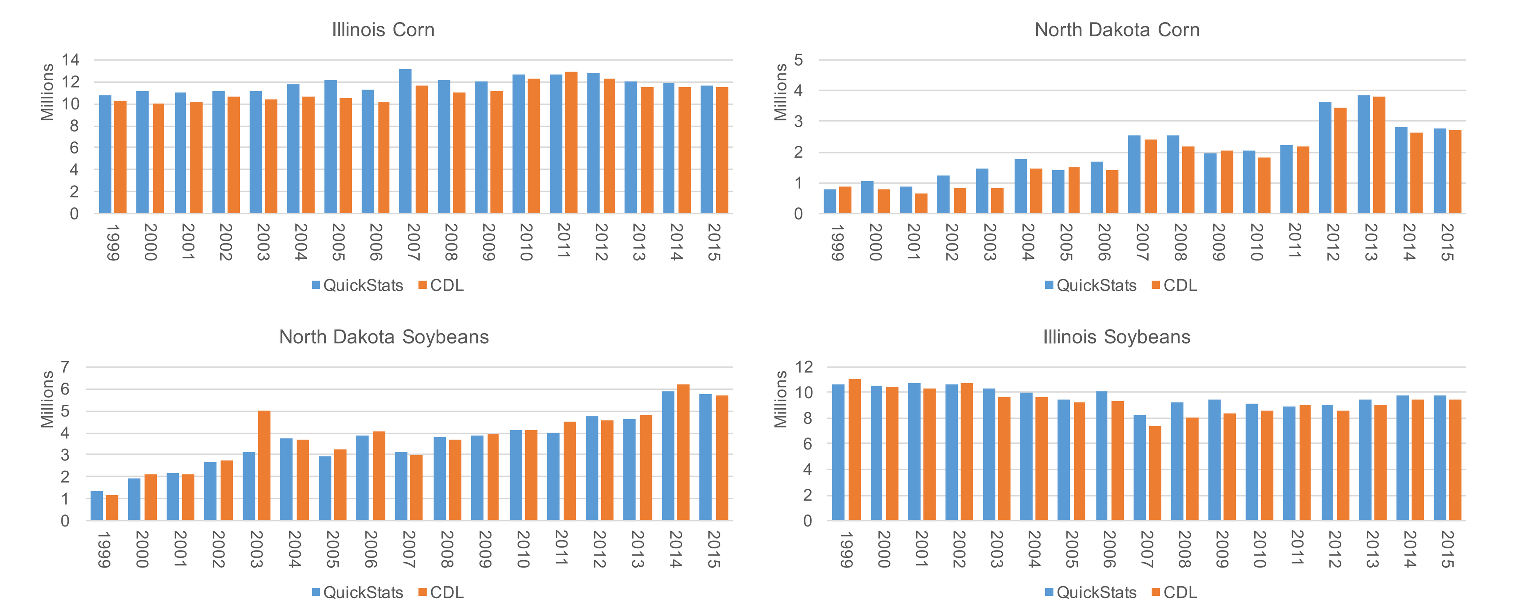
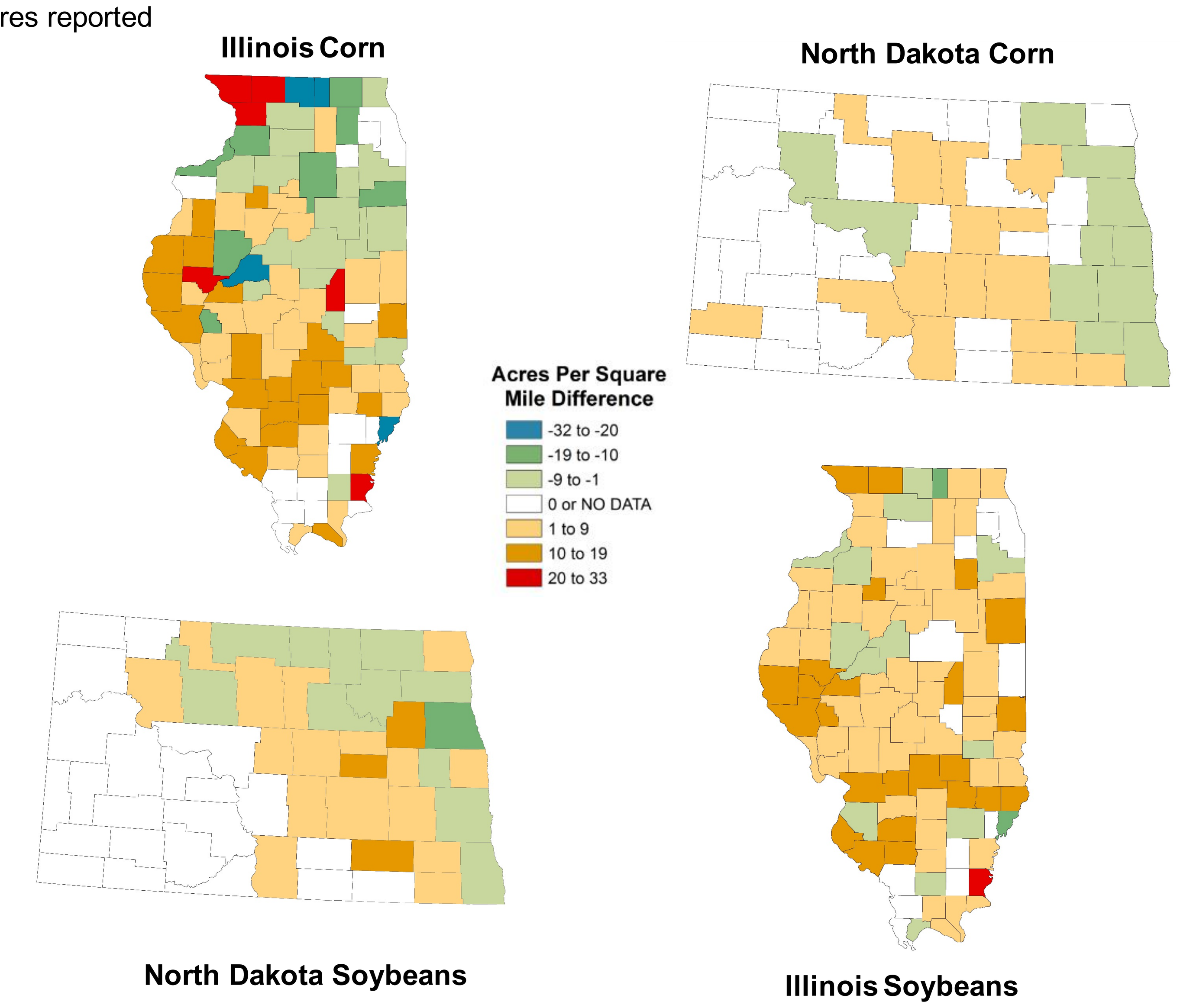
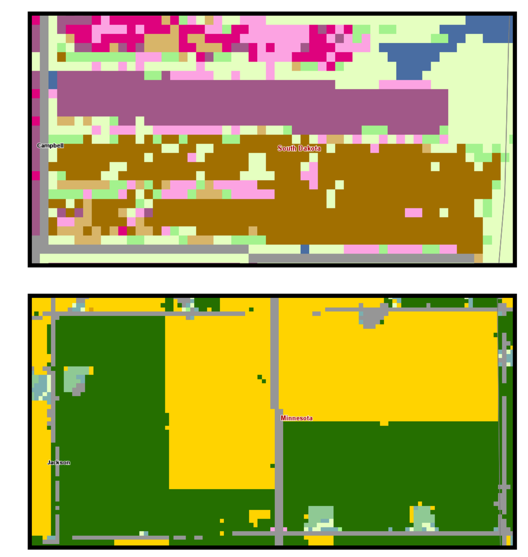


Figure 4 (above) & 5 (below): Annual changes in corn and soybeans in North Dakota and Illinois from 1999 to 2015 using both the NASS June Survey (QuickStats) and Cropland Data Layer (CDL) data. When looking at the percentage difference between the data reported from each source, Illinois corn was off by an average of 7.27% and soybeans an average of 5.55%, while North Dakota corn was off by 18.17% and soybeans 6.82%.

While no discernible patterns exist when mapping differences between the June Survey and CDL acreages for planted corn and soybeans in Illinois and North Dakota, counties in shades of orange/red indicated higher June Survey acres reported, while those in shades of blue/green indicated higher CDL acres reported



Analysis



Landscapes comprised of crops that are relatively easy to distinguish from one other, such as corn and soybeans, tend to have lower error rates versus landscapes where small grain and grass/pasture are more common. At the state level corn acreage reported by CDL and NASS were typically less different from one another (R2=0.99) whereas wheat (a small grain) acreage, was more variable (R2=0.87) and NASS June Survey data acres – reported by farmers – was higher in all but two states (Figure 7, lower right). Which shows that CDL was under-counting acres of oats due to confusion with other crops.

Figure 6: Spectral confusion in CDL between barley, oats, winter/spring wheat, alfalfa, pasture/hay (left, top) and the relative spectral clarity between corn and soybeans (left, bottom).

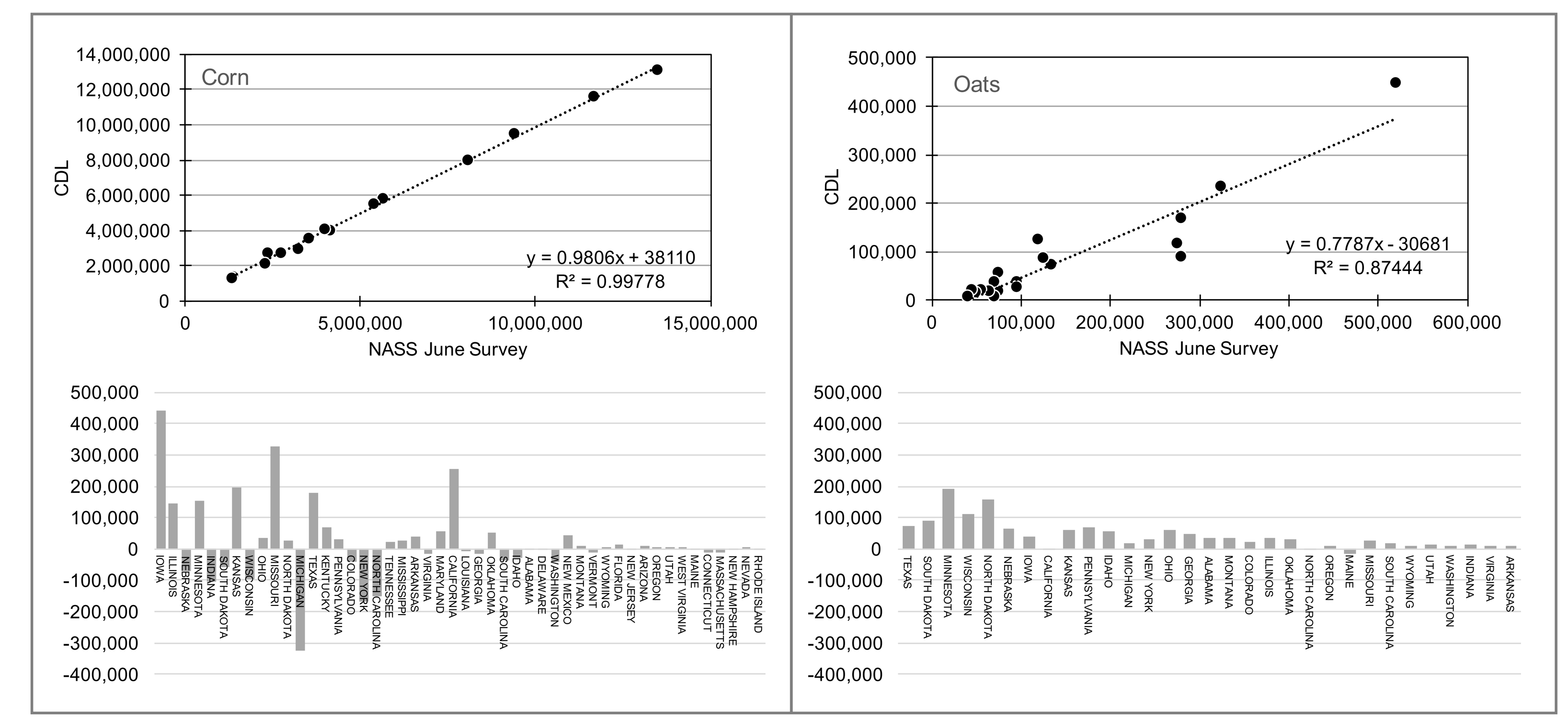


Figure 7: Scatterplots comparing acreage data from CDL and the NASS June Survey for Corn and Oats in 2015 (for the leading states of production equaling 90% of national total). Bar charts beneath each shows differences in acreages by state. Positive values indicate NASS June Survey acres > CDL acres.

Conclusions

Varied results were found when the data were compared with one another, which stressed the importance that data users must consider and question how the data were created, processed, and disseminated before sweeping conclusions can be made with respect to what those data reveal. Data and maps should continually be scrutinized for suspicious patterns (Gallant 2009), such as those found when mapping and analyzing land cover data using disparate datasets, and great care must be taken, if those data are used, when describing the methods by which those data were acquired, and any shortcomings that may be present in them. As creators and users of spatial data and map-based products – hardcopy, virtual, or otherwise – we all have a three-fold responsibility to 1) the data, 2) to the map reader(s), and 3) to anyone who might be affected by a foreseeable misinterpretation of the data or our conclusions (Gersmehl 1985).

References

- Boryan, C., Z. Yang, R. Mueller, and M. Craig. 2011. Monitoring U.S. Agriculture: the U.S. Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International* 26(5): 341-358.
- Craig, M. 2010. *A History of the Cropland Data Layer at NASS*. http://www.nass.usda.gov/Research_and_Science/Cropland/CDL_History_MEC.pdf. USDA,Washington, DC.
- Gallant, A.L. 2009. What You Should Know About Land-Cover Data. *The Journal of Wildlife Management* 73(5):796-805.
- Gersmehl, P.J. 1985. The Data, The Reader, and the Innocent Bystander – A Parable for Map Users. *The Professional Geographer* 37(3):329-334.
- Johnson, D.M. 2013. A 2010 Map Estimate of Annually Tilled Cropland Within the Conterminous United States. *Agricultural Systems* 114: 95-105.
- Johnson, D.M., and R. Mueller. 2010. The 2009 Cropland Data Layer. *Photogrammetric Engineering and Remote Sensing* 76(11):1201-1205.
- Laingen, C.R. 2015. Measuring Cropland Change: A Cautionary Tale. *Papers in Applied Geography* 1(1):65-72.
- Larsen, A.E., B.T. Hendrickson, N. Dedeic, and A.J. MacDonald. 2015. Taken as a Given: Evaluating the Accuracy of Remotely Sensed Crop Data in the USA. *Agricultural Systems* 141: 121-125.
- USDA, NASS. 2015. *June Area*. http://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/June_Area/. USDA-NASS, Washington, DC.
- Xie, Y., Sha, Z., & Yu, M. 2008. Remote Sensing Imagery in Vegetation Mapping: A Review. *Journal of Plant Ecology,* 1(1): 9-23

Faculty Mentor: Christopher R. Laingen