Uncertainty in Cropland Data Layer derived land-use change estimates: putting corn and soy expansion estimates in context

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Ryan R. Noe

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Jason Hill

January 2015

Acknowledgements

This research was funded in part by the USDA CenUSA Bioenergy grant. Bonnie Keeler aided in selecting the research question, and provided feedback throughout. Marvin Bauer provided comments on an early draft. I cannot thank Brian Krohn and Kimberly Mullins enough for their unmitigated willingness to provide feedback on any and all questions I have encountered since I started working for Jason. Their moral support and intellectual guidance was invaluable for completing this research and has helped me grow as a scientist. I thank Jasmine Cutter for proof-reading, comments, and encouragement. I am grateful to Natalie Hunt, Nat Springer, and all members of the Hill lab group past and present for their feedback and support. I am indebted to my family and friends for their endless encouragement. Lastly, I am grateful to my adviser, Jason Hill, for the resources and freedom to learn new skills and explore this and other research questions.

Abstract

Increased demand for corn for ethanol and the subsequent record high commodity prices has resulted in rapid expansion of corn and soy in the Midwestern U.S. Whether or not this expansion is replacing existing agricultural production or is expanding onto previously uncultivated grass or pasture land has profound implications for ecosystem services such as soil carbon storage, soil erosion prevention, and water quality. Several studies have used the Cropland Data Layer (CDL) to track fine scale land-use changes driven by corn and soy. However, these studies rarely account for the variability in data quality throughout the CDL's history. Here I compare established techniques as well as the application of USDA's Common Land Unit (CLU) data for removing 'noise' from change rasters and quantifying the land covers lost to corn and soy expansion. I compare these estimates to equivalent measures from the National Agricultural Statistics Services (NASS) and use the discrepancy between them to identify spatial and temporal variability in CDL data that could influence land-use change study results. The CLU results differed little from established techniques, but both improved over direct comparisons of CDLs. Comparison to NASS data revealed pre-2010 versions of the CDL underestimate corn and soy area much more than later versions, leading to the detection of illusory land-use change when they are compared to post-2010 versions. Spatial and temporal variability resulted in errors that were several times larger than the trends the data are being used to detect. According to the CDL, approximately five million hectares of corn and soy expanded onto the grass, pasture, hay, and wheat between 2007 and 2012. However, over

the same time period, the CDL overestimated the amount of corn and soy expansion by 3.5 million hectares in the unmodified treatment and by 1.5 million with cleaning methods applied. This work suggests that studies that use the CDL should test for and report variability and uncertainty in their results.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Observed recent land-use change

Increased demand for food and fuel has resulted in elevated commodity prices, motivating producers to increase production (Gecan et al. 2009; Claassen & Nickerson 2011). In the last decade, the area devoted to corn and soy production in the Midwestern United States (defined as IA, IL, IN, KS, MN, MO, ND, NE, OH, SD, and WI) has reached record highs eight times. In this domain, production of these two crops increased by 6.25 million ha between 2006 and 2014 (USDA-NASS 2014b). National Agricultural Statistics Service (NASS) census data show large increases in the area of cropland and even larger decreases in the area of pasture between 2007 and 2012 (Figure 1). While this suggests possible corn and soy expansion onto pasture, it could also be explained by increasing pressure from urban expansion or a transition from pasture to hay.

Though NASS data show unambiguous increases in the area devoted to corn and soy production, a lack of spatial detail and data on natural vegetation prevent it from being used for a comprehensive understanding of land-use change. Importantly, the lack of a sub-county spatial component precludes identifying what specific transitions are occurring. It is possible to know that the area of one crop is increasing and another is decreasing, but whether the land-use changes of those crops is direct or indirect cannot be ascertained from NASS data. Simply tracking change in total area can hide a complex web of land-use change and the environmental consequences of those transitions. At the county level, the finest spatial resolution available, only a few major crops are reported on an annual basis, and pasture area data are only available in census releases every five years. Furthermore, strict privacy concerns result in data being redacted if there are only a few producers in a given county. Inconsistent temporal reporting of many crops and the sparse non-cropland data make even tracking correlations between increases of one land cover and decreases of another difficult.

1.2 Implications for ecosystem service provisioning

Whether corn and soy expansion is occurring on existing agricultural land or on natural vegetation has a spectrum of consequences for ecosystem services. Grassland and other natural vegetation enable the accumulation and long-term storage of soil carbon. Many studies have shown at least a 50% loss of soil carbon when converting grass or pasture land to cropland (Bowman et al. 1990; Gebhart et al. 1994; Unger 2001). Soil carbon losses are even higher in areas with precipitation matching that of the Dakotas, Nebraska, and Kansas (Guo & Gifford 2002). Given that one of the goals of U.S. biofuel policy is to reduce the carbon intensity of our transportation fuel, it is especially important to understand if soil carbon is being lost due to increased demand for corn and soy. Emissions from soil carbon under corn and soy expansion have the potential to undermine the already uncertain carbon emission reductions from corn-based biofuels (Mullins et al. 2011; Fargione et al. 2008).

Natural vegetation is also effective at preventing soil erosion and reduces the flow of nutrients into waterways that cause both local pollution and larger-scale eutrophication. Corn, and to a lesser extent soy, require extensive fertilizer and pesticide application relative to grass or pasture land. In the Upper Mississippi River Basin, corn and soy production alone is responsible for 52% and 25% of nitrogen and phosphorous pollution, respectively (Alexander et al. 2008). Excess nitrogen can contaminate local water supplies, increasing the risk for methemoglobinemia (blue baby syndrome) or require costly removal technologies (Wolfe & Patz 2002). Increased nutrient loading has far reaching economic implications. Algal blooms in the Gulf of Mexico results in millions of dollars of damage to the tourism and fishing industries (Downing et al. 1999). Additionally, the shallow root system of corn makes the land used to produce it vulnerable to soil erosion. Sediment run-off not only decreases the long term productivity of the land, it increases turbidity in local waterways which can change the species composition (Fargione et al. 2009).

Perhaps most difficult to quantify, grasslands can provide services such as pollination, pest control, and wildlife habitat (Werling et al. 2014). A comprehensive assessment of land-use change in the Midwest is necessary to understand how elevated commodity prices translates into changes in ecosystem service provisioning.

1.3 Application and uncertainty of the Cropland Data Layer

1.3.1 Diverse applications of the CDL

Several studies have sought to overcome the limitations of NASS data by using the Cropland Data Layer (CDL) (Mueller & Harris 2013). With a minimum of 56m spatial resolution, eight years of annual Midwest releases, and 133 land covers tracked, the CDL is the ideal dataset to answer questions about rapid agricultural land-use change and the associated environmental consequences. In addition to traditional land-use change studies, discussed in detail shortly, the relatively high spatial and temporal resolution of the CDL enables it to track changes in rotations on a field by field basis (Yost et al. 2014; Stern et al. 2012; Long et al. 2014; Plourde et al. 2013). The crop specific land covers have made it popular with those predicting landscape changes in response to biofuel policy (Li et al. 2012; Elliott et al. 2014). Others use the CDL as an input for water quality, yield, and ecosystem service modeling studies (Karcher et al. 2013; Resop et al. 2012; Meehan et al. 2013).

1.3.2 Uncertainty and sensitivity

While all of these studies are sensitive to variability in the quality of CDL data, the vast majority of these studies make only minor adjustments to compensate for obvious issues (e.g., mismatched resolutions, changing classifications). The ‘as-is’ use of the CDL is sometimes justified by its 85% to 95% overall accuracy for major crops, and often 97% producer and user accuracy for corn and soy (Boryan et al. 2011; USDA-NASS 2014a). However, errors of just a few percent can translate to millions of hectares of improperly classified land when examining major crops across several states. Spatial variability in data quality can also result in estimates that seem reasonable in aggregate, but reveal over- and under- estimates when viewed at finer scales. Furthermore, comparison of multiple datasets multiplies their error rates. All of these potentials for error propagation are especially troublesome given that these data are being used to detect changes on the order of hundreds of thousands of hectares in a 100 million hectare study area (Wright &

Wimberly 2013). While the reported changes may be occurring on the landscape, the data used to detect them are laden with unreported uncertainty.

Accuracy assessments of CDL-derived estimates would add confidence to these results, however the quantity of data and labor required to assess fine scale changes taking place at regional extents over several years is typically unobtainable. Additionally, auxiliary sources of information on grass and pasture are rare. As an alternative, the most prominent studies on the subject have used a combination of conservative assumptions and qualitatively driven processing techniques to produce estimates from the CDL. These assumptions and processing decisions are usually well defended, but the results are not presented in the context of other data sources, or other assumptions. This context is vital to understand how results compare to potential errors in the data.

1.3.3 Wright and Wimberley (2013) assumptions and results

The most cited paper on the subject, Wright and Wimberley (2013), examined a broad area of rapid transition, the five western Corn Belt states (ND, SD, NE, MN, and IA), between 2006 and 2011. They found that the areas of most intense grassland to corn and soy conversion were in the eastern Dakotas and southern Iowa. However, they examined only transitions between corn-soy and grassland. Their reported 530,000 hectares of net grassland to corn-soy conversion does not fully answer where the 1.75 million hectares of new corn-soy production that occurred over this period went, nor does it examine possible indirect effects of corn and soy expansion.

The narrow focus of their study required creating only two aggregate classes; one comprised of corn and soy, and another comprised of the grass/pasture, other hay/non-alfalfa, and fallow/idle cropland classes. This is the only major study to include fallow/idle in an aggregate grass category. They did not include double-cropped winter wheat and soy in the corn and soy class.

Comparisons between the 2006 and 2011 CDLs were made after resampling the 2011 CDL to 56m. To mitigate the numerous individual change pixels that arose from noise in the inputs, they employed a simple but aggressive cleaning method. The change raster was processed with a 5x5 majority filter. This choice of filter was qualitatively justified by arguing that the output resembled the size and shape of fields. The area of detected change removed by the filter is not reported, nor is a comparison to other potential filters presented.

Making minor changes to the assumptions proposed by Wright and Wimberley can drastically change the results. For example, changing the baseline year from 2006 to 2007 quadruples the amount of conversion in the data (Figure 3). The results are also more sensitive to the inclusion of 'Fallow/Idle cropland' when using 2006 as a baseline. While this alone provides little information on which assumptions are more accurate, the magnitude of the sensitivity provides perspective on the robustness of the results.

1.3.4 Johnston (2013) assumptions and results

Johnston's (2013) work expanded on Wright and Wimberley's by considering annual transitions over seven years and by looking at exchanges between six aggregate land classes instead of two. However, the study area was confined to the Dakota Prairie Pothole Region. While an ecologically sensitive area, the non-political boundaries prevent any comparisons to NASS data.

Johnston found increasing grassland coverage between 2006 and 2010, and a sharp decline between 2011 and 2012 resulting in grassland coverage 9% lower than 2006 levels. In total, 800,000 hectares of land converted to corn-soy, 80% of which was formally split evenly between grassland and wheat. Grassland was also lost indirectly as displaced small grains expanded onto grassland. The addition of annual transitions demonstrates that a four year trend can be completely reversed and hidden in a single year. Her work also demonstrates that examining just one transition can hide complex indirect interactions.

This study used the most common definition for its aggregate grass class: ‘Grass/pasture and ‘Other Hay/Non-alfalfa’. Change detection did not employ any cleaning methods to restrict transitions by field size or shape. However, due to poor alignment of roads, any land ever classified as developed was masked out. This assumption may further overestimate the area of the already inflated 56m rasterized roads.

1.3.5 Environmental Working Group and Farm Bureau assumptions and results

In the past two years, three reports by have been released by advocacy organizations with opposing views on corn’s value as a biofuel. While these reports were carried out by professional GIS analysts and make defensible assumptions, the variability in the CDL allowed them to come to very different conclusions.

The Environmental Working Group (EWG) released two reports on the drivers, spatial distribution, and environmental implications of agricultural expansion (Cox & Rundquist 2013; Faber et al. 2012). In their 2012 agricultural expansion report they used the 2008 through the 2011 national CDLs to track expansion of several commodity crops onto natural vegetation. To limit the analysis to realistic management units and mitigate the effects of varying resolutions, the EWG employed filters that eliminated land covers and transitions smaller than 4 hectares. Using these assumptions the EWG found 9.3 million ha of natural vegetation loss to seven crops in just five years. They, however, did not compare this estimate to an independent source, such as NASS data. Over the same time period NASS data show the total area of the same crops decreasing by 700,000 ha.

The Farm Bureau's multi-state land-use change report (2013) examined the extent of land-use change in the Midwest under elevated commodity prices. In the seven state region, they found 3.45 million hectares of net land-use change away from grassy land covers, 2.34 million hectares of which was to corn or soy. Their results are not directly comparable to others because their definition of grassland included wetlands and herbaceous wetlands. They did not employ any filters or cleaning methods. While their findings support the perception of grassland loss to corn and soy expansion, they argue that variability in the data quality of the CDL preclude it from being used for policy making.

1.4 Changes in CDL methodology

Though iterations of the CDL were first produced in 1997, national coverage did not begin until 2008, and coverage for the entire agriculture-dominated Midwest was only available shortly before, in 2006. Despite spanning only eight years in the Midwest, the CDL has undergone many methodological changes. Changes in classifications, satellite sensors, and resolution have expanded the capabilities of the CDL and improved its accuracy, but they also confound analysis of land cover changes in ways for which it is difficult to account and correct.

The desire for more frequent imagery over a broader area prompted the acquisition of imagery from the India Resource Satellite's Advanced Wide Field Sensor (AWFiS) beginning in 2006 (Bailey & Boryan 2010). This enabled broader coverage, but at a coarser 56m resolution compared to 30m for Landsat. AWFiS imagery was used until 2011 when imagery from the satellites Deimos-1 and UK-DMC-2 enabled the production of a consistent national 30m CDL. While this represents and improvement in data quality, it also makes aligning and comparing older and newer datasets more difficult.

In addition to changes in imagery source, the CDL has had varied auxiliary classification data. While the CDL captures and classifies non-agricultural imagery each year, the training data required to perform the classification comes from the less frequently released National Land Cover Dataset (NLCD) (Jonhson & Mueller 2009). Using already-classified data for training increases the potential for error propagation because classification errors that exist in the NLCD would carry over to the CDL.

Additionally, the NLCD is released every five years, but the data that go into a given release can span several years before the release, and the product is not available until two years after its namesake year. This results in using imagery that is between five and ten years old to classify non-agricultural land covers. The five year release cycle also results in sudden, large fluctuations in non-agricultural land covers, which can mask actual land-cover change.

With one exception, since 2006 the CDL has used a consistent collection of commercial software products to classify its input imagery. However, there are classes in later versions of the CDL that did not exist in the earlier versions. For example, cells with the classification 'Sweet Corn' first appeared in 2008. Reclassification into broader categories is the typical practice to correct for these changes, but no guidance is provided on whether sweet corn grown in 2007 was most likely classified as 'Corn', 'Other Crops', or 'Misc Vegs & Fruits'. These discrepancies between years typically affect classes that have relatively small area, but they add uncertainty to comparisons over longer time periods.

The 2006 Nebraska CDL differs from other CDLs in that it was classified with an older method that limited the number of scenes that could be used to make the classification (USDA-NASS 2014a). It is also unique in that the non-agricultural land covers are taken directly from the 2001 NLCD, rather than using the NLCD for training data. Furthermore, it is the only modern CDL to have 'Undefined/Non-agricultural' cells. Of the over 900,000 hectares of undefined land, approximately 580,000 and 732,000 ha were classified as ‘Grass/Pasture’ in 2005 and 2007 respectively (Figure 2). These factors contribute to anomalies that make the 2006 Nebraska CDL difficult to compare even to other states of that year.

1.5 Research Objectives

I attempt to answer two questions that will improve our understanding of land-use change in the Midwest. First, what is the quantity and prior land cover of new corn and soy land? Second, are those changes large enough to be distinguished from, or potentially influenced by, variability in the CDL? I test the application of Common Land Unit (CLU) data and established raster-based techniques for detecting land-use change with the CDL, and compare previous studies’ estimates to equivalent estimates from NASS at regional, state, and county levels. With both spatial and temporal comparisons between CDL-derived estimates and NASS, I identify anomalous areas and years of the CDL. I put my land-use change estimates, as well as previous published estimates, in the context of the uncertainty demonstrated in the data.

2. Methods

2.1 Study area data preparation

The 11-state study area was selected to encompass the states that make up the core of the Corn Belt and to also capture transitions that are happening at its periphery. These states made up approximately 80% of 2013 U.S. corn and soy production (USDA-NASS

2014b). Their selection was also contingent on data availability. States were excluded if they did not have a 2007 CDL. The state of Michigan and the Minnesota counties of Polk and Otter tail were excluded due to incomplete CLU data. The 11-state (minus two counties) study area was extracted from seamless national CDL files for 2008 to 2013. Individual state files from 2006 and 2007 were first projected to Albers Equal Area Conic and then mosaicked together. Alignment of roads was checked visually and by using the ArcGIS Combine tool to determine if alignment could be improved by shifting the cells. The 2007 CDL was found to have better alignment when shifted north one cell (56m). CDLs with 30m resolution were resampled to 56m for comparisons to earlier CDLs. However, for CLU analysis and comparisons between 30m CDLs, the higher resolution was retained.

2.2 Software and data sources

Seamless national CDL files for 2008-2013 were obtained from the USDA CDL website (USDA-NASS 2013). Individual state CDLs for 2007 and 2008, state boundary, and county boundary shapefiles were obtained from the geospatial data gateway. The CLU data used in this analysis were limited to what was publically available in 2008, and were obtained from www.geocomunity.com. Later versions of the CLU were not made public. NASS data were all obtained using the quickstats query tool available at quickstats.nass.usda.gov (USDA-NASS 2014b). All GIS operations were performed using ESRI ArcGIS 10.2 (ESRI 2013). Operations on tabular data were performed with the Python data analysis package Pandas 0.15 (McKinney 2010).

2.3 Treatments and metrics

Treatments were selected to represent common applications of the CDL while also maintaining comparability with NASS data. To test the influence of temporal variability in the CDL's methodological history, two combinations of comparison years were selected. The comparison between 2007 and 2012 captures much of the range and subsequent variability of the CDL's history and also allows for direct comparison with NASS pasture data. The years 2011 and 2013 were selected for comparison because, despite the short time frame, there was significant corn and soy expansion over that time period, but more importantly those CDLs were produced with the most consistent data sources and methodology.

To test the ability of cleaning methods to improve the accuracy of change estimates, no cleaning was compared to a method prescribed by Wright and Wimberley as well as a method employing a dataset of land management units derived from high resolution aerial imagery. These methods are discussed in detail in section 2.5.

Due to limitations of NASS data, the accuracy of a given treatment is determined by the net change in area of corn and soy. The net change metric is not ideal for assessing the environmental consequences of transitions, because the consequences of those transitions may not be equivalent. For example, converting pasture to corn may release soil carbon faster than it is sequestered by converting corn to pasture.

The net change metric is also limited in its ability to assess spatial accuracy. While the CDL may estimate the change in corn and soy area for a given state perfectly relative to NASS data, there is no guarantee that the location of those changes or their prior land covers are accurate. To improve the spatial component of this metric, comparisons of change in total area are made at regional, state, and county levels to reveal potential spatial trends in error that would be hidden in an aggregate estimate.

2.4 Aggregation and re-classification

2.4.1 Importance of aggregation and re-classification

Aggregation and re-classification is a vital part of any multi-year study that uses the CDL because some classifications are now in use that did not exist in earlier years. Maintaining comparability with independent datasets such as NASS requires ensuring that any aggregation is equivalent in both datasets. Classification ambiguity, both between studies that use the CDL and within the CDL itself, is especially common in grassland classes. Despite its importance, there is usually very little justification, quantitative or otherwise, for the re-classification decisions in a given study. Here, quantitative and qualitative justifications are provided for classifications that are inconsistently called grassland between studies.

2.4.2 Treatment of 'Other Hay/Non-Alfalfa'

Some re-classifications are straightforward, for example, the class 'Other Hay/Non Alfalfa' was not used before 2009, but 87% of land with that classification in 2009 had been classified as grass/pasture in at least two of the three prior years. Given the prior land cover, one can infer that 'Other Hay/Non Alfalfa' should be considered a grass land cover. Although including this class in the aggregate grassland class increases the area of grassland, including it will always result in a lower net grassland loss estimate. Because this land cover did not exist earlier time period, it impossible for it to contribute to grassland to corn-soy conversion, but it can contribute to the reverse and thus reduce the net loss estimate. ‘Other Hay/Non-Alfalfa' is also necessary to include because it is inconsistently included on a state-by-state basis, even in recent iterations of the CDL.

2.4.3 Treatment of ‘Alfalfa’

Even though alfalfa has many similarities with other hay and pasture, it was excluded from the aggregate grassland class because of its tendency to be grown in rotation with corn. Unlike other hay, alfalfa experiences autotoxicity and does not grow well on land previously used to produce alfalfa (Chon et al. 2003). When a stand of alfalfa becomes less productive, it is common practice to grow corn before returning to alfalfa (Miller 1983). This frequent exchange between alfalfa and corn would confound estimates of grassland to corn conversion.

2.4.4 Treatment of ‘Fallow/Idle Cropland’

A more ambiguous classification is ‘Fallow/Idle Cropland’. There is evidence that this class has been inconsistently classified between 2006 and future years of the CDL. In 2006 its area was twice as large as any future year. Half of this land was then classified as grassland in at least two of the next three years. Normally this would be an indication that it should be included in an aggregate grassland class. However, looking at the prior classification of 2013 ‘Fallow/Idle Cropland’ indicates that it was most often classified as wheat before 2013.

For further understanding of inconsistencies in ‘Fallow/Idle Cropland’ classification, its effects on amount of net grassland to corn and soy conversion detected were compared in Western Corn Belt states. Definitions of grassland that included ‘Fallow/Idle Cropland’ doubled the amount of net grassland to corn-soy conversion found in comparisons that use 2006 as the baseline, but have almost no effect in comparison that use 2007 as a baseline (Figure 3). Therefore it is likely that this classification was treated differently in 2006. Because 2006 is not being used as one of the comparison years in this study, ‘Fallow/Idle Cropland’ was not included in the aggregate grassland class. However, studies that use the 2006 CDL extensively may achieve better results by including it.

2.4.5 Other classification considerations

Agricultural re-classification decisions were made based on the prominence of the crop and comparability with NASS data. ‘Alfalfa’, ‘Fallow/Idle Cropland’, as well as the ten field crops with the greatest area were maintained. All agricultural covers deemed minor or incomparable between datasets were assigned to ‘Other Agriculture’. Due to low changes rates, non-grass natural vegetation was aggregated into a single category.

2.4.1 Treatment of double cropping

Although most double crop classifications are not present in early versions of the CDL, and were minor in later versions, they were maintained to make accurate comparisons with NASS data. If land other than hay or alfalfa is harvested multiple times in the same year, NASS counts the area twice, once for each crop (USDA-NASS 2012). Because it is not possible to count a raster cell twice, double-crop classifications where neither classification is 'Other Agriculture' are not aggregated with their respective classes until the data have been converted to tabular form, then their area is added to both of the crops grown on that land. For analyses that focus on expansion of individual crops (e.g., corn and soy expansion), double-cropped systems that contained the crop were re-classified to that crop.

2.5 Strategies for minimizing noise

A limitation of several land-use change studies is the identification of change between rasters on an unrealistic cell by cell basis (Johnston 2013; Farm Bureau 2013). Raster cleaning methods have been applied to remove presumably erroneous individual change pixels and visually approximate fields (Wright & Wimberly 2013; Cox & Rundquist 2013), but the ability of these methods to improve the quantitative accuracy of estimates is rarely assessed. I attempt to improve on past approaches by applying the Common Land Unit (CLU). The CLU delineates discreet units of uniformly managed land from permanent features visible in aerial photography (Adkins 2013). By combining the CDL and the CLU one can evaluate how land-use is changing on a field-by-field basis each year. Here, the qualitative considerations behind these methods are discussed, and the ability of the methods to decrease discrepancies from NASS in corn and soy change estimates is compared.

2.5.1 Raster-based moving window filters

A long-used method for removing spurious values and smoothing rasters of all types is a moving window filter. In this class of filters each cell's value is determined by a mathematical function of the surrounding cells. For continuous data, this could consist of taking the average the surrounding eight cells, but for categorical data operations are limited to what can be determined by counting the cells of different categories in each window. Typically, the majority of the surrounding cells is assigned to the analysis cell. Wright and Wimberley (2013), among others, selected a 5x5 window for their majority filter. This requires that of the 24 cells forming a square around a given cell, 13 of them had to be classified as change for the center cell to be classified as change. At 56m resolution this is equivalent to requiring a cell to be touching or one cell away from 4 hectares of changing land. While this is very effective and removing the obviously erroneous cells, it also has a tendency to round the corners of fields and remove small or thin parcels of detected change that could be actual land-use change.

The major advantage of this approach is that it does not require any auxiliary data sources, and is still able to generate a more realistic looking landscape. However, as with any simple algorithm, it is unable to replicate the many factors that go into delineating a landscape.

2.5.2 Common Land Unit based filtering

In an effort to introduce more information into the filtering methodology, I applied the CLU polygon dataset. The CLU improves on the analysis by adding boundaries professionally delineated from aerial imagery to within three meters accuracy. This incorporates the complex patterns that are used to managed the landscape rather than simplify them. The polygons can conform to much finer features than 56m cells, and their area can be calculated to avoid biases from cell counting.

The CLU also provides several benefits when comparing multiple years of data. Foremost, the fixed boundaries allow for tracking changes on a field-by-field basis from year to year with greatly reduced edge effects. Because the CDL is only used to determine the land cover of a field, and not its shape or size, rasters of different resolutions can be compared without information loss from re-sampling. Additionally, the tabular format of the data can be quickly manipulated and queried in ways that would require generating several intermediate files using a raster-based approach.

While the CLU improves on directly comparing cells, it does have some drawbacks. Due to restrictions in the 2008 Farm Bill, the most recent available version of these data is from 2008. Furthermore, the available data have been stripped of their meta-data, which prevents identifying the exact year the units were delineated. Even with somewhat outdated CLUs, the boundaries are likely still appropriate because they are drawn to conform to non-changing features (Adkins 2013). However, in an exceedingly dynamic time for agricultural production, these units may need updating to capture subdivisions of previous uncultivated units. Changes in field boundaries, such as adding stream buffers or additional rows of production to the edges of fields would not be detected by this CLU snapshot. Even small buffers and contours that already exist in the CLU dataset may not be detected or may be improperly classified because of the coarse resolution of the CDL.

Lastly, there is uncertainty with regard to whether or not CLUs truly have a uniform land cover. In contrast to the Farm Service Agency's definition for a field, their definition for a CLU does not preclude multiple crops within a CLU. A CLU is delineated along crop boundaries if the same boundary is used for multiple years (USDA Farm Service Agency 2012). Researchers with full access to the data have suggested that the amount of CLUs with mixed covers is as high as 50%, but no context is given on the size or distribution of these inclusions (Bailey & Boryan 2010; Gelder et al. 2008). This presents a problem because the CLU filter assumes that the majority land cover within it is the land cover of the entire unit. Despite this uncertainty, the CLU adds valuable boundary information, especially when crop lines are consistent between years.

Frequent occurrences of CLUs with mixed land covers may upwardly bias area estimates of more prevalent crops. Additionally, in the unlikely event of a tie, the lower land cover value is assigned, which would be biased toward corn. Ultimately, while the CLU offers new and relevant information to detecting land-use change, it too introduces assumptions and variability that cannot fully be accounted for without an extensive accuracy assessment.

2.6 Common Land Unit preparation

Although the CLU is nominally distributed on a county basis, each file typically had dozens to hundreds of CLUs located outside of the county boundary. To ensure that every CLU in my study area was used, a script was used to perform a spatial join operation between every CLU and a shapefile of states and counties. CLUs were assigned attributes for the state and county in which their centroid fell, regardless of what county file they were distributed in. Additionally, a unique ID was assigned to every CLU to facilitate joins between multiple datasets derived from the CLU. Next, the CLUs that fell within the study area were selected and exported to new shapefiles. This created a library of only CLUs that fell in the study area, even if they were distributed with a county outside of the study area.

2.7 Common Land Unit based analysis

To create tabular land cover data, the library of Midwest CLUs was iterated over and the ArcGIS Zonal Statistics tool was used to determine the majority land cover of each CLU in each year. The land cover in each year was exported to a tabular format for analysis in Pandas.

Within Pandas, corn and soy expansion was quantified by first re-classifying all double crop classifications that contained corn or soy to corn and soy. Next, all CLUs that were classified as anything other than corn or soy in the first comparison year, but were classified as corn or soy in the second comparison year, were selected and exported. The process was reversed to identify the CLUs that had transition from corn or soy to anything else. Both of these subsets were aggregated to the county level using the 'groupby' operation. Lastly, the expansion and loss were summed for each county, providing the net change in corn and soy area in each county between the comparison years, which can be directly compared to NASS data. These steps were repeated for different combinations of crops and comparison years.

2.8 Cropland Data Layer raster based analysis

To calculate corn and soy expansion, the input land cover rasters were first reclassified into binary (corn-soy or not corn-soy) rasters. While it is possible to calculate net change by simply comparing the totals on a county level between two years, such a method does not allow for either the identification of the specific transitions that are occurring or the application of raster cleaning techniques.

The ArcGIS Combine tool was used to create a change raster that could identify land transitioning both to and from corn and soy between comparison years. The change raster values were aggregated to the county level using the ArcGIS Zonal Histogram tool.

As a means of reducing noise and restricting the analysis to only realistically sized changes, Wright and Wimberley and others have applied filters to change rasters. To test the influence of this method, the ArcGIS Focal Statistics tool was used to calculate the majority value of the change raster in a 5 by 5 cell moving window. The filtered change raster values were similarly aggregated to the county level.

Prior land cover of corn and soy expansion was identified with the ArcGIS Zonal Histogram tool by using the change raster to define the zones and the CDLs to identify the land cover. Net expansion was calculate for all crops by subtracting the quantity of land that expanded onto corn and soy on a crop by crop basis.

2.9 NASS data preparation

Using the NASS quickstats tool, county and state level census and survey data for major commodity crops and for pasture were downloaded. Except for hay and alfalfa, which are only reported as harvest area, the planted area attribute was used. Planted area was preferred over harvested area because the CDL has no mechanism for identifying failed crops. Using harvested area has the potential to underestimate agricultural expansion because, for example, the 2012 drought would reduce the harvested area more than in a typical year. Unfortunately this prevented the use of census data for many crops, because the census only reports harvested area.

Using a similar approach as a study that compared NASS agricultural area estimates to those of the NLCD, survey data for corn and soy were compiled at the county level and joined to a shapefile for comparison with other estimates and visualization (Maxwell et al. 2008). Counties that did not have estimates for both crops in all years of interest were excluded. State level estimates were downloaded separately rather than summing county estimates to minimize the effect of data withheld for privacy reasons.

3. Results and Discussion

3.1 Prior land cover of new corn and soy land

A simple intersection of the corn and soy change raster cells with the 2007 CDL reveals the distribution of prior land covers for corn and soy expansion (Figure 4). While this method shows alarming amounts of land-use change, it is also vulnerable to the many anomalies in CDL data. To put these changes in context, the net change from 2007 to 2012 in corn and soy area according to the unmodified CDL was 8.8 million ha, while NASS reports only 5.4 million. Despite re-classification and re-sampling to make the CDL and NASS derived estimates comparable, the change estimates for the most accurately detected crops, vary by 3.4 million ha, or 63% of the NASS reported change.

The CDL documentation cautions against comparing area estimates derived from cell counting to those reported by NASS due to a downward bias for major land covers reported in several studies (Gallego 2004; Gallego et al. 2008; USDA-NASS 2014a). This bias results from smaller features being generalized under moderate resolution imagery (Czaplewski 1992). However, unless otherwise noted, all comparisons presented here are of change in area. Any bias from cell counting should be equal in all inputs, thus minimizing differences from NASS change estimates.

3.2 Comparison of prior land cover by treatment

Between 2007 and 2012 corn and soy expanded primarily onto grass/pasture and wheat (Figure 5). There was generally strong agreement on the quantity and distribution of prior land covers of new corn and soy land between the majority filter and CLU, with the exception of the open space developed category. This category consists mostly of roads, which were poorly detected and aligned in early 56m data. Both the majority filter and CLU treatments were effective at avoiding these errors. The majority filter cleaning method detected one million fewer hectares of net grassland to corn and soy conversion than the unmodified CDL and the CLU. Given the unmodified CDL's tendency to overstate changes, this could indicate that mixed CLUs are being erroneously classified as all corn. Of course, the majority filter has a tendency to aggressively remove change pixels, so the truth very well may lie between these points. Wheat losses were about a half to a third as large as those of grassland. Concerns of grassland loss due to corn and soy expansion are well founded, but other crops are absorbing some of the expansion as well.

The distribution of prior land covers differed in the 2011 to 2013 comparison and the quantity of change was an order of magnitude lower, but was still dominated by grass/pasture and wheat (Figure 6). More crops expanded onto corn and soy land as the prices declined, but wheat and grass/pasture still saw losses to corn and soy. Notably, the unmodified CDL detected about half as much grassland to corn and soy conversion as the two cleaned treatments in the 2011 to 2013 comparison. It is possible that cleaning methods that were applicable to comparisons of older versions of the CDL may introduce unnecessary changes in later versions.

3.3 Comparison to NASS data

Direct comparisons of 2007 and 2012 versions of the CDL resulted in large errors. Cleaning methods were effective at reducing those errors, though they may not be applicable in all cases. In 2007 to 2012 comparisons, the change in corn and soy area was best estimated by the CLU filter, while this method performed the worst in comparisons between 2011 and 2013 (Figure 7). The difference between the unmodified and cleaned estimates was much larger in the 2007 to 2012 treatment than in 2011 to 2013, indicating that the cleaning techniques play a more important role when working with older versions of the CDL. In the 2007 to 2012 treatment, the difference between the majority filter and the CLU was minor, but both offered a large improvement over no cleaning. Even after cleaning, users tracking corn and soy between these two years have over a million hectares of illusory land-use change to contend with. Errors in the 2011 to 2013 treatment were smaller, but the amount of land-use change was as well. On a percentage basis, performance between comparison year groups was similar for the majority filter, better for no cleaning, and much worse for the CLU (Figure 8). The variability in response to cleaning methods reflects temporal variability in the input data.

Error assessments that depend on large area aggregation of change estimates are especially vulnerable to spatial variability in accuracy. Overestimates in one area can cancel underestimates in another, leading to incorrect distributions of land-use change, but correct area change estimates. To mitigate this, study area wide estimates are broken down to state and county aggregates to determine if the trends remain. This is also useful for identifying reliable subsets to study further or hot spots for illusory land-use change. For example, looking at the state level errors for 2011 to 2013 reveals that direction and magnitude of errors can vary widely by state (Figure 9).

The state level variation poses a challenge for simple cleaning methods, and as demonstrated here, even more sophisticated datasets such as the CLU. The tendency for these methods to increase or decrease area estimates uniformly across the landscape does little to address spatial variability. A case by case assessment of the performance of a cleaning method is necessary for determining which to use in a given geography.

3.4 Sources of variability

The biggest driver of errors in corn and soy area estimation is temporal variability. Though one would not expect the CDL estimate to have perfect agreement with NASS estimates, there is a temporal trend to the size of the error. The size of the underestimate decreases with later years (Figure 10). This improvement in area estimation has the unfortunate side-effect of artificially inflating corn and soy expansion estimates between strongly underestimating years and more accurate years by millions of hectares. This land-use change is indistinguishable from actual land-use change happening at the same time and in some cases is much larger.

A possible explanation for this jump is the change in resolution from 56m to 30m, which happened at the same time. However, re-sampling the later imagery to 56m has little effect on the total area of corn and soy. It could also be explained simply by better classification techniques and imagery sources.

Spatial variability also plays an important role, but it is more difficult to quantify. Even for the relatively accurately classified corn and soy, very few counties are able to estimate the change in area to within 1,000 hectares of NASS estimates (Figure 11 Figure 16). Congruent with the total area estimate errors discussed earlier, comparisons between 2007 and 2012 tend to overestimate the amount of change, while comparisons between 2011 and 2013 underestimate it. However, there is clear spatial variability at the county level. Cleaning methods had little influence on the spatial distribution of the errors, and only marginally reduce the magnitude across the landscape.

Interestingly, several counties in eastern North and South Dakota had large underestimates of corn and soy expansion under all years and treatments. This is especially striking because other studies that have used the CDL have highlighted this as an area of corn and soy expansion, but this indicates even those estimates are too low. It is also worth noting that southeastern Minnesota, an area being studied for the rapid expansion of corn and soy demonstrated by the CDL, is an area of overestimation in all treatments.

Comparison at the county level of majority filter and CLU cleaning methods show no noticeable improvements with the CLU. Given the restrictions on the CLU, majority filters are an appealing alternative. However, accurate area estimates from a majority filter at the aggregate level does not mean the underlying spatial arrangement is correct. The aggressive filter may be reducing the size of some fields, and not removing erroneous ones elsewhere (0). For applications that require precise assessment of fields, the CLU may be a better choice.

3.5 Replication of Wright and Wimberley

Wright and Wimberley's study is a useful case study of the strengths and weaknesses of the CDL because it asks a simple yet important question of the data, but its results are very sensitive to variability in the CDL. While the previously discussed sensitivity analysis indicates that Wright and Wimberley may have under estimated the amount of grassland to corn and soy conversion (Figure 3), examining NASS data suggests the opposite. Although corn and soy area definitely increased between 2006 and 2011 in their study area, the area devoted to all principal field crops actually declined by 900,000 hectares (USDA-NASS 2014b).

Unfortunately, the net grassland to corn and soy conversion metric employed by Wright and Wimberley is not comparable to NASS data. The most that can be done is to put the net grassland to corn and soy estimates in the context of the discrepancy between CDL and NASS corn and soy expansion estimates using their cleaning methods (**Error! eference source not found.**). In North and South Dakota, reportedly hot spots for change, the CDL's overestimate of corn and soy expansion eclipses the reported grassland to corn and soy conversion. Nebraska contends with the opposite problem, only a small amount of conversion is reported in the face of a large underestimate of corn and soy expansions in the state. Although corn and soy expansion is not directly comparable to Wright and Wimberley's net grassland to corn and soy conversion metric, the magnitude of the errors relative to the reported conversion calls into question the reliability of the results.

4. Conclusions

Corn and soy expansion is happening at a rapid rate in the Midwest, and the CDL indicates that the majority of that expansion is happening at the expense of grass, pasture, hay, and wheat land. This pattern was apparent under all cleaning methods and year combinations compared here. According to the CDL, between 2007 and 2012 approximately five million hectares of corn and soy expanded onto the grass, pasture, hay, and wheat. However, over the same time period the CDL overestimated the amount of corn and soy expansion by 3.5 million hectares in the unmodified treatment and 1.5 million with cleaning methods applied.

Variability in the CDL can reveal either staggering land-use change trends or overwhelming uncertainty when in reality there is some of both. The environmental and policy questions being answered with the CDL cannot wait for perfect data. While any analysis should examine and respond to variability specific to the question at hand, there are a few guidelines for CDL based analysis that can be gleaned from this research.

First, while of intense interest for both policy and environmental questions, studying land-use change with the CDL can be unreliable. If the research question can be answered with NASS data, then a great deal of uncertainty can be avoided by forgoing the higher resolution of the CDL. If the CDL must be used, comparisons between pre and post 2010 versions of the CDL should be avoided. Improvements in the CDL's accuracy make it difficult to separate actual land-use change from improvements in classification.

The CDL's inclusion of non-agricultural classifications is tempting, but they should be used cautiously. Their classification is dependent on many assumptions and their accuracy is unreported. Given the variability with the most accurately classified crops, it is difficult to be confident in conclusions involving specific non-agricultural land covers.

Lastly, imperfect assumptions are impossible to avoid when working with CDL data. In the absence of a reliable data source to compare against, performing a sensitivity analysis on those assumptions can reveal potential pitfalls and add confidence to an assumption.

The discussion and assessment of variability in CDL data quality presented here should not deter people from using the CDL. The frequency, resolution, extent, and variety of land covers tracked are unparalleled by other data sources and provide an invaluable resource for understanding the consequences of a dynamic agricultural landscape. However, these impressive characteristics and high classification accuracies can inspire analyses that sometimes go beyond the capability of the data. The land-use change trends we are most interested in are often on the scale of thousands to tens of thousands of hectares, but error propagation from spatial and temporal variability introduce uncertainty that can quickly overwhelm the results.

5. Tables

Table 1. Reclassification crosswalk

CDL Classification	New Classification	CDL Code	New Code	CDL Classification	New Classification	CDL Code	New Code
Background	No Data	0	0	Clouds/No Data	No Data	81	0
Corn	Corn	1	1	Developed	Developed	82	122
Cotton	Other Agriculture	2	2	Water	Water	83	111
Rice	Other Agriculture	3	2	Wetlands	Other Natural	87	141
Sorghum	Sorghum	4	4	Nonag/Undefined	No Data	88	0
Soybeans	Soybeans	5	5	Aquaculture	Water	92	111
Sunflower	Sunflower	6	6	Open Water	Water	111	111
Peanuts	Other Agriculture	10	2	Perennial Ice/Snow	Water	112	111
Tobacco	Other Agriculture	11	2	Developed/Open Space	Open Space Developed	121	121
Sweet Corn	Other Agriculture	12	2	Developed/Low Intensity	Developed	122	122
Pop or Orn Corn	Other Agriculture	13	2	Developed/Med Intensity	Developed	123	122
Mint	Other Agriculture	14	2	Developed/High Intensity	Developed	124	122
Barley	Barley	21	21	Barren	Developed	131	122
Durum Wheat	Wheat	22	22	Deciduous Forest	Other Natural	141	141
Spring Wheat	Wheat	23	22	Evergreen Forest	Other Natural	142	141
Winter Wheat	Wheat	24	22	Mixed Forest	Other Natural	143	141
Other Small Grains	Other Agriculture	25	2	Shrubland	Other Natural	152	141
Dbl Crop WinWht/Soybeans	Double Wheat-Soybeans	26	26	Grassland/Pasture	Grass/Pasture/Hay	176	176

Rye	Other Agriculture	27	2	Woody Wetlands	Other Natural	190	141
Oats	Other Agriculture	28	2	Herbaceous Wetlands	Other Natural	195	141
Millet	Other Agriculture	29	2	Pistachios	Other Agriculture	204	2
Speltz	Other Agriculture	30	2	Triticale	Other Agriculture	205	2
Canola	Canola	31	31	Carrots	Other Agriculture	206	2
Flaxseed	Other Agriculture	32	2	Asparagus	Other Agriculture	207	2
Safflower	Other Agriculture	33	2	Garlic	Other Agriculture	208	2
Rape Seed	Other Agriculture	34	2	Cantaloupes	Other Agriculture	209	2
Mustard	Other Agriculture	35	2	Prunes	Other Agriculture	210	2
Alfalfa	Alfalfa	36	36	Olives	Other Agriculture	211	2
Other Hay/Non Alfalfa	Grass/Pasture/Hay	37	176	Oranges	Other Agriculture	212	2
Camelina	Other Agriculture	38	2	Honeydew Melons	Other Agriculture	213	2
Buckwheat	Other Agriculture	39	2	Broccoli	Other Agriculture	214	2
Sugarbeets	Sugarbeets	41	41	Peppers	Other Agriculture	216	2
Dry Beans	Beans	42	42	Pomegranates	Other Agriculture	217	2
Potatoes	Other Agriculture	43	2	Nectarines	Other Agriculture	218	2
Other Crops	Other Agriculture	44	2	Greens	Other Agriculture	219	2
Sugarcane	Other Agriculture	45	2	Plums	Other Agriculture	220	2
Sweet Potatoes	Other Agriculture	46	2	Strawberries	Other Agriculture	221	2
Misc Vegs & Fruits	Other Agriculture	47	2	Squash	Other Agriculture	222	2
Watermelons	Other Agriculture	48	2	Apricots	Other Agriculture	223	2
Onions	Other Agriculture	49	2	Vetch	Other Agriculture	224	2
Cucumbers	Other Agriculture	50	2	Dbl Crop WinWht/Corn	Double Wheat-Corn	225	225
Chick Peas	Other Agriculture	51	2	Dbl Crop Oats/Corn	Double Oats-Corn	226	226
Lentils	Other Agriculture	52	2	Lettuce	Other Agriculture	227	2
Peas	Other Agriculture	53	2	Pumpkins	Other Agriculture	229	2

Tomatoes	Other Agriculture	54	2	Dbl Crop Lettuce/Durum Wht	Wheat	230	22
Caneberries	Other Agriculture	55	2	Dbl Crop Lettuce/Cantaloupe	Other Agriculture	231	2
Hops	Other Agriculture	56	2	Dbl Crop Lettuce/Cotton	Other Agriculture	232	2
Herbs	Other Agriculture	57	2	Dbl Crop Lettuce/Barley	Barley	233	21
Clover/Wildflowers	Other Agriculture	58	2	Dbl Crop Durum Wht/Sorghum	Double Wheat-Sorghum	234	234
Sod/Grass Seed	Other Agriculture	59	2	Dbl Crop Barley/Sorghum	Double Barley-Sorghum	235	235
Switchgrass	Other Agriculture	60	2	Dbl Crop WinWht/Sorghum	Double Wheat-Sorghum	236	234
Fallow/Idle Cropland	Fallow/Idle	61	61	Dbl Crop Barley/Corn	Double Barley-Corn	237	237
Forest	Other Natural	63	141	Dbl Crop WinWht/Cotton	Wheat	238	22
Shrubland	Other Natural	64	141	Dbl Crop Soybeans/Cotton	Soybeans	239	5
Barren	Developed	65	122	Dbl Crop Soybeans/Oats	Double Soybeans-Oats	240	240
Cherries	Other Agriculture	66	2	Dbl Crop Corn/Soybeans	Double Corn-Soybeans	241	241
Peaches	Other Agriculture	67	2	Blueberries	Other Agriculture	242	2
Apples	Other Agriculture	68	2	Cabbage	Other Agriculture	243	2
Grapes	Other Agriculture	69	2	Cauliflower	Other Agriculture	244	2
Christmas Trees	Other Agriculture	70	2	Celery	Other Agriculture	245	2
Other Tree Crops	Other Agriculture	71	2	Radishes	Other Agriculture	246	2
Citrus	Other Agriculture	72	2	Turnips	Other Agriculture	247	2
Pecans	Other Agriculture	74	2	Eggplants	Other Agriculture	248	2
Almonds	Other Agriculture	75	2	Gourds	Other Agriculture	249	2
Walnuts	Other Agriculture	76	2	Cranberries	Other Agriculture	250	2
Pears	Other Agriculture	77	2	Dbl Crop Barley/Soybeans	Double Barley-Soybeans	254	254

6. Figures

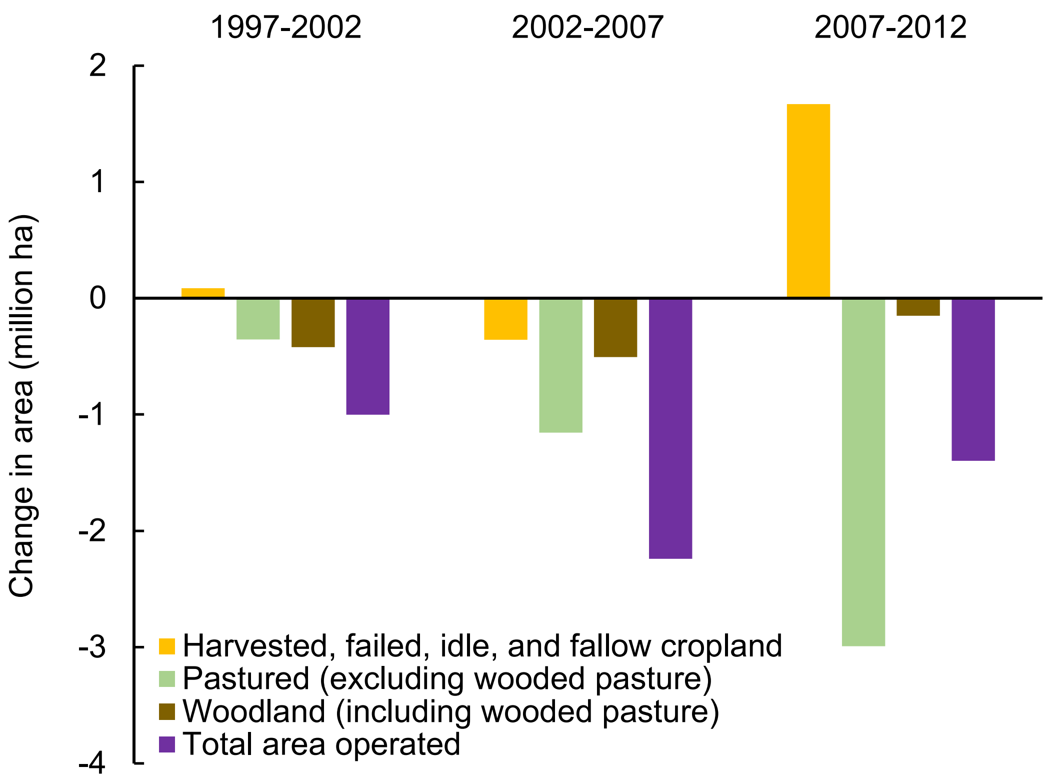

Figure 1 Changes in area of different types of agricultural land between NASS censuses in the Midwest.

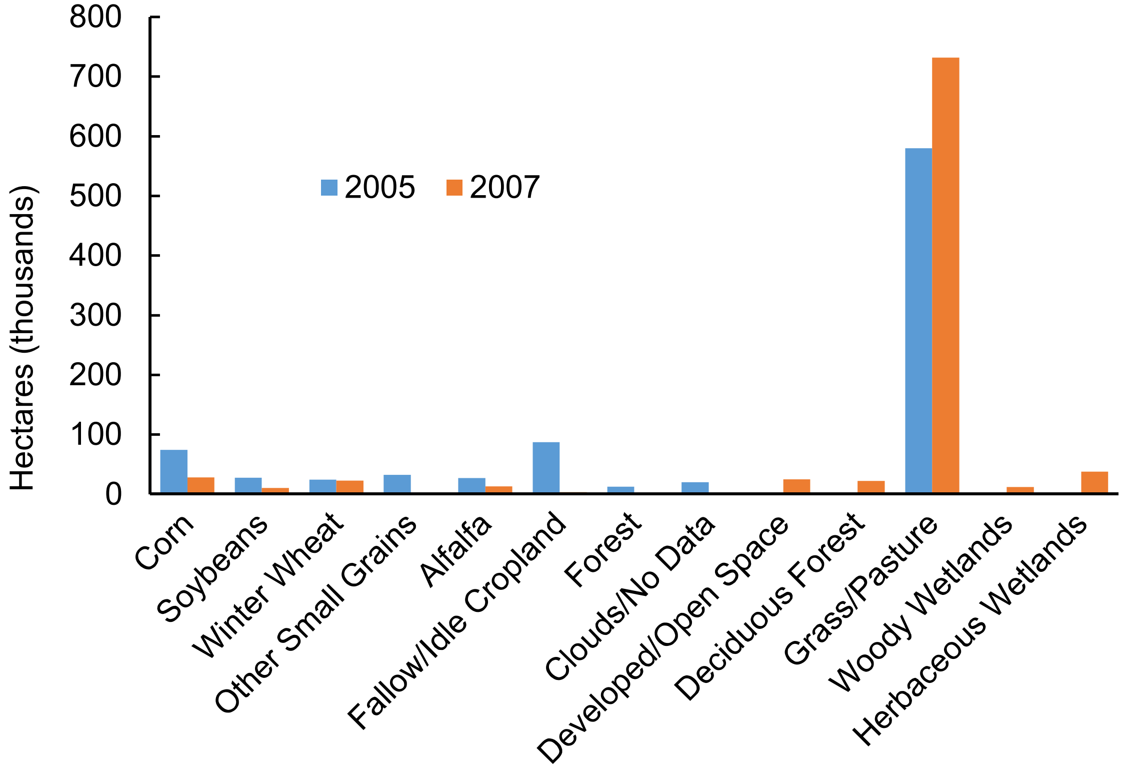

Figure 2 Classification of 2006 Nebraska ‘Non-ag/undefined’ cells in 2005 and 2007. Only transitions larger than 10,000 ha are shown here.

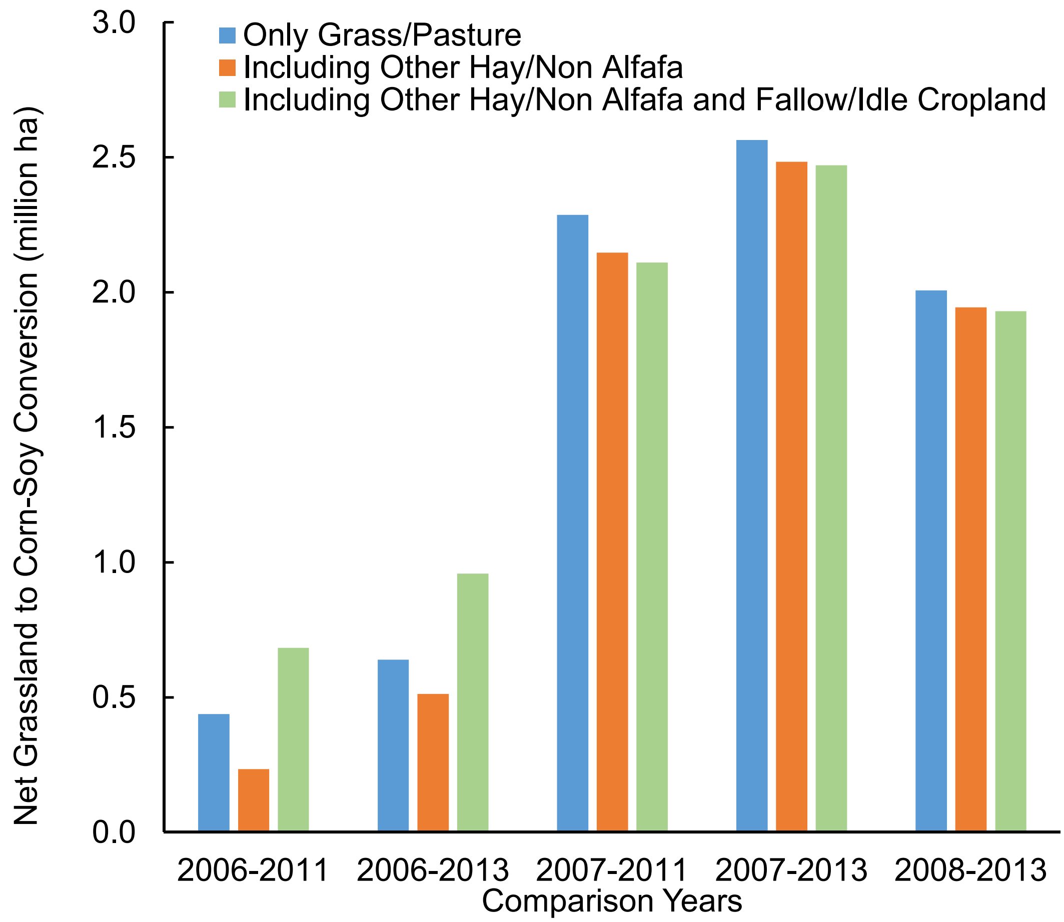

Figure 3 Influence of grassland definition and comparison years on net grassland to corn and soy conversion detected in IA, MN, ND, NE, and SD.

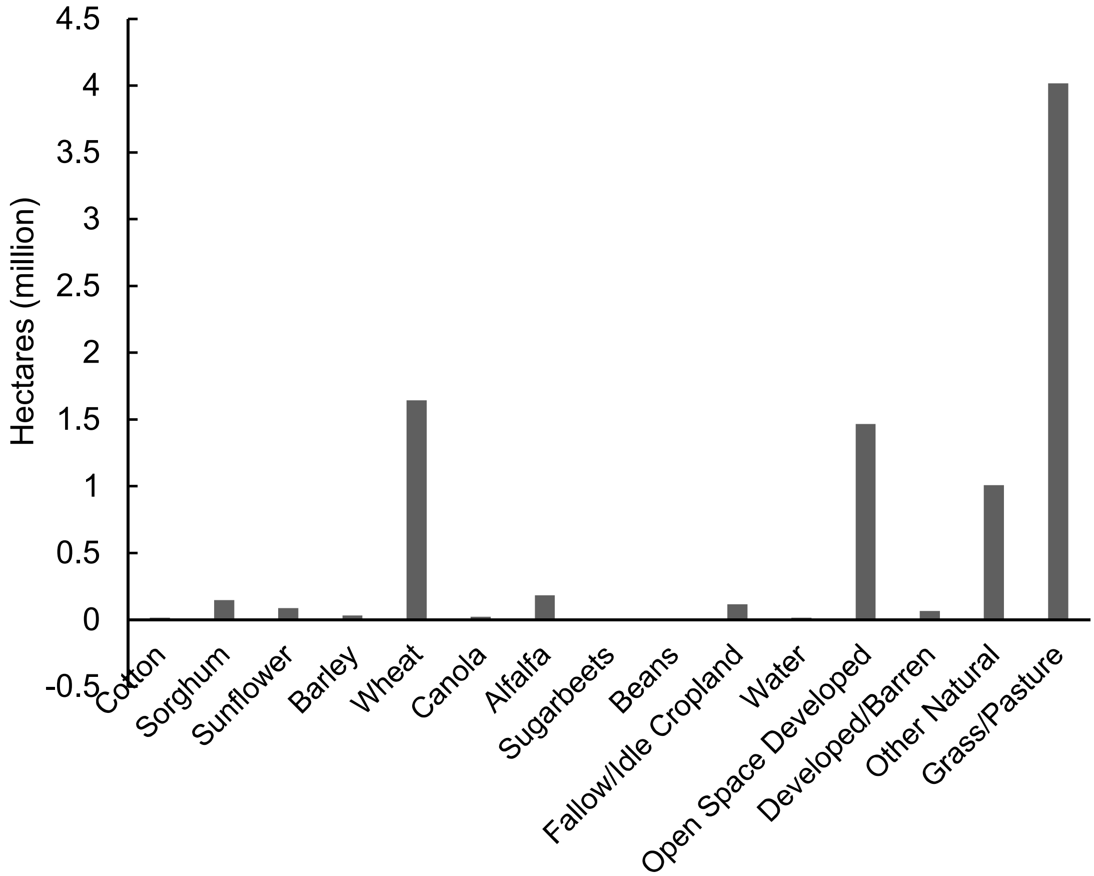

Figure 4 Land cover net loss to corn and soy in the Midwest from 2007 to 2012 with unmodified CDL.

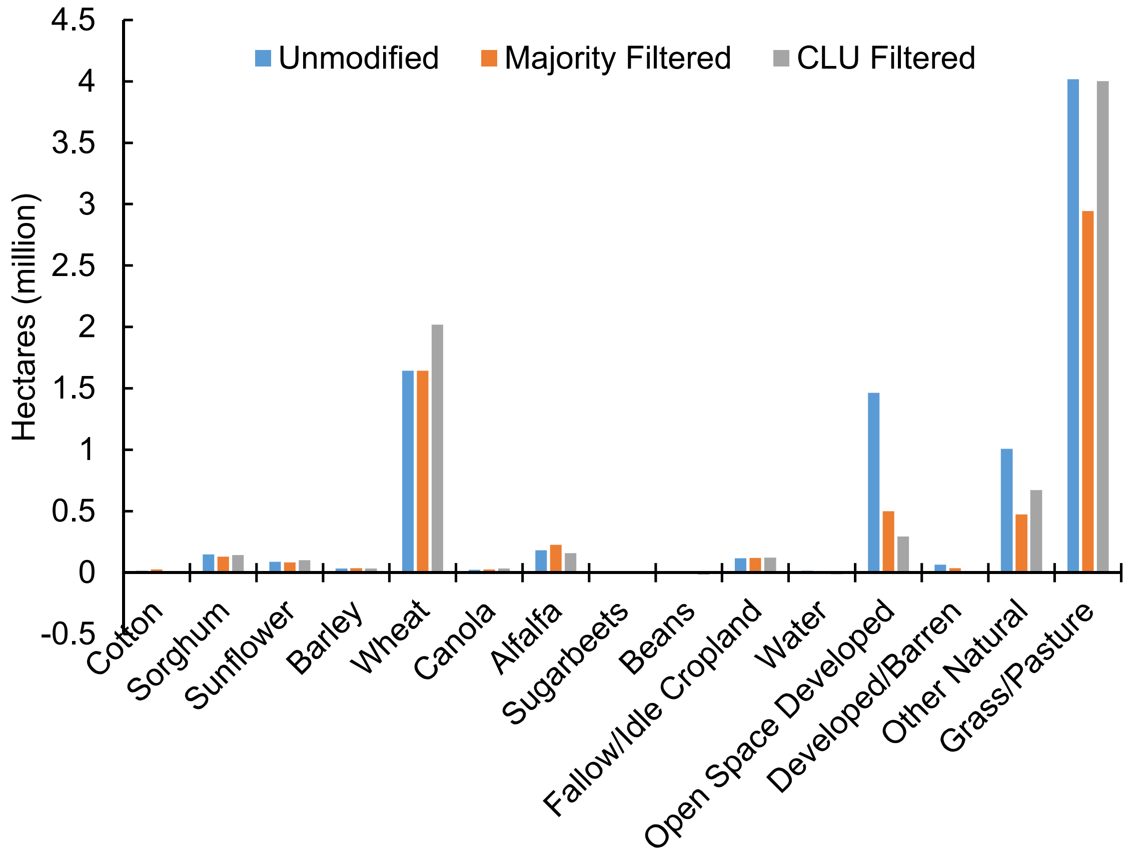

Figure 5 Prior land covers of net corn and soy expansion in the Midwest 2007 to 2012.

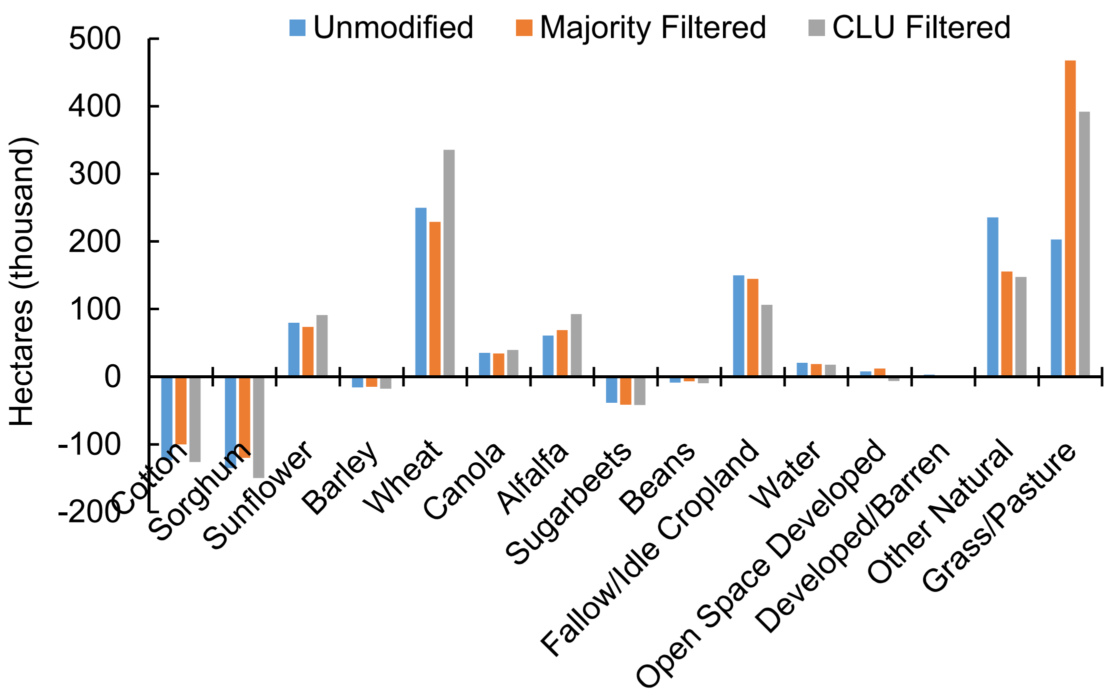

Figure 6 Prior land covers of net corn and soy expansion in the Midwest 2011 to 2013.

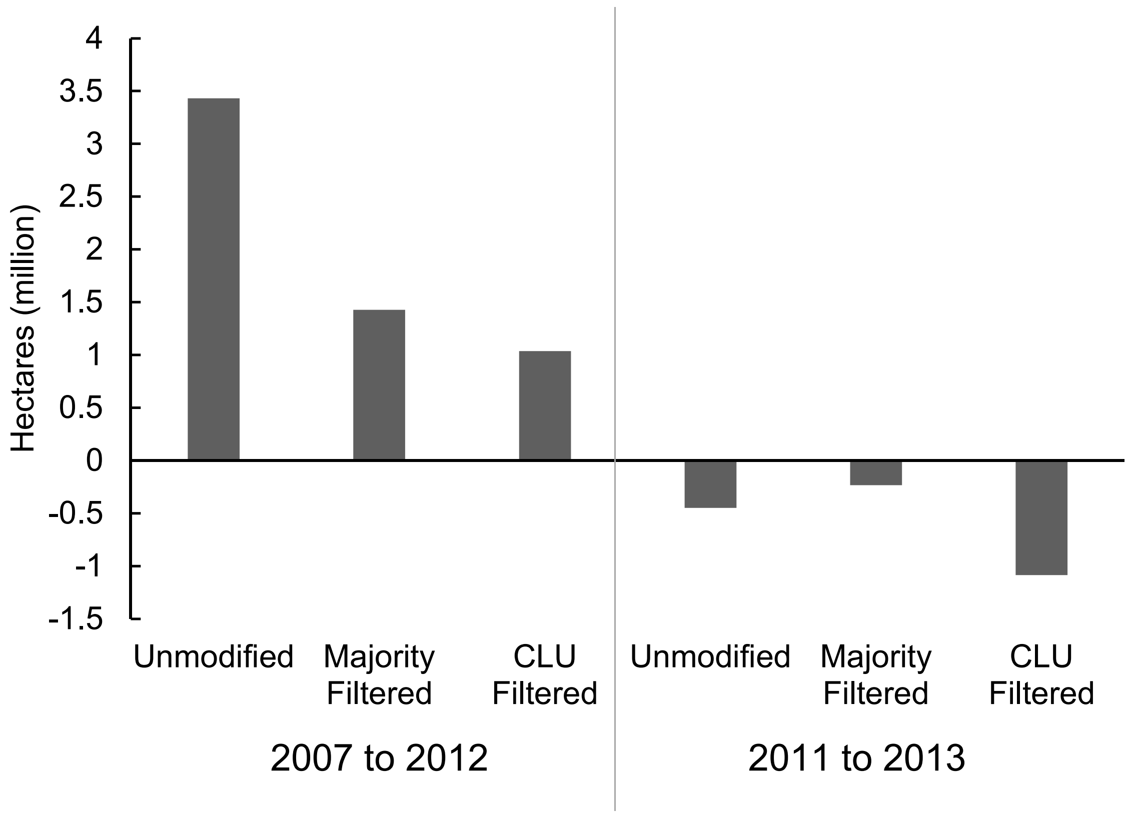

Figure 7 Absolute error between CDL-derived estimates of corn and soy expansion and NASS estimates in the Midwest under two cleaning methods and year combinations.

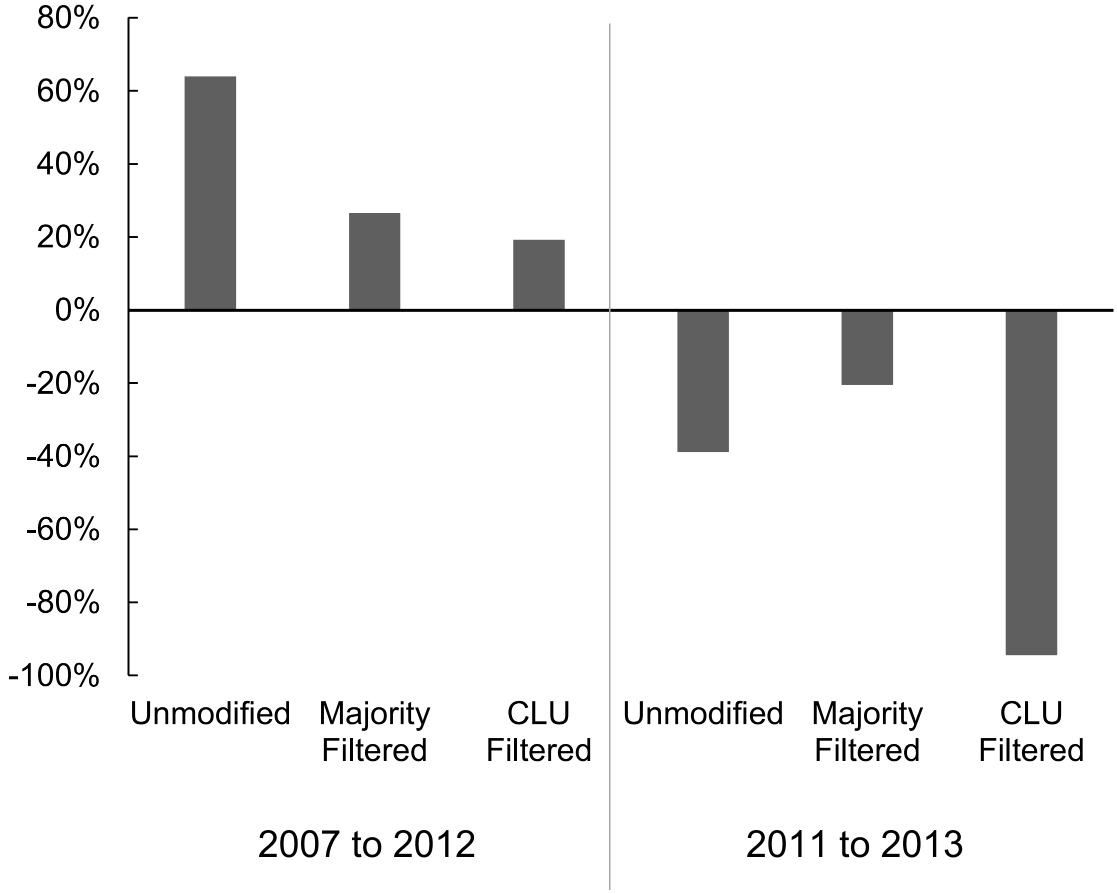

Figure 8 Relative error between CDL-derived estimates of corn and soy expansion and NASS estimates in the Midwest under two cleaning methods and year combinations.

Figure 9 Error in corn and soy expansion estimates between CDL-derived estimates and NASS on a state basis between 2011 and 2013.

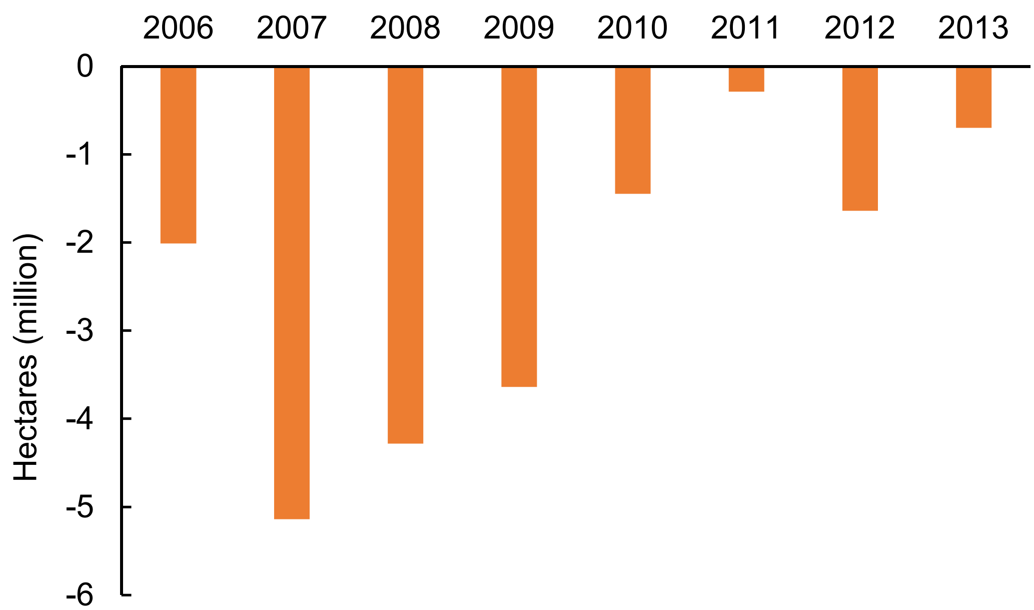

Figure 10 Underestimate of CDL Midwest corn and soy area relative to NASS.

The difference between a later and earlier year's value is the minimum amount of illusory corn and soy expansion in a comparison.

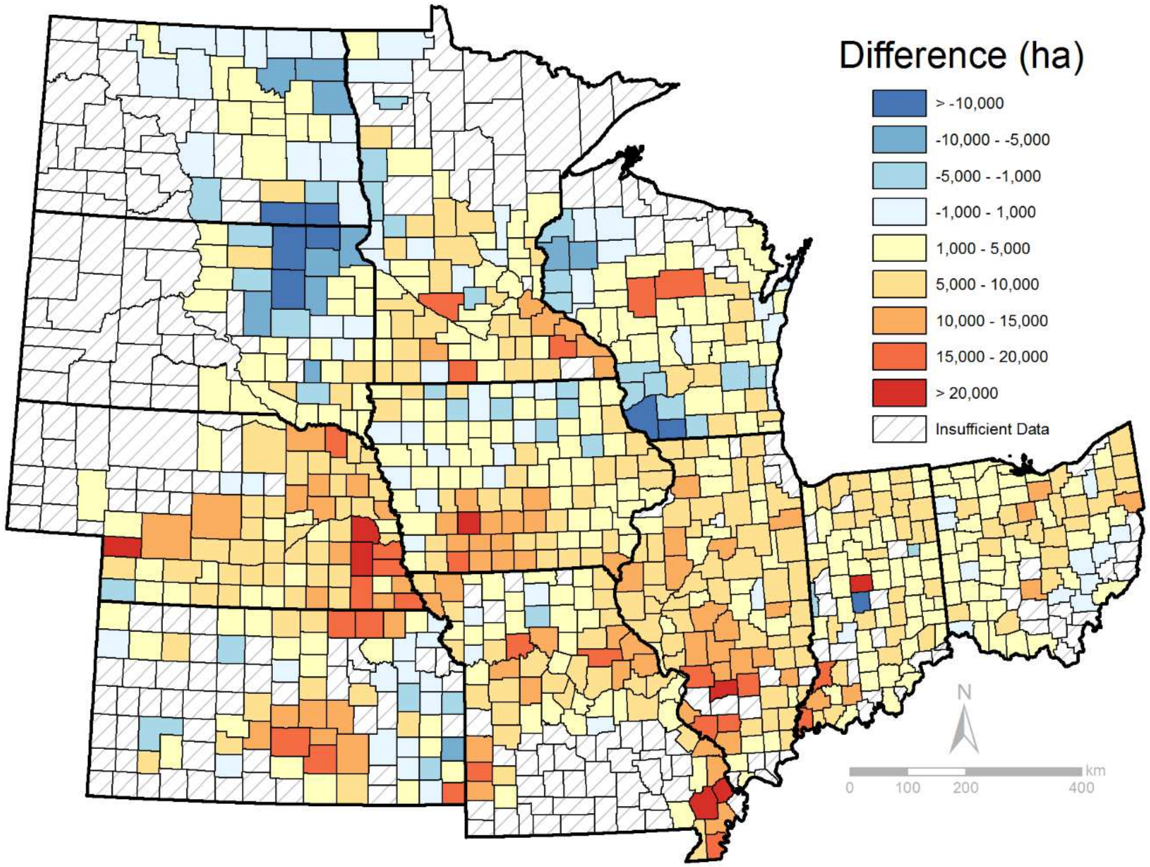

Figure 11 Difference of 2007 to 2012 corn and soy change estimate between NASS and unmodified CDL.

Negative values indicate the CDL derived value underestimated the amount of change relative to the NASS value, while positive indicates an over estimate.

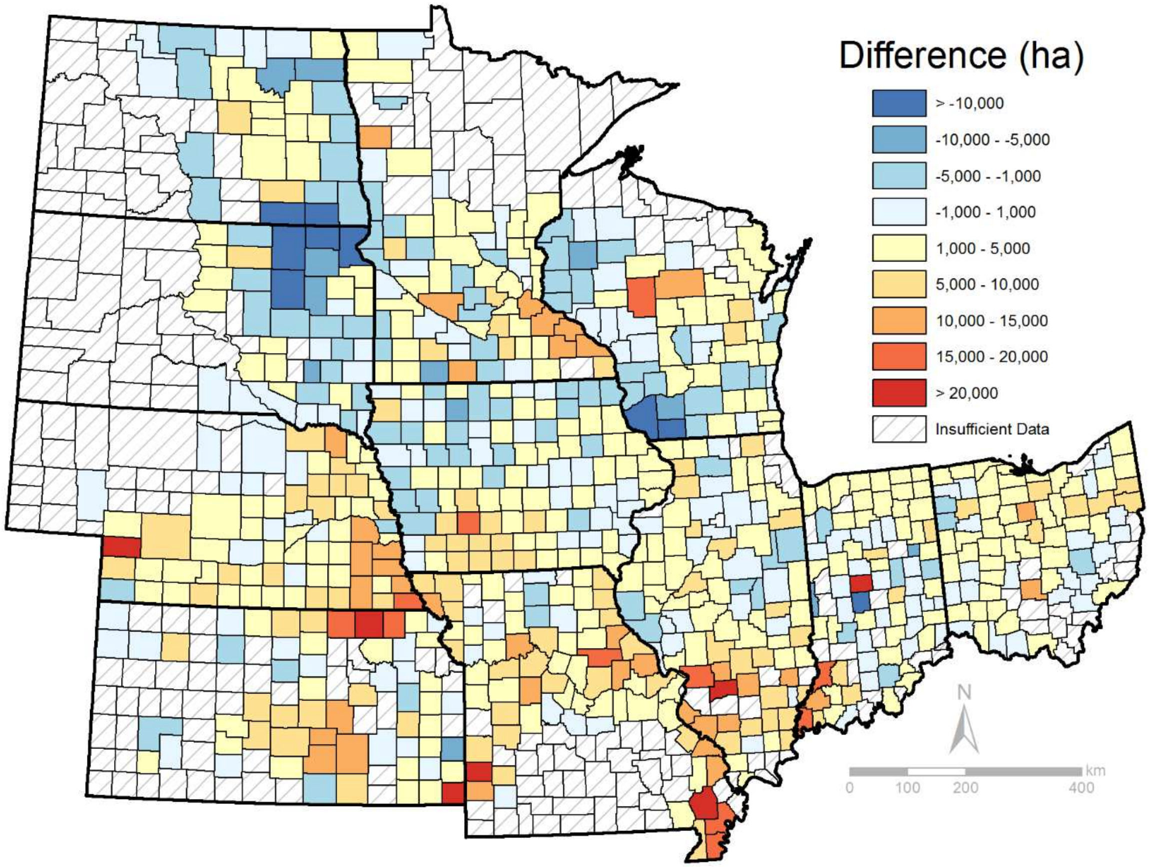

Figure 12 Difference of 2007 to 2012 corn and soy change estimate between NASS and 5x5 majority filtered CDL.

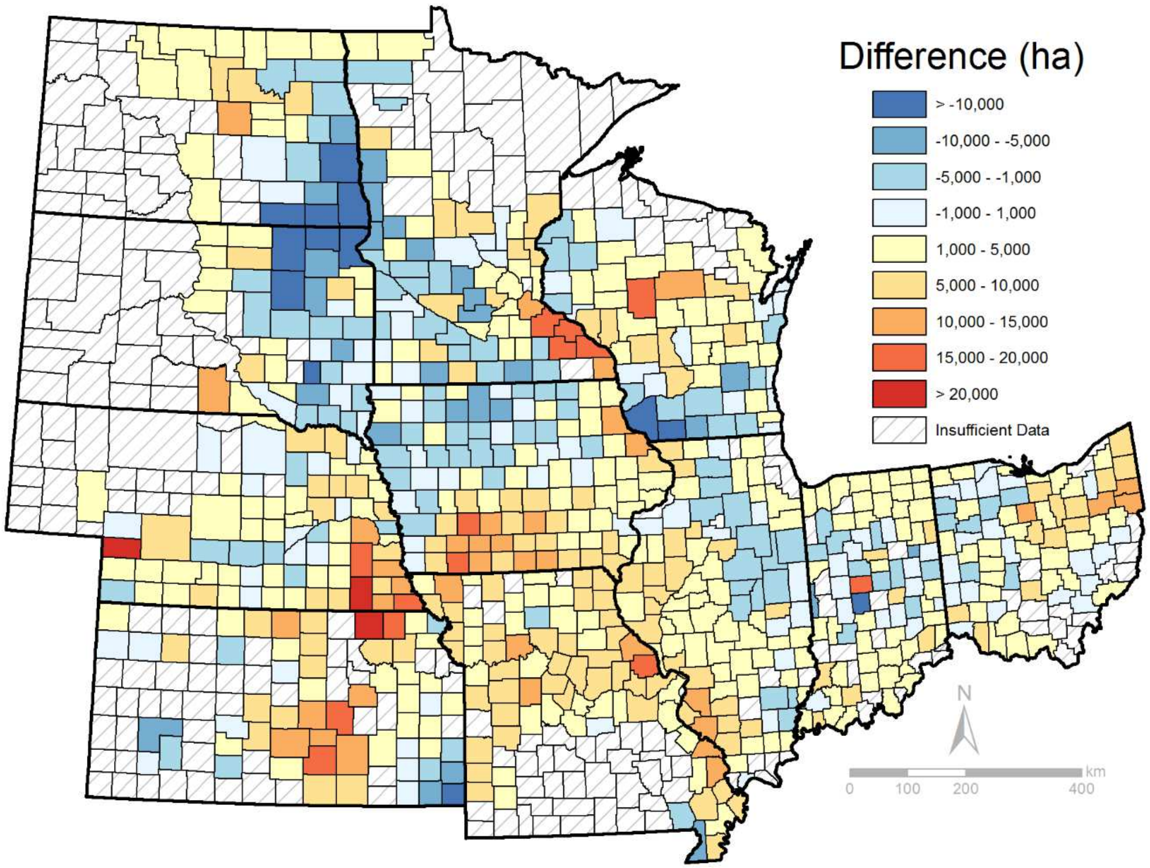

Figure 13 Difference of 2007 to 2012 corn and soy change estimate between NASS and CLU filtered CDL.

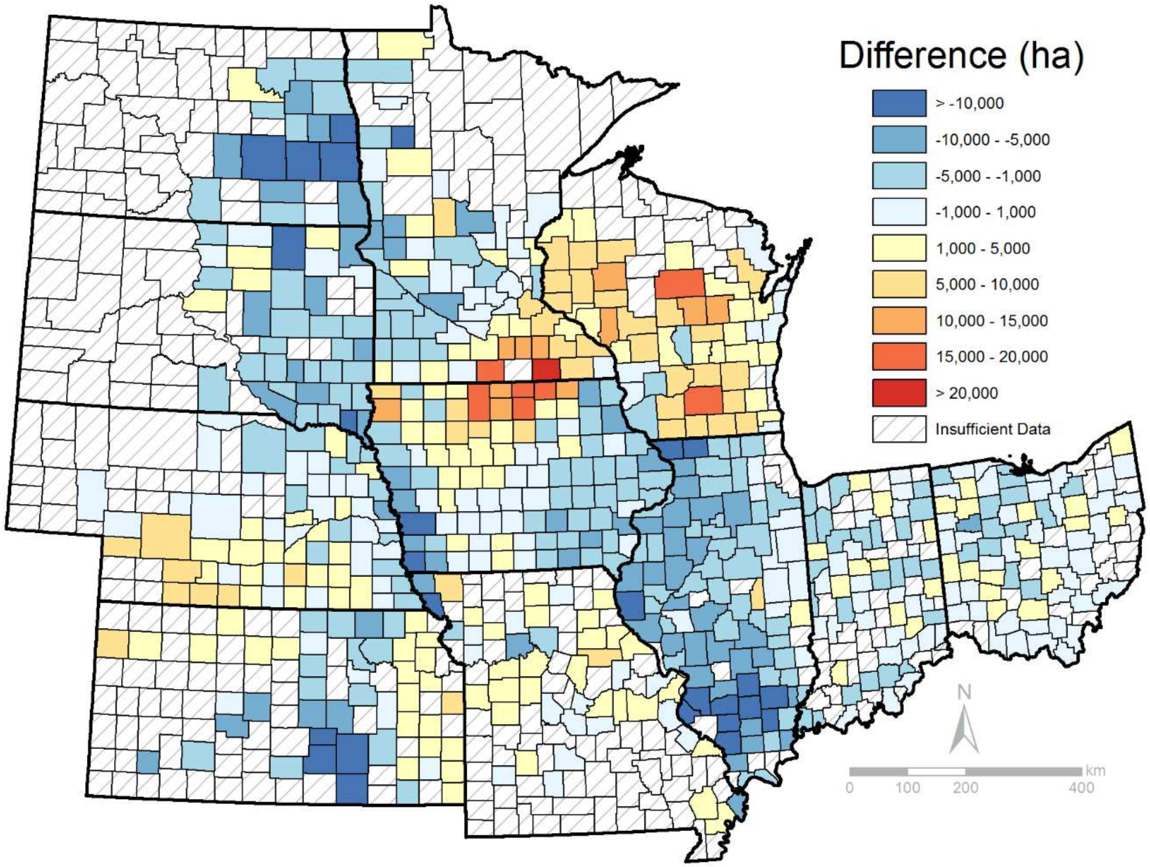

Figure 14 Difference of 2011 to 2013 corn and soy change estimate between NASS and unmodified CDL.

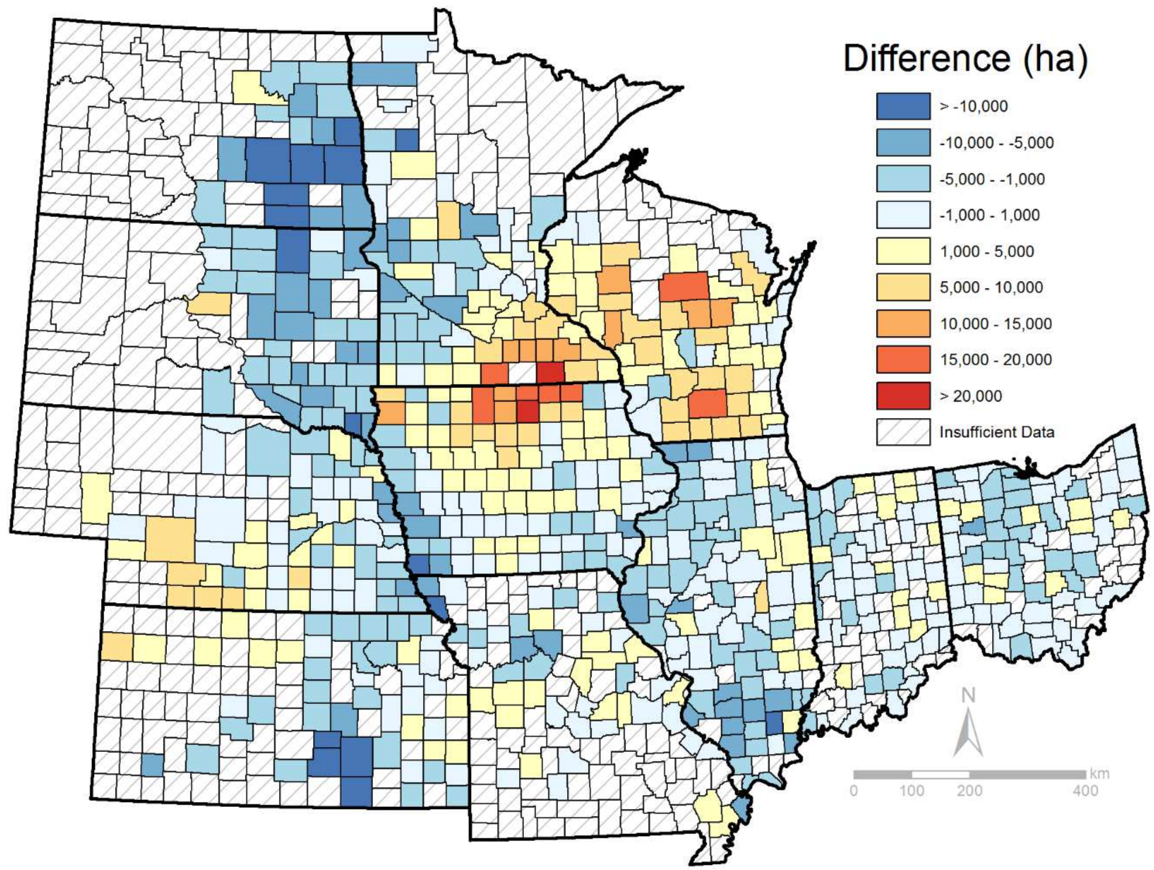

Figure 15 Difference of 2011 to 2013 corn and soy change estimate between NASS and 5x5 majority filtered CDL.

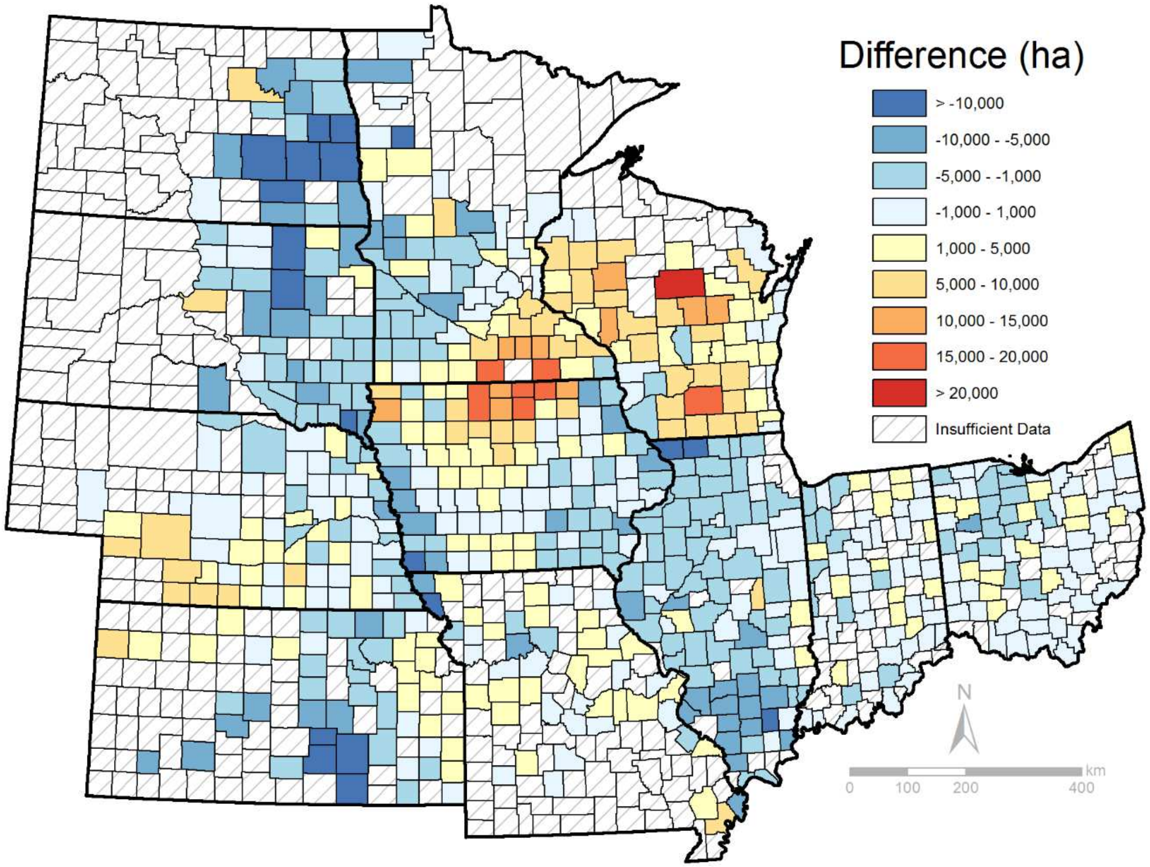

Figure 16 Difference of 2011 to 2013 corn and soy change estimate between NASS and CLU filtered CDL.

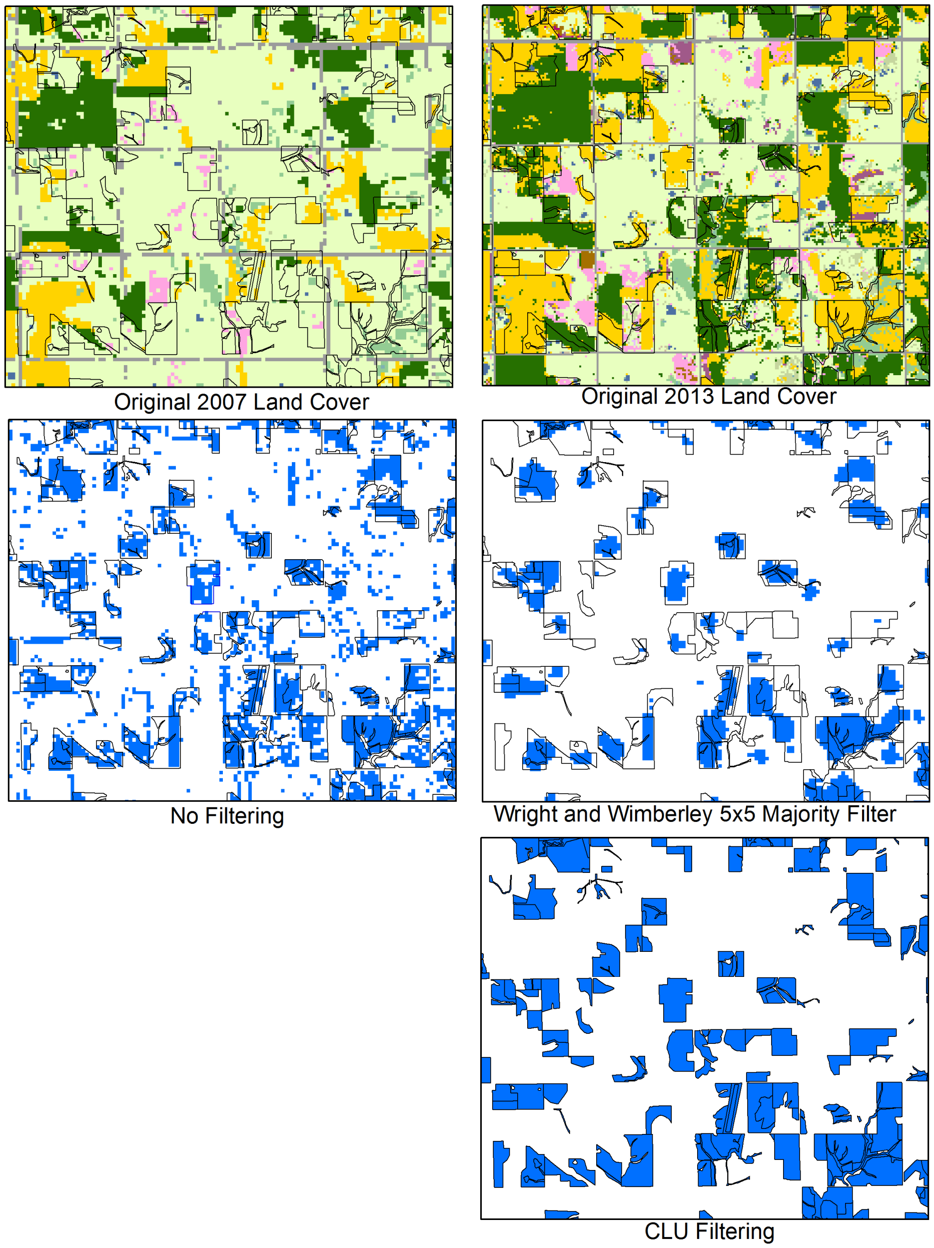

Figure 17 Comparison of no filtering, a 5x5 majority filter, and the CLU.

Black lines represent the boundaries of CLUs that transitioned from grass to corn or soy between 2007 and 2013. Blue represents the same transition under different cleaning methods.

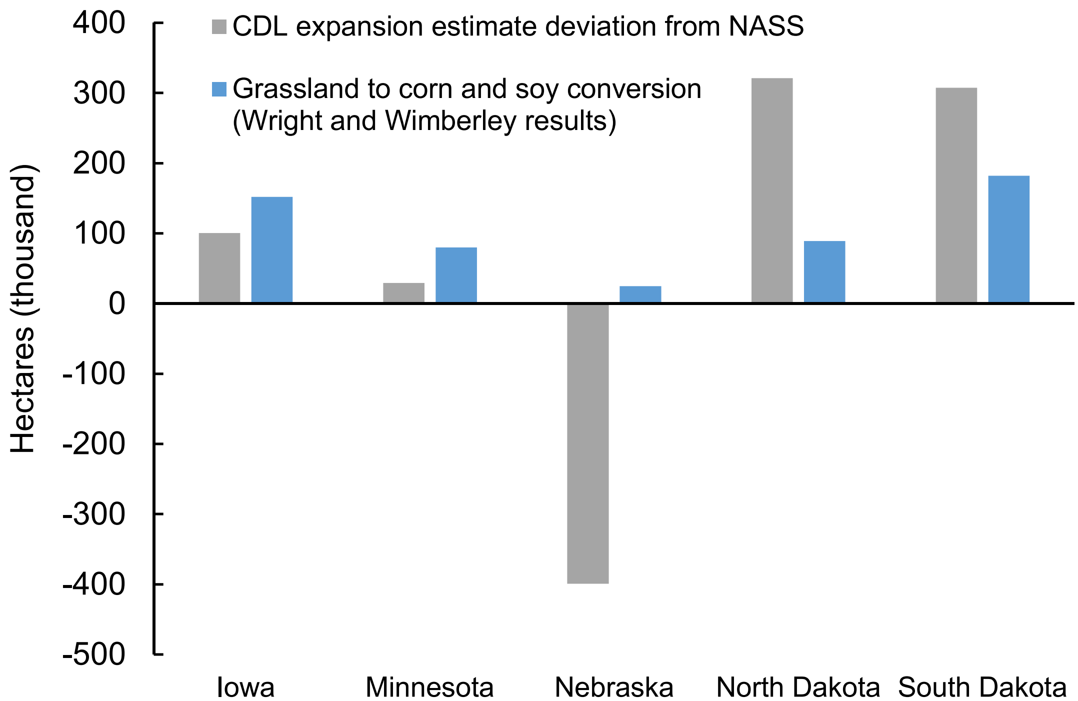

Figure 18 Wright and Wimberley's grassland loss results in the context of the CDL's deviation from NASS data on corn and soy expansion in the Midwest.

7. Literature Cited

Adkins, Z., 2013. Common Land Unit (CLU) Information Sheet. , (June). Available at: http://www.fsa.usda.gov/Internet/FSA_File/clu__infosheet_2013.pdf.

Alexander, R.B. et al., 2008. Differences in Phosphorus and Nitrogen Delivery to The Gulf of Mexico from the Mississippi River Basin. *Environmental Science & Technology*, 42(3), pp.822–830.

Bailey, J. & Boryan, C., 2010. Remote sensing applications in agriculture at the USDA National Agricultural Statistics Service.

Boryan, C. et al., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, 26(5), pp.341–358.

Bowman, R.A., Reeder, J.D. & Lober, R.W., 1990. Changes in soil properties after 3, 20, and 60 years of cultivation. *Soil Science*, 150(6).

Chon, S., Nelson, C.J. & Coutts, J.H., 2003. Physicological assessemtn and path analysis to improve evaluation of alfalfa autotoxicity. , 29(11), pp.2413–2424.

Claassen, R. & Nickerson, C., 2011. An Expansion of U.S. Corn Production, 2000-09. *ERS Economic Information Bulletin*, (79).

Cox, C. & Rundquist, S., 2013. *Going, going, gone! Millions of acres of wetlands and fragile land go under the plow*, Washington, DC.

Czaplewski, R.L., 1992. Misclassification bias in areal estimates. *Photogrammetric Engineering and Remote Sensing*, 58(2), pp.189–192.

Downing, J.A. et al., 1999. *Gulf of Mexico Hypoxia : Land and Sea Interactions*, Council for Agricultural Science and Technology.

Elliott, J. et al., 2014. A Spatial Modeling Framework to Evaluate Domestic Biofuel-Induced Potential Land Use Changes and Emissions.

ESRI, 2013. ArcGIS Desktop: Release 10.2. *Redlands CA*.

Faber, S., Rundquist, S. & Male, T., 2012. *Plowed under: how crop subsidies contribute to massive habitat losses*, Washington, DC.

Fargione, J. et al., 2008. Land clearing and the biofuel carbon debt. *Science*, 319(February), pp.1235–1238.

Fargione, J.E. et al., 2009. Bioenergy and Wildlife: Threats and Opportunities for Grassland Conservation. *BioScience*, 59(9), pp.767–777.

Farm Bureau, 2013. *2013 Multi - State Land Use Study : Estimated Land Use Changes 2007 - 2012*, Gallego, F.J., 2004. Remote sensing and land cover area estimation. *International Journal of Remote Sensing*, 25(15), pp.3019–3047.

Gallego, J.F. et al., 2008. *Best practices for crop area estimation with remote sensing*,

Gebhart, D.L. et al., 1994. The CRP increases soil organic carbon. *Journal of Soil and Water Conservation*, 49, pp.488–492.

Gecan, R., Johansson, R. & FitzGerald, K., 2009. *The Impact of Ethanol Use on Food Prices and Greenhouse-Gas Emissions*,

Gelder, B.K., Cruse, R.M. & Kaleita, a. L., 2008. Automated determination of management units for precision conservation. *Journal of Soil and Water Conservation*, 63(5), pp.273–279.

Guo, L.B. & Gifford, R.M., 2002. Soil carbon stocks and land use change: a meta analysis. *Global Change Biology*, 8, pp.345–360.

Johnston, C. a., 2013. Wetland Losses Due to Row Crop Expansion in the Dakota Prairie Pothole Region. *Wetlands*, 33(1), pp.175–182.

Jonhson, D.M. & Mueller, R., 2009. The 2009 Cropland Data Layer. *Photogrammetric Engineering and Remote Sensing*, pp.1201–1205.

Karcher, S.C. et al., 2013. Alternative Land-Use Method for Spatially Informed Watershed Management Decision Making Using SWAT. , (December), pp.1413–1423.

Li, R., Guan, Q. & Merchant, J., 2012. A geospatial modeling framework for assessing biofuels-related land-use and land-cover change. *Agriculture, Ecosystems & Environment*, 161, pp.17–26.

Long, J. a. et al., 2014. Changes in field-level cropping sequences: Indicators of shifting agricultural practices. *Agriculture, Ecosystems & Environment*, 189, pp.11–20.

Maxwell, S.K., Wood, E.C. & Janus, a., 2008. Comparison of the USGS 2001 NLCD to the 2002 USDA Census of Agriculture for the Upper Midwest United States. *Agriculture, Ecosystems & Environment*, 127(1-2), pp.141–145.

McKinney, W., 2010. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*. pp. 51–56.

Meehan, T.D. et al., 2013. Ecosystem-service tradeoffs associated with switching from annual to perennial energy crops in riparian zones of the US Midwest. *PloS one*, 8(11), p.e80093.

Miller, D., 1983. Allelopathic effects of alfalfa. *Journal of Chemical Ecology*, 9(8).

Mueller, R. & Harris, M., 2013. Reported Uses of CropScape and the National Cropland Data Layer Program. In *Sixth International Conference on Agricultural Statistics*. Rio de Janeiro, Brazil.

Mullins, K. a, Griffin, W.M. & Matthews, H.S., 2011. Policy implications of uncertainty in modeled life-cycle greenhouse gas emissions of biofuels. *Environmental science & technology*, 45(1), pp.132–8.

Plourde, J.D., Pijanowski, B.C. & Pekin, B.K., 2013. Evidence for increased monoculture cropping in the Central United States. *Agriculture, Ecosystems & Environment*, 165, pp.50–59.

Resop, J.P. et al., 2012. Combining explanatory crop models with geospatial data for regional analyses of crop yield using field-scale modeling units. *Computers and Electronics in Agriculture*, 89, pp.51–61.

Stern, A.J., Doraiswamy, P.C. & Raymond Hunt, E., 2012. Changes of crop rotation in Iowa determined from the United States Department of Agriculture, National Agricultural Statistics Service cropland data layer product. *Journal of Applied Remote Sensing*, 6(1), p.063590.

Unger, P., 2001. Total carbon, aggregation, bulk density, and penetration resistance of cropland and nearby grassland soils. *Soil Science*.

USDA Farm Service Agency, 2012. *Common Land Unit*, Washington, D.C.

USDA-NASS, 2012. 2012 Census Full Report - Appendix B. General Explanation and Census of Agriculture Report Form.

USDA-NASS, 2014a. Cropland Data Layer Metadata. Available at: http://www.nass.usda.gov/research/Cropland/metadata/meta.htm.

USDA-NASS, 2014b. NASS QuickStats Ad-hoc Query Tool. Available at: http://quickstats.nass.usda.gov/.

USDA-NASS, 2013. USDA National Agricultural Statistics Service Cropland Data Layer.

Werling, B.P. et al., 2014. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4), pp.1652–7.

Wolfe, A.H. & Patz, J. a, 2002. Reactive nitrogen and human health: acute and long-term implications. *Ambio*, 31(2), pp.120–5.

Wright, C.K. & Wimberly, M.C., 2013. Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences of the United States of America*, 110(10), pp.4134–9.

Yost, M. a. et al., 2014. Alfalfa Stand Length and Subsequent Crop Patterns in the Upper Midwestern United States. *Agronomy Journal*, 106(5), p.1697.