



Review

Evaluation of survey and remote sensing data products used to estimate land use change in the United States: Evolving issues and emerging opportunities

Minzi Wang^{a,b}, Michelle Wander^{a,*}, Steffen Mueller^c, Nico Martin^d, Jennifer B. Dunn^{e,f,g}

^a Department of Natural Resources and Environmental Sciences, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

^b Department of Resources and Environment, Moutai Institute, Renhuai 564507, China

^c Energy Resources Center, University of Illinois at Chicago, Chicago, IL 60607, USA

^d Department of Crop Sciences, University of Illinois, Urbana, Champaign 61801, USA

^e Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA

^f Northwestern-Argonne Institute of Science and Engineering, Evanston, IL 60208, USA

^g Center for Engineering Sustainability and Resilience, Northwestern University, Evanston, IL 60208, USA

ARTICLE INFO

Keywords:

Land use change
Survey data
Thematic maps
Remote sensing
Land use classification
Error
Uncertainty
Biofuel policy

ABSTRACT

Transparent, consistent, and statistically reliable land use/ land cover area estimates are needed to assess land use change and greenhouse gas emissions associated with biofuel production and other land uses that are influenced by policy. As relevant studies have increased rapidly during past decades, the methods used to combine data extracted from land use land cover (LULC) surveys and remote sensing-based products and track or report sources of uncertainty vary notably. This paper reviews six data sources that are most commonly used to investigate LULC and change in the contiguous U.S. by highlighting the main characteristics, strengths and weaknesses and considering how uncertainty is assessed by the June Area Survey (JAS), the Census of Agriculture (COA), the Farm Survey Agency (FSA) acreage, the National Resources Inventory (NRI), the National Wetlands Inventory (NWI), and the Forest Inventory and Analysis (FIA); and two remote sensing-based data products, the Cropland Data Layer (CDL) and the National Land Cover Database (NLCD). The summary and conclusion identify important research gaps or challenges limiting current land use/land cover and change studies (e.g., lack of high-quality reference data and uncertainty quantification, etc.) and opportunities and emerging techniques (data fusion and machine learning) that will improve reliability of land use/land cover assessments and associated policies. Blended approaches that marry high quality ground truth data that are more finely resolved than data supplied by government surveys with multitemporal imagery are needed track use of non-agricultural lands vulnerable to agricultural expansion. These considerations are notably important as the U. S. considers the renewal and possibly revision of its Renewable Fuel Standard, which includes provisions that require monitoring of agricultural land expansion.

1. Introduction

In the U.S., efforts to develop low carbon fuels and reduce greenhouse gas (GHG) emissions from transportation have formalized the role of biofuels in the United States' Renewable Fuel Standard (RFS) and state-level programs including California's Low Carbon Fuel Standard (LCFS) (Lepitzki and Axsen, 2018; Souza et al., 2017). As these policies took shape and demand for corn and other biofuel feedstocks grew, so did concern about the potential for agricultural land to expand into

natural lands such as forests, grasslands, and wetlands around the globe (Searchinger et al., 2008). Similar concerns have arisen in the development of the Carbon Offsetting and Reduction Scheme for International Aviation (International Civil Aviation Organization, 2019). Quantification of land use change (LUC) stemming from renewable fuel policies is both difficult and contentious because it can cause changes in carbon stocks and, on net, produce GHG emissions that would reduce or even negate biofuels' promise to stem transportation GHG emissions (Malins et al., 2020). Beyond the potential to generate GHG emissions,

* Corresponding author.

E-mail address: mwander@illinois.edu (M. Wander).

<https://doi.org/10.1016/j.envsci.2021.12.021>

Received 26 August 2021; Received in revised form 4 December 2021; Accepted 17 December 2021

Available online 28 December 2021

1462-9011/Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

losses of natural lands to agriculture also can negatively influence biodiversity among other important ecosystem services (Behrman et al., 2015). To preemptively minimize agricultural land expansion in response to the RFS, the policy excluded biofuels produced from feedstocks grown on “virgin agricultural lands cleared or cultivated after December 19, 2007.” The policy permits feedstock production on pastureland, rangeland, and land in the conservation reserve program (CRP). Along with this important designation, the RFS also requires the U.S. Environmental Protection Agency (EPA) to track and characterize agricultural expansion (Souza et al., 2017). This review considers whether and how existing and evolving data sources can allow the agency to fulfill this requirement with sufficient accuracy. This question is very important currently, as the RFS2 will be replaced or revised in 2022 (Congressional Research Service, 2020). Calls to repeal or revise the legislation justify review of many aspects of the legislation, including the requirements used to evaluate agricultural land expansion.

Foundational data sources that have been available for evaluating agricultural expansion since the RFS’ inception (2007): the Cropland Data Layer (CDL), the National Land Cover Database (NLCD), U.S. Forest Service (USFS) data products, and several U.S. Department of Agriculture (USDA) data sources including the Census of Agriculture (COA). Estimates of land under grassland cover or wetlands must be drawn from other sources (e.g., (Hoekman and Broch, 2018; Morefield et al., 2016; Wright et al., 2017). The challenges associated with using these data sources to evaluate agricultural expansion have become clear as efforts to quantify it have grown (Copenhaver et al., 2021).

For example, Lark et al.’s (2015) estimate of LUC between 2008 and 2012 relied primarily on the CDL. These authors noted that satellite revisit time used to make comparisons could vary by season, and that changing classification methodologies could create a false impression of change where land normally rotates into and out of crop production over several years. The US EPA’s 2018 Triennial Report to Congress has provided national guidance on LUC assessment by encouraging harmonization of methods to address our current inability to accurately quantify LUC that is in part caused by differences in the definitions and methodologies used by important sources used to classify land. That report also noted how evolving data features and techniques influence LUC estimates. For example, improvements in methodology in the COA and the USFS data products caused an increase in grassland area and a decrease in forest area on paper that did not actually occur on-the-ground. Lark et al. (2017) also showed that crop area underestimation needs to be accounted for and corrected when calculating changes to crop and cropland areas using the CDL along with other sources. They also acknowledged how methodological differences among studies, including their use of different reference data, complicates comparison of results. These observations underline the need to track and document methodological changes over time to provide accurate estimates of LUC. While noting there is not a national set of recommended guidelines for use of remote sensing data in LUC studies, Lark et al. proposed a set of good practices for post-classification LUC estimation that are relevant to any remotely sensed land cover dataset including the NLCD, Moderate Resolution Imaging Spectroradiometer (MODIS)-based land cover products, as well as other land use/land cover (LULC) classification maps.

In addition to the fact that data resources used in LUC analyses differ in classification, methods, reporting frequency and goals (Falcone, 2015), the analytical methods used can differ widely in approach. A very important difference among LUC studies is whether they address uncertainty, which is uncommon (Ayele et al., 2018). If uncertainty is addressed, methods to quantify it differ. Recommendations for best practices for uncertainty (e.g., making sure the reference classification is of higher quality; using more rigorous sampling strategy to collect reference classification; using unbiased area and accuracy estimators, etc. Olofsson et al. (2014) often do not consider critical spatial differences (Wickham et al., 2018).

Clearly, a better understanding of data sources used in LUC analyses

and their associated uncertainties is essential for policy makers and researchers who investigate the benefits and drawbacks of biofuels. This understanding could inform policy choices about how we use our land and enable analysts to draw more transparent and defensible conclusions. Now is the time to have this discussion and allow for tailoring of the RFS or a follow-up policy so that natural lands are protected as the U. S. pursues renewable fuels to help address climate change (Rogers et al., 2017).

Accordingly, this review summarizes the strengths, weaknesses, and differences among dominant LUC analysis data sources, for analysts, policy makers and the broader sustainability community interested in biofuels and GHG emissions associated with land use by considering the key characteristics of methodology, classification schemes, reporting scale and frequency, and error quantification used by survey and remote sensing data sources and products. Next, we examine differences in reported area of land categories among data sources, provide an in-depth consideration of how data sources differ in their treatment of grasslands, and quantify the error associated with different types of land by state and region. We conclude by highlighting emerging data and analytical capabilities that could enable the next stages of RFS implementation to better evaluate agricultural expansion. It is possible that data and methods have caught up to the needs of biofuels policies to monitor the agricultural expansion they may trigger.

2. Survey and remote sensing data sources and data products

Important characteristics of the most influential field survey and RS data products used to assess LUC in the continental U.S. (CONUS) are summarized in Table 1.

2.1. Federal field survey data sources

Among the USDA’s numerous surveys and resulting datasets, four (JAS, COA, FSA-crop acreage, NRI) are commonly used to estimate agricultural land expansion in the context of biofuel policies. We expand on important characteristics of these products and key complementary datasets (NWI and FIA) in the following sub sections.

2.1.1. NASS statistics (JAS and COA)

Arguably the two most important surveys of agricultural land use are USDA’s JAS and COA. While both surveys offer national coverage and are administered by the National Agricultural Statistics Service (NASS), their objectives, collection frequency and methods differ (Table 1). The JAS data provide direct crop-specific acreage estimates for major U.S. crops annually for each state in the U.S. except Alaska (Abreu et al., 2018). Data are aggregated to protect producers’ privacy and then published in every late June to support crop forecasting (USDA-NASS, 2019a). Due to its frequency, use of a national area sampling frame (Table 1), and the fact that the data are collected at the field-level, the JAS is widely viewed to be of high accuracy and so it is commonly used as reference data to train and validate cropland area estimates produced by other sources, such as the CDL (Table 1) (Boryan et al., 2011).

The COA is also unique because it is legally mandated and provides the longest historical record of land use of any source. Producers must share information about harvested cropland acres, land use, ownership, operator characteristics, production practices, income, and expenditures (Lamas et al., 2007; USDA-NASS, 2020). As with the JAS, farm-specific data and farm number are also collected at finer scale but aggregated and reported at the state-level (USDA, 2019b).

2.1.2. USDA administrative statistics (FSA acreage data)

The crop reporting statistics provided by NASS are complimented by the Farm Service Agency (FSA) and USDA Risk Management Agency (RMA) data. These entities administer U.S. conservation, farm commodity and disaster programs including the Agriculture Risk Coverage program. Unlike the COA and the JAS, the FSA does not follow any

Table 1

Commonly Used Datasets in Land Use Change Research. Spatial and temporal coverage, reporting frequency, land use types or categories of land cover tracked, methods used by LULC survey sources and remote sensing data products commonly used in the U.S.

| Dataset and Source | Frequency | Sampling method/ scale | Coverage | Accuracy | |
|--|---|--|--|---|--|
| June Area Survey (JAS) USDA | 1 year (1975-current) | Area sampling frame (ASF) divided among land use types before systematic sampling to generate classes with similar proportions of cropland use that are divided into primary sampling units (PSU= 1 mi ²) from which a subset are randomly selected for enumeration by interview of all farm operator within its boundaries. | Provided at the state level for all states except for Alaska | Varies by crops. The relative standard error from the 2019 survey are 1.2%, 1.3%, 9.8%, 3.0%, 8.6%, 4.1%, and 2.2% for corn, soybeans, barley, cotton, sorghum, spring wheat, and winter wheat. | |
| Census of Agriculture * (COA), USDA National Ag Statistics Service (NASS) | 10 years (1840–1920) 5 years (1920–2017) | Farms are defined as area producing \$1000 worth of ag products during a census year. Based on list containing all known and potential farms. The area frame from JAS is then used to account for the under-coverage of the list. | Provided at the county level for all 50 states | If the data are measured using the Census response rate: 71.8% for 2017 census 74.6% for 2012 census 78.2% for 2007 census | |
| National Field Data | Crop Acreage Data, USDA Farm Service Agency | 1 year (2007-current) | Based on self-reporting from farm operators at local administrative office for planted, failed, or prevented for fields within a Common Land Unit (CLU) delineated by a permanent, contiguous boundary, a common land cover and land management, a common owner and a common producer in agricultural land associated with USDA farm programs. | Provided at county for all land enrolled in the USDA programs unless CLUs for land use type are too few to obscure identity of producer | A 90% confidence interval with a tolerance of three meters from ground features visible on the photography is used during digitization of the CLU tract and field boundaries based on photography; |
| National Resources Inventory (NRI), USDA-NRCS | 5 years (1977–2015) | Longitudinal sample based on a two-stage stratified area sample with the primary sampling units (PSUs) being segments of townships within each county. Segments are first randomly sampled by stratum (typically are two by six square miles) before three sample points are selected for data collection. | Provide at the state level for all states except for Alaska | | |
| Forest Inventory and Analysis (FIA), FSA National Wetlands Inventory (NWI) | | Data are collected at county and sub-county level followed the U.S. Army Corps of Engineers (ACOE) methodology. | | CONUS plus territories and 35% Alaska** | |
| 1 year (mid-1980 s-present) open-source products derived from RS data | Cropland Data Layer (CDL) | 1 year (1997–2018) | 30 m | Certain states (1997–2007) At state level for CONUS (2008–2018) | Varies by states and years, and changes annually. E.g. 2009 CDLs have overall accuracy of 85% – 95% for major crops |
| | National Land Cover Database (NLCD) | 2 or 3 years (2001–2016) | 30 m | CONUS | Overall agreement: 71~97% |
| | GlobeLand30 | 2000, 2010 | 30 m | Global | (1) ~ 80% (overall accuracy) with 150, 000 validation sites; (2) ~ 84% for entire China (Wang et al., 2018); (3) ~ 78% for entire Iran (Arsanjani et al., 2016); |

Notes: JAS = June Area Survey; COA = Census of Agriculture; FSA = Farm Service Agency; CLU = Common Land Unit; CDL = Cropland Data Layer; NLCD = National Land Cover Database; INHS =

Notes:

a Principal crops included are corn, sorghum, oats, barley, rye, winter wheat, durum wheat, other spring wheat, rice, soybeans, peanuts, sunflower, cotton, dry edible beans, potatoes, sugar beets, canola, and proso millet.

b Specified Crops include 1) a variety of field crops, e.g. corn, cotton, soybean, etc.; 2) field and grass seeds, forage, and hay; and 3) other specified crops. For more details of which specific crop is being included, refer to the Table 35 in the 2017 COA full report.

c Grasslands in the data from the FSA include 1) grassland in farms; 2) grassland enrolled in USDA's Conservation Reserve Program (CRP) and Grassland Reserve Program (GRP); and 3) land used as emergency watershed/floodplain. In addition, Shrubs/forbs and Kochia (Prostrata) are also included in the Grasslands categories used for data comparison analysis in this paper.

d During the NRI sampling, the land of a county is divided by township (usually are six-miles square areas), then the land of each township is subdivided into "section" (commonly defined as 1-mile square area). Each stratum contains 12 such sections and a quarter of a section is considered as a segment.

* Collection of data is mandatory and required by law.

** No data quality information provided.

specific sampling methods for crop acreage estimates or attempt to report total area sums. Instead, the local FSA office collects cropland use information including area planted, planted but failed to yield, and unplanted or abandoned for common land units (CLUs). Information is gathered annually from producers participating in government programs tied to social insurance, subsidies, and taxation (Carfagna and Carfagna, 2010). Due to their comprehensive coverage, FSA data are unique and considered to be complete and highly trustworthy (Mueller et al., 2009). This makes them invaluable for ground truthing other data products. Unfortunately, as with other USDA data sources, FSA information is only available to the public in aggregate at either the county or state level (USDA-NRCS, 2020). Since 2010, data for selected crop sectors are released and updated throughout the year after new information is received (Good, 2014).

2.1.3. NRCS statistics (NRI)

Information about non-agricultural lands such as forests, wetlands and grasslands, which is not covered by the agriculturally focused surveys previously reviewed, is provided by the National Resources Inventory (NRI) (USDA-NRCS, 2020). A strength of this key resource is that it contains longitudinal data collected annually with consistent methodology from a subsample of fixed, non-Federal sampling sites that are revisited over time. Since 1977, the resulting data are summarized and reported every five years at the state level but more resolved data can be obtained upon request (USDA-NRCS, 2020). The NRI provides the longest history (1982–2015) for examining non-agricultural use trends at the national scale. Unfortunately, very little information is given for points on federal lands covered by the inventory.

2.1.4. U.S. Fish and Wildlife Service and the Forest Service statistics (NWI and FIA)

Both the National Wetlands Inventory (NWI), which is administered by U.S. Fish and Wildlife Service (USFWS or FWS) Agency, and the Forest Inventory and Analysis (FIA) program, administered by the U.S. Forest Service (USFS), provide more in-depth survey data covering the nation's wetlands and forestry resources. Both sources provide broad temporal coverage and rely on periodic resampling of plots. The FWS shares the Wetlands Data Layer for the National Spatial Data Infrastructure (NSDI), which is compiled using information provided by 160 organizations that collect state and regional inventory data following established data standards (Stout, 2007). The FIA data covers federal and non-federal lands by drawing on a combination of remote sensing imagery and field survey carried out by the USFS.

2.2. Remote sensing data products

Successful efforts to use RS imagery to generate unbiased statistical estimates of crop area at the state and county levels provide an important way to reduce variance and uncertainty in acreage estimates obtained from farmer reported surveys (Craig and Atkinson, 2013). Remote sensing based products offer greater spatial and temporal coverage and higher resolution than national surveys including the CDL and the National Land Cover Database (NLCD) which are both federally produced and the most frequently used for national LULC assessment.

2.2.1. The CDL and NLCD

The CDL program produces crop-specific thematic maps using a combination of FSA CLU data and RS imagery (Boryan et al., 2011). It also relies on some ancillary data sets including the USGS' National Elevation Data, forest canopy and the NLCD (Boryan et al., 2011). Multiple data foundations for the CDL include Landsat imagery, Sentinel, and the MODIS among others. The [supplementary information](#) (SI) provides more detail about imagery used for producing the CDL (Appendix A). Data are translated into CDL's thematic classes using a decision tree classifier that ground-truths categories. NASS used JAS data as ground truth from 1997 to 2005, before shifting to rely primarily

upon FSA's CLUs for classifier training and testing purposes (Boryan et al., 2011). Ancillary ground truth data varies by State. Both the resolution, which increased from 56 m to 30 m in 2010 and, methods to develop the CDL have improved over time. These improvements can cause inaccurate estimates of change when historical CDL maps are compared with newer maps. Several efforts using the CDL data set in LUC studies (e.g., Lark et al., 2015; Wright and Wimberly, 2013; Wright et al., 2017) encountered this problem after applying the CDL outside of its' intended use, which is to provide accurate, timely, and useful agricultural information for major crops of a year. Updated versions of historical maps are needed to enable more accurate temporal LUC analyses.

The NLCD is a land cover product of the Multi-Resolution Land Characteristic Consortium that tracks trends in all lands and provides detailed categorical land cover information for 16 modernized Anderson Level II classifications covering the entire U.S. (<https://www.mrlc.gov>). In this resource, cropland is combined into a single category of cultivated crops. Data are generated at 30 m ground resolution using Landsat imagery and reported every five years (Homer et al., 2015, 2004). Beginning in 2006, NLCD added national scale LULC information and in doing so pioneered change estimate methodologies (Wickham et al., 2013). Newly developed products are released as updates of previous versions that are shared along with original data (Wickham et al., 2017). The newly released NLCD 2016 not only updates all previous versions of NLCD 2001, 2006, and 2013, but also offers four additional land cover versions for years 2003, 2008, 2013 and 2016 (<https://www.mrlc.gov/>). By updating versions of the dataset, NLCD allows for repeatable multi-temporal LULC change estimates at 2–3-year intervals between 2001 and 2016 (Wang and Atkinson, 2018). These data are invaluable and so widely used to assess wetland loss caused by cropland expansion (e.g., Johnston, 2013) and changes in crop rotation (Stern et al., 2012).

3. Uncertainty and LULC data sources

3.1. Error sources

All of the data sets we include in this review contain error and associated uncertainty. Understanding the origin of uncertainty in survey, administrative, and RS data can improve understanding of error inherent in LUC analyses that use them.

The types of error contributing to uncertainty vary by source. Survey-based error is generally derived from sampling and non-sampling processes while errors in remotely sensed data originate from positional and thematic/classification errors (SI Table B1). Other sources of error and discrepancies among surveys result from differences in statistical design (Table 1). Survey-based results also contain non-sampling errors that represent the difference between data collected and their ground truth values resulting from data gathering, processing, and recording. Response rates are one source of non-sampling error. Declining response rates for agricultural surveys have been observed for decades with no sign of trend reversal (Johansson et al., 2017). Other examples of non-sampling error including enumerator error and data processing error are explored in Appendix B. When remote sensing data are turned into a map product, errors can occur in many steps from imagery acquisition, processing, analysis and assessment as well as visual presentation.

Using RS products for LUC analyses introduces additional error beyond the error inherent in thematic maps themselves. To detect change in land cover based on bi- and multi-temporal RS imagery error must consider differences that come from the change in detection algorithms as well as the classification schemes used to produce the maps. Both the quality, alignment of the imagery (Faiza et al., 2012) and the classification scheme may differ between RS products for the start and end years of an LUC analysis. Spectral heterogeneity within land cover classes can also add uncertainty (Gomez-Chova et al., 2015). All these types of error are introduced and propagated into the change estimates

in an unknown way. These issues can contribute to false positives in the change map and thus may significantly “devalue” the use of remote sensing in LUC monitoring unless corrected (Lark et al., 2017; Wang and Atkinson, 2018). Efforts to update the CDL and NLCD to draw upon improving imagery provide examples of corrected data sources that increase confidence in LUC assessments.

Other changes in methods or norms can improve the quality of data sources used for LUC assessments. Spatial summaries of LULC or RS products that draw upon multiple sources of information must overcome challenges in data’s scale of resolution. Valuable information is lost when finer-scale data are combined with resources reported at a coarser scale. These limitations are commonly overcome using data fusion to gap fill and improve and validate data sources. For example, improvements are made to make final official cropland use area estimates based on the CDL using the JAS data to calibrate and adjust the area estimates obtained by regressing against pixels on CDL maps. Optimal statistical methods vary with crop and map scale (Butler et al., 2018). Map error based on remote-sensing data can and will also improve with reference data quality (ground-truth data collected through field survey or extracted from higher quality imagery) (Congalton, 2001; Khatami et al., 2017). Ground-based field surveys commonly suffer from infrequent reporting and inconsistent or incomplete coverage (Larsen et al., 2015). In addition to increasing the frequency and intensity of sampling, investigators can note sampling design used to collect ground truth data and use well-document accuracy assessment techniques to improve the quality, transparency, and reproducibility of efforts (Olofsson et al., 2013; Stehman, 2013; Wickham et al., 2021).

3.2. Importance of quantifying uncertainty in LULC data

As previously noted, the goals, methods, and accuracy of agricultural survey data and remote sensing data sources vary. To illustrate how these variations influence estimates of land area by type (cropland, grassland, wetland, woodland) among the data sources, we conducted a cross-source comparison of these estimates (Fig. 1). One of the most important differences among these data sources is how they define and treat each land category. Appendix C, Table C1 provides a detailed summary of JAS, COA, FSA, NRI, CDL, and NLCD land category definitions and how they were combined into broad groups summarized in Fig. 1. Methods used to calculate the error associated with categories reported by each data source are described in Appendix B.

3.2.1. Croplands, forests and wetland discrepancies

There are some important differences in coverage of cropland by the data sets we have described, which is driven in part by the methodologies used to develop them. For example, even though the FSA covers ~95% of croplands based on farmer reporting, it often misses small or specialty crops that aren’t subject to government benefits. On the other hand, the JAS is a statistical-based survey that uses a list frame method to cover all U.S. croplands without as detailed spatial information as the FSA contains. Because the FSA does not cover all croplands, it estimates a lower cropland area than the COA and JAS (Fig. 1, SI:Table C2) even though the FSA CLU is more comprehensive. In addition, whereas the COA and FSA include unplanted or non-harvested areas within their cropland estimates, the JAS does not account for failed or non-harvested areas. The treatment of idle lands is also important because these lands regularly flip in and out of agricultural use. Additionally, what is considered idle varies among sources. For example, FSA treats non

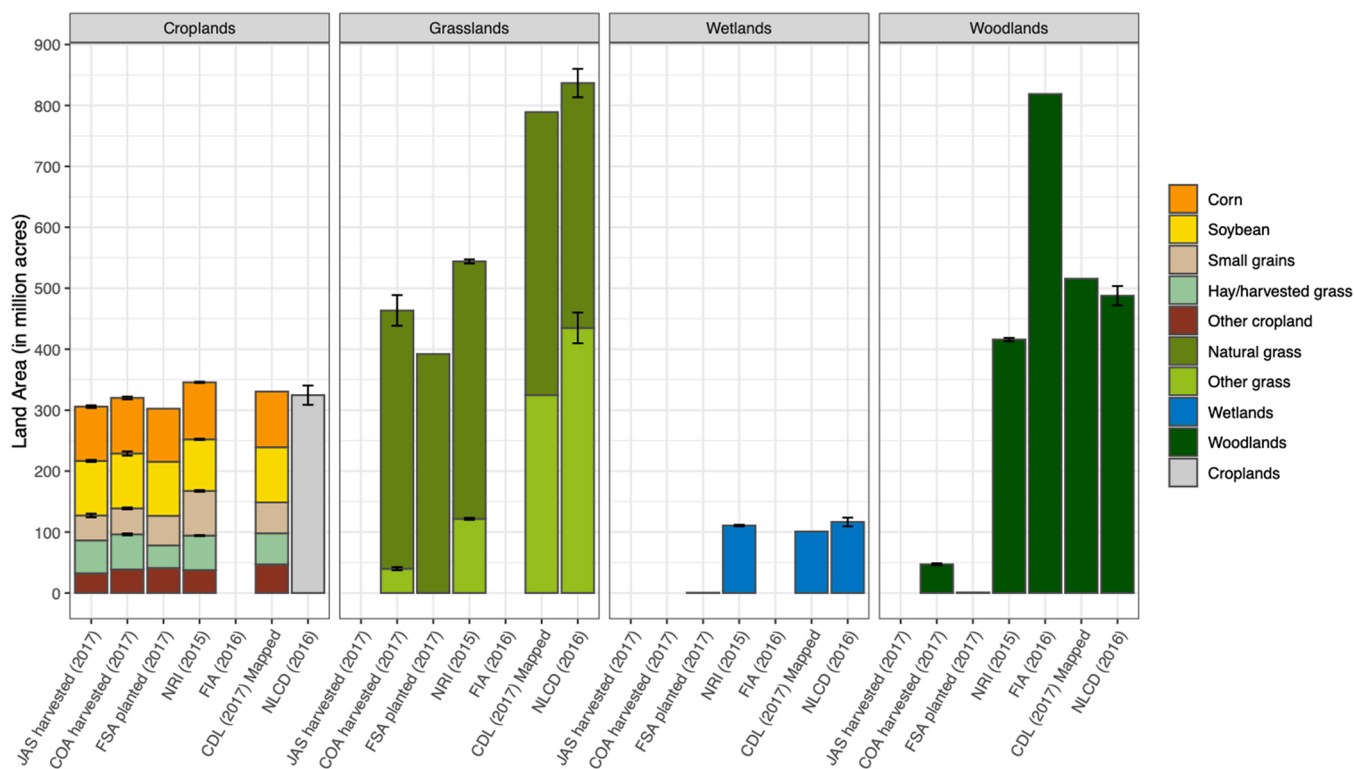


Fig. 1. Cropland refers to harvested area for JAS, COA, and NRI, planted area excluding failed ha for FSA, and mapped area for CDL and NLCD excluding idle/fallow categories. See SI: Appendix C; Table C2 for more details.

(a) Summary of estimated and reported land-use/land-cover (LULC) area and associated error for the June Area Survey (JAS), Census of Agriculture (COA), Farm Services Agency (FSA), Forest Service Inventory (FIA) National Resources Inventory (NRI), Cropland Data Layer (CDL), and National Land Cover Database (NLCD)) using records for or as near as possible for 2017 that are separated into major categories including croplands, grasslands, wetlands, and woodlands (left to right). (b) Where possible, the margin of error (MOE) was computed for these four major classes along with several sub-classes after reclassification of original source categories using population statistics reported by JAS.

harvested crops including cover crops as idle land but NASS counts this area as cropland (USDA-NASS, 2020). Despite this, we found FSA estimates of fallow or idle acreage are about one fifth (~5 M ha) the amount estimated by the COA (~25 M ha) (SI:Table C2) due to differences in sources' focus. Comparisons among estimates suggest NASS survey estimates overestimate 1 ~ 7 M ha of cropland due to double counting of idle/cover cropped acres that may be planted but unharvested (SI: Table C2).

Data sources can also differ in wetlands and woodlands estimates because of classification discrepancies (SI:Table C1). Area estimates for wetlands are quite similar for CDL (40.8 M ha), NRI (44.8 M ha) and NLCD (47.1 M ha) (Table C2). But, despite consistent use of the Cowardin classification system (Dahl et al., 2020) by both the NWI and NRI, these sources differ slightly in their estimates of wetland area due to different legislative mandates, sampling methodology, data collection and estimation procedures (USDA-NRCS, 2018; <https://www.fws.gov/wetlands/>). Estimates of woodland area vary more widely – from 197.4 (NLCD) to 331.4 (FIA) M ha – than estimates of wetland area. Inclusion of shrublands within woody grassland (176 M ha shrub/scrub) within “other” grassland may only partially explain why NLCD woodland area estimates are only ~60% of the area reported by the FIA (Fig. 1). Additionally, whereas the FIA uses a relatively broad woodlands definition and prioritizes woodland tracking, the NLCD uses a narrow woodlands definition, classifying woodlands as areas with at least 20% vegetation with trees taller than 5 m (Homer et al., 2012).

Interrelationships between data sources should also be recognized. The CDL relies heavily on the NLCD to map natural lands (Boryan et al., 2011). Even though NASS surveys that are used to supplement the CDL completely cover the CONUS, not all land use types are covered equally. Accordingly, the NLCD data have been used since 2006 with FSA-CLU data to assess the mapping accuracy for non-agriculture lands in the CDL (e.g. forest, water, grass, cities) (Boryan et al., 2011).

3.2.2. Estimates of grassland area vary widely

Inconsistent use of terms used to classify and track grasslands, which include any land on which grasses are the dominant vegetation, may be the most pronounced and consequential for LUC assessments. While a strictly land cover-based definition includes meadows, rangeland, prairies, and tundra, grasslands must be subdivided according to management and cover to properly account for land use (Sanderson et al., 2009). Grasslands managed for silage, pasture, and hay all need to be defined from a land-use perspective. Surveys are nearly uniquely positioned to provide information about grassland use and management that remote sensing products struggle to capture.

Variability in classification of natural and native grasslands and confusion about where to place intermittently or sparsely managed grasslands hinders studies of agriculturally driven grassland loss. For example, the COA subdivides grasslands that are considered working lands within its ‘other’ and ‘cropland’ subclasses into 10 subclasses based on survey responses gathered by FSA. It defines areas in permanent pasture (any enclosed area of land that is for grazing and not suitable for cropland or woodland pasture) and rangeland as ‘grasslands’ (USDA-NASS, 2019b). Similarly, both pasture (uncultivated) and rangeland terms are classified as grassland categories by the NRI, but these are defined in a slightly different way than done by the COA (USDA-NRCS, 2018). The NRI reports rangeland and uncultivated pastureland as grassland. It also introduces the term “grass” in its definition of hayland, which is a subcategory of non-cultivated cropland and is part of the broader category of cropland. In this example, the complication deepens when considering that the NRI considers hayland and pastureland that are in rotation with row crops to be cultivated cropland. In contrast to the COA and NRI, the JAS records non-harvested areas as broad land use categories (wetland, woods, pasture/rangeland) in some states (e.g., Illinois) but does not report these areas given the nature of that survey (Gerling et al., 2015). Survey-based estimates thus range from 158.6 M ha in the FSA to 220.1 M ha in the NRI.

Differences among survey-based estimates were small compared to difference between survey- and RS- based estimates. The latter suggest as much as twice as much area under grass depending on the comparison (Fig. 1). Area estimates for grasslands vary greatly from 158.6 M ha to 338.6 M ha (SI: Table C1) with RS-based sources reporting much larger overall areas (> 100 M ha) in permanent grass (Fig. 2). For example, the CDL estimates 319.3 M ha grasslands, with a high level of uncertainty (~ 20% misclassification), while the FSA reports 158.5 M ha (SI: Table C1). The RS products offer less grassland detail than the COA and NRI, even though they include a broader range of grassland types. Neither the NLCD nor the CDL distinguish among uses of grasslands (e.g., for grazing versus CRP) or provide grassland classifications like permanent pasture (Fig. 2). The NLCD only reports three grassland subclasses (Fig. 2; SI:Table C2). The CDL collapsed three historical categories: “Pasture/Grass”, “Grassland/Herbaceous”, and “Pasture/Hay” in 2014, limiting analysts’ ability to ascertain changes in the area of these different types of grassland. The CDL is the only national data source updated annually that includes both agricultural and non-agricultural lands. Its’ reliance on less-frequent NLCD for non-agricultural land data of course introduces inaccuracies.

Extremely large differences between the NRI and RS-based CDL and NLCD (SI: Table C4) reflect differences in the definitions of grassland used by survey and satellite imagery classifications and the purpose of the resource. Complicating LUC analyses of grasslands is the fact that ‘natural’ and ‘other’ grassland categories account for a substantial proportion of non-agricultural lands described by the NRI, CDL, and NLCD (SI:Table C4). Based on these examples of differences among data sources discussed in this section, users of multiple sources for comparison of grassland area within and across years must equate categories or generate aggregate classes that may lead to contradictory findings and add uncertainty.

The treatment of CRP lands which, like idle lands, can move into and out of agricultural use deserves special attention. Despite their importance for assessments of agricultural expansion, only the COA, FSA, and NRI consider them specifically (Fig. 2). All three surveys reporting CRP lands rely on aggregate information provided by USDA’s FSA, and NRCS. This broad category includes areas enrolled in continuous CRP and Conservation Reserve Enhancement Program (CREP) that target the most vulnerable areas and account for about a quarter of the area enrolled in conservation contracts and, lands enrolled through the general the CRP or other initiatives (e.g., Farmable Wetland Program (FWP) and State Acreage for Wildlife Enhancement (SAFE) initiatives), that commonly enroll whole fields. While both managed and idle grasslands and CRP lands are counted as cropland in EPA’s RFS (US EPA, 2010), we categorized them as grassland rather than cropland because of previously noted discrepancies among the data sources’ idle land’s classification schemes. Even though these land categories account for only a small share of all land (< 2% based on the FSA, CDL, NRI) and cropland (~6%) (SI:Table C3), fluctuations in program size that result in reversion to cropland can result in substantial losses of carbon and ecosystem services (Li et al., 2017).

3.2.3. Regional variability

We conducted a reanalysis of the CDL data to explore how accuracy varies across the CONUS and found that uncertainty, based on probability of misclassified pixels, declines as area covered by a land use category increase (Fig. 3). When the share of total land in any land category within a state exceeds 10%, error rapidly declines with its share of total land and then levels off (Fig. 4). For example, the error is relatively small for corn area in IL (4%) and IA (2%) where corn is a dominant crop. The error for corn area reaches 29% in NM, where corn accounts for just 0.1% of the state (Fig. 4).

In general, major crops such as corn, soybean and small grains tend to have smaller errors than non-agricultural lands regardless of dominance in the landscape. For example, soybean error exceeds 25% in only five states (CO, GA, MT, OR, and WA) (Fig. 4). Hay/harvested grass also

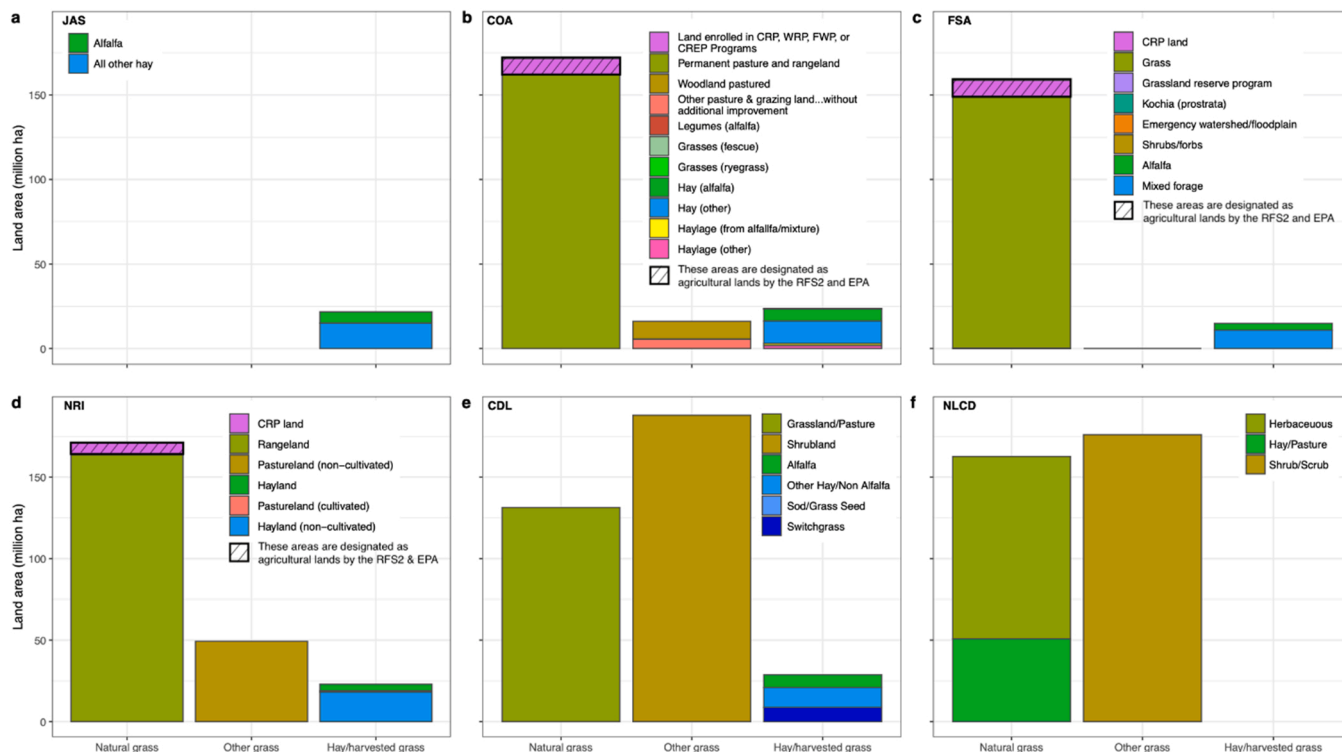


Fig. 2. (a) Summary of national grassland area reported for natural, other, and cropland subcategories, hay or harvested grassland, and associated nomenclature used by survey sources and data products. (b) (a) June Area Survey (JAS), (b) Census of Agriculture (COA), (c) Farm Services Agency (FSA), (d) National Resources Inventory (NRI), (e) Cropland Data Layer (CDL), and (f) National Land Cover Database (NLCD).

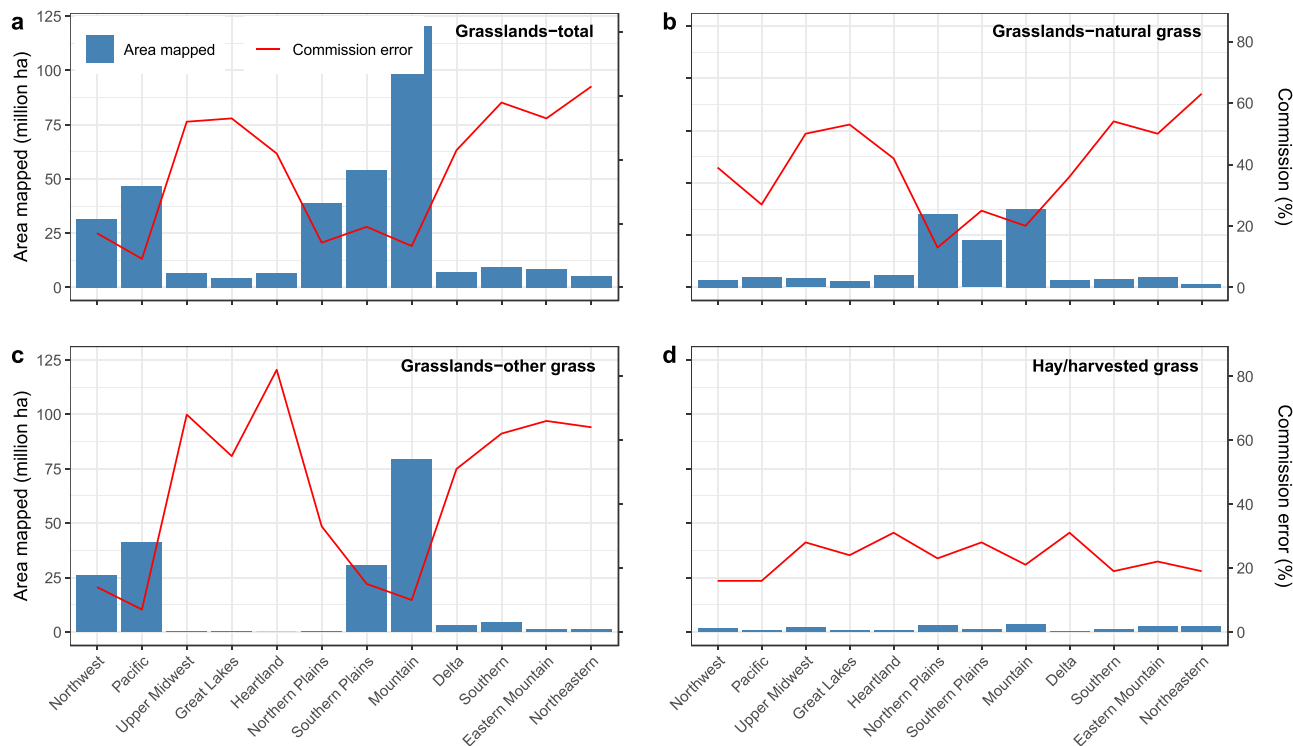


Fig. 3. National grassland area and associated map errors reported by the Cropland Data Layer (CDL) 2017 for (a) grasslands category, (b) natural grassland, (c) other grassland, and (d) hay or harvested grassland subcategories across the National Agricultural Statistics Service (NASS) regions. For this figure, the map errors refer to commission errors for corresponding main and sub-categories. Thus, the map error of a category indicates the rate of misclassifying pixels that belong to other land use/cover categories into that category.

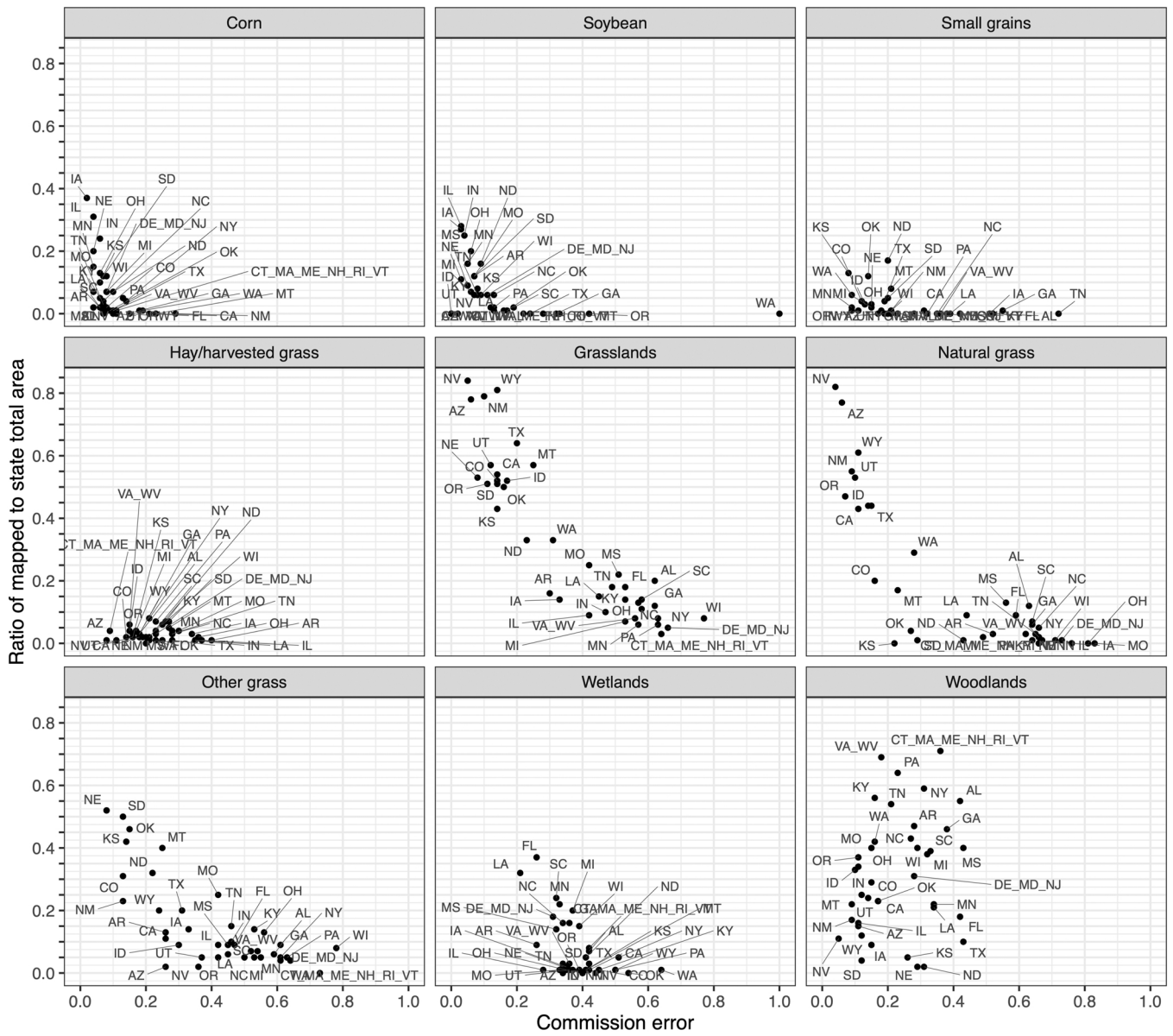


Fig. 4. Relationship between ratio of mapped area to state total area and map error (commission error) reported by the Cropland Data Layer (CDL) 2017 for the constructed main- and sub-categories. For this figure, main categories include Croplands, Grasslands, Wetlands, and Woodlands. Subcategories include Corn, Soybean, Small grains, Hay or harvested grass under Croplands, Natural grass and Other grass under Grasslands.

has a relatively low error across all states though it only occupies less 9% of any state’s total land area. In contrast, map error for total-, natural- and other- grasslands is much higher and does not decline below ~25% until it occupies ~40% of land area in a state (Fig. 4). This error increases markedly in the Midwest and eastern regions, with rates exceeding 50% for most states in those regions (Fig. 3). Trends are similar for wetlands and woodlands, where the commission error remains greater than 25% in many states. This high error holds even when these classes occupy a large proportion of area because these non-agricultural land categories are not CDL’s mapping focus.

4. Summary and conclusions

Mitigating agricultural expansion that increased demand for biofuels could trigger is a core component of the RFS. As we reach an inflection point punctuated by calls for repeal or reform of the RFS and how it is implemented, the need for a common understanding among policy makers, analysts, and other stakeholders of the strengths and

weaknesses of data sources frequently used to evaluate agricultural expansion becomes clear. Numerous spatial and temporal factors, including the nomenclature used to separate LULC categories, contribute to variability. Blended products, like the CDL and NLCD, that combine different types of data and consider both agricultural and non-agricultural land are critical resources. The CDL and NLCD in particular have been subject to evolving and improving methods with older data sets being revised with updated methodologies. Improvements have reduced discrepancies among estimates of agricultural land that arise from comparisons of thematic maps of different vintage. As new data and techniques emerge, the ability to identify land by cover type (cropland, wetland, grassland, woodland) will improve and bring quantitative accuracy to levels that bring quantitative accuracy espoused by policy within reach. While at present the accuracy associated with cropland identification is relatively high, there is an urgent need for improved accuracy in improving tracking of marginal lands classified as grassland-other, or CRP and wetlands, that are vulnerable to agricultural expansion. Improving and emerging data sources can help

achieve that objective.

Since the inception of the RFS, we have entered an era of big data. Increased volumes and quantities of data offer new opportunities and resources to substantiate sustainability claims for biofuels and other land use categories (Calvin et al., 2021). For example, products such as the National Agriculture Imagery Program (NAIP), GlobeLand30 and Sentinel-2 are evolving and improving rapidly in terms of spatial resolution, revisit time, and sensor capability to enhance identification of land by type (Drusch et al., 2012; Knight and Kvaran, 2014; Maxwell et al., 2017). Studies from recent years show the importance of NAIP in tracking wetlands (Wu et al., 2017; Xie et al., 2019) and also demonstrate the potential of the Sentinel-2 data in mapping and detecting grassland (Griffiths et al., 2019b; Koleccka et al., 2018; Rapinel et al., 2019b), wetlands (Araya-López et al., 2018; Arroyo-Mora et al., 2018; Ludwig et al., 2019; Rapinel et al., 2019a), and cropland (Defourny et al., 2019; Griffiths et al., 2019a). Of course, sensors with higher spatial resolution like those in Sentinel-2 can increase image processing and computational resource demand which is a factor that must be considered.

Beyond these public sector initiatives, the private sector is also expanding rapidly to provide high-quality LULC data. The Satellite Imaging Corporation's RapidEye (5-m resolution, revisit time between 1 and 5.5 days) and WorldView-2 (2-m resolution, revisit time 1 day) satellites are two important examples of higher resolution data (Urdike and Comp, 2010). Worldview-2 has great potential in mapping wetlands (Tarantino et al., 2016) and to identify grassland changes (Tarantino et al., 2012). While powerful, these high-resolution data have drawbacks including cost and their relative short archive which may limit their use for studying historical agricultural expansion over large geographical areas (Defourny et al., 2019).

High quality ground truth data that are vital to assess the accuracy of RS and other data products are also evolving rapidly. Application-specific, open-source products such as very fine field or sub-field level data obtained from individualized input using handheld devices or laptops have great potential to replace crude aggregated agricultural survey data. For example, phones and tablets collect geolocation data, planted and harvested areas by crop type, irrigation amounts, fertilizer application rates, and tillage practices. These data can be combined to create a shared, crowd-sourced resource (Laso Bayas et al., 2017). Some claim that these methods provide information as or more accurate than data in RS products (Estes et al., 2016). Some researchers seek to mitigate concerns about data reliability by offering monetary rewards and alternatives including of authorship scientific papers to encourage data contributors to provide data of sufficient quality and frequency (Laso Bayas et al., 2017). Similar platforms can leverage individual input on land use in the form of citizen science (Deguines et al., 2012; Walter et al., 2018). In addition, synergistic efforts are also drawing upon unmanned Aerial Vehicles and robot platforms to cover small areas and complement satellite data. Though they have limited spatial coverage, they do in fact offer revisit flexibility and the potential for high temporal resolution and can offer very fine-scale data (e.g., field and even plant-specific details). In programs to audit claims that biofuel feedstock production is not inducing agricultural expansion, in Europe but also globally, field-specific data is often collected. These efforts could provide a template to follow in the United States. While all of these emerging data sources offer the potential to contribute to studies of agricultural expansion as, at a minimum, one type of ground truth, their value will depend upon the extent to which data are compiled and shared using transparent and consistent methodologies. To make full use of these data both the high costs of ground truth measurement and privacy concerns must be addressed.

The rapid growth in data availability for LULC and change studies call for advanced data analysis techniques and encourage the use of powerful data processing, computing techniques and storage tools. These include but not limited to data fusion, machine learning techniques, high performance computer cloud computing, Google Earth

Engine, and so on. Examples include fusion of Sentinel-2, the RapidEye, and Landsat-8 OLI to help detect LULC and associated change (Amani et al., 2018); combination of LiDAR data with multispectral Earth Observation data to improve wetland mapping accuracy (Rapinel et al., 2019). Integration of ground-based field survey and RS derived data (land cover maps) can improve the mapping reliability and detail (Henrys and Jarvis, 2019). The FSA data are used in this way to enhance the CDL (Craig, 2005). Furthermore, advances in RS increasingly look to automatization of algorithms used to map LULC and identify LUC. ML approaches, such as Support Vector Machine and Random Forest (RF), decision trees, etc., are often combined with fusion algorithms to enhance performance (Debolini et al., 2015; Weiss et al., 2020). For example, by combining multispectral Landsat and corresponding NDVI data, Huang et al. (2017) use Google Earth Engine cloud calculations to find major land types including wetlands. Similar techniques could be applied in the US.

All in all, data products that can inform our understanding of agricultural expansion and the corresponding loss of grasslands, wetlands, and forests have come a long way since the RFS' inception. It is now possible to conduct detailed, frequent, and high-accuracy evaluation of agricultural expansion that is needed to support biofuels policy. As the RFS evolves, it would be advisable to consider how advances that have been made and 'best practices' that have been observed can be incorporated into new policies to make full use of the above-described sources. There will still be a need, however, to grapple with these sources' differences in spatial and temporal coverage, taxonomy and uncertainty.

There are several high-priority needs to foster further advance. In particular, to overcome significant shortcomings in grassland characterization, analysts will need to track lands enrolled in conservation programs or of similar quality that are reported in average by surveys due to privacy laws. RS and crowd-sourcing that combine survey, remote sensing, and high-resolution aerial imagery data can provide higher accuracy techniques to identify CRP and grasslands or wetlands at risk of succumbing to agricultural expansion. Relatedly, finer-scale data are needed that go beyond land cover but also indicate land management pertaining to grasslands. These types of data can clarify the status of grasslands and whether they are managed, unmanaged, or in transition. Second, the role of multi-temporal imagery, especially in hotspots, should be evaluated. Existing public high-resolution imagery sources, like NAIP, are produced infrequently and can't provide insights into agricultural expansion into grasslands and wetlands on policy-relevant timescales. Another high priority is continued development of techniques that apply big data and computing power to the growing volumes of large imagery and RS datasets. Deploying artificial intelligence and machine learning can speed data interpretation and subsequent action.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge Joshua Pritsolas and Randy Pearson from GeoSpatial Mapping, Applications, and Research Center at Southern Illinois University Edwardsville. This research was supported by U.S. Department of Agriculture National Institute of Food and Agriculture award 2018–10008-28530.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.envsci.2021.12.021](https://doi.org/10.1016/j.envsci.2021.12.021).

References

- Abreu, D.A., Lawson, L.A., Hickman, S., 2018. Assessment of a review process for the 2017 census of agriculture. *Proc. Jt. Stat. Meet. 2018 Surv. Res. Methods Sect. Vanc., Br. Columbia, Can.* 1742–1750. July 28 - August 2, 2018.
- Amani, M., Salehi, B., Mahdavi, S., Brisco, B., 2018. Spectral analysis of wetlands using multi-source optical satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 144, 119–136. <https://doi.org/10.1016/j.isprsjprs.2018.07.005>.
- Araya-López, R.A., Lopatin, J., Fassnacht, F.E., Hernández, H.J., 2018. Monitoring Andean high altitude wetlands in central Chile with seasonal optical data: a comparison between Worldview-2 and Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* 145, 213–224. <https://doi.org/10.1016/j.isprsjprs.2018.04.001>.
- Arroyo-Mora, J.P., Kalaska, M., Soffer, R., Ifimov, G., Leblanc, G., Schaaf, E.S., Lucanus, O., 2018. Evaluation of phenospectral dynamics with Sentinel-2A using a bottom-up approach in a northern ombrotrophic peatland. *Remote Sens. Environ.* 216, 544–560. <https://doi.org/10.1016/j.rse.2018.07.021>.
- Ayele, G.T., Tebeje, A.K., Demissie, S.S., Belete, M.A., Jemberie, M.A., Teshome, W.M., Mengistu, D.T., Teshale, E.Z., 2018. Time series land cover mapping and change detection analysis using geographic information system and remote sensing, Northern Ethiopia. *Air, Soil Water Res.* 11. <https://doi.org/10.1177/1178622117751603>.
- Behrman, K.D., Juenger, T.E., Kiniry, J.R., Keitt, T.H., 2015. Spatial land use trade-offs for maintenance of biodiversity, biofuel, and agriculture. *Landscape Ecol.* 30, 1987–1999. <https://doi.org/10.1007/s10980-015-0225-1>.
- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int.* 26, 341–358. <https://doi.org/10.1080/10106049.2011.562309>.
- Butler, E., Howell, N., Guerrero, B., Mulamba, O., 2018. Enhancing crop acreage estimation within a semi-arid watershed via statistical assessments and techniques. *Agron. J.* 110, 2400–2407. <https://doi.org/10.2134/agronj2018.01.0065>.
- Calvin, K., Cowie, A., Berdes, G., Arneith, A., Cherubini, F., Portugal-Pereira, J., Grassi, G., House, J., Johnson, F.X., Popp, A., Rounsevell, M., Slade, R., Smith, P., 2021. Bioenergy for climate change mitigation: scale and sustainability. *GCB Bioenergy* 1–26. <https://doi.org/10.1111/gcbb.12863>.
- Carfagna, E., Carfagna, A., 2010. Alternative sampling frames and administrative data. What is the best data source for agricultural statistics? In: Benedetti, R., Bee, M., Espá, G., Piersimoni, F. (Eds.), *Agricultural Survey Methods*. John Wiley & Sons Ltd, Chichester, pp. 45–62. <https://doi.org/10.1002/9780470665480.ch3>.
- Congalton, R.G., 2001. Accuracy assessment and validation of remotely sensed and other spatial information. *Int. J. Wildl. Fire* 10, 321–328. <https://doi.org/10.1071/wf01031>.
- Congressional Research Service, 2020. *Renew. fuel Stand. (RFS): Overv. R43325*.
- Copenhaver, K., Hamada, Y., Mueller, S., Dunn, J.B., 2021. Examining the characteristics of the cropland data layer in the context of estimating land cover change. *ISPRS Int. J. Geo-Inf.* 10. <https://doi.org/10.3390/ijgi10050281>.
- Craig, M., 2005. Using FSA administrative data in the NASS cropland data layer. *NASS*. Craig, M., Atkinson, D., 2013. *A Literature Review of Crop Area Estimation*. Fao, pp. 1–22.
- Dahl, T.E., Dick, J., Swords, J., Wilen, B.O., 2020. *Data collection requirements and procedures for mapping wetland. Deep., Relat. Habitats U. S. (Version 3) undefined 91*.
- Debolini, M., Valette, E., François, M., Chéry, J.P., 2015. Mapping land use competition in the rural-urban fringe and future perspectives on land policies: a case study of Meknès (Morocco). *Land Use Policy* 47, 373–381. <https://doi.org/10.1016/j.landusepol.2015.01.035>.
- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabatte, T., Savinaud, M., Udrou, C., Valero, S., Bégué, A., Dejoux, J.F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneau, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* 221, 551–568. <https://doi.org/10.1016/j.rse.2018.11.007>.
- Deguines, N., Julliard, R., de Flores, M., Fontaine, C., 2012. The whereabouts of flower visitors: contrasting land-use preferences revealed by a country-wide survey based on citizen science. *PLoS One* 7, 1–9. <https://doi.org/10.1371/journal.pone.0045822>.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>.
- Estes, L.D., McRitchie, D., Choi, J., Debats, S., Evans, T., Guthe, W., Luo, D., Ragazzo, G., Zemleni, R., Caylor, K.K., 2016. A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environ. Model. Softw.* 80, 41–53. <https://doi.org/10.1016/j.envsoft.2016.01.011>.
- Faiza, B., Sophiayati Yuhani, S., Zaiton Mohd Hashim, S., Kalema AbdulRahman, K., 2012. A review and analysis of image misalignment problem in remote sensing. *Int. J. Sci. Eng. Res.* 3.
- Falcone, J.A., 2015. *U. S. Conte wall-to-wall Anthropog. Land Use Trends (NWALT) 1974–2012*.
- Gerling, M., Lawson, L., Wilson, E., 2015. Field Data Collection Using Geographic Information Systems Technologies and iPads on the USDA's June Area Frame Survey.
- Gomez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal classification of remote sensing images: a review and future directions. *Proc. IEEE* 103, 1560–1584. <https://doi.org/10.1109/JPROC.2015.2449668>.
- Good, D., 2014. Comparing NASS and FSA Planted Acreage Data farmdoc daily (4):9, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, January 21, 2014. <https://farmdocdaily.illinois.edu/2014/01/comparing-nass-fsa-planted-acreage-data.html>.
- Griffiths, P., Nendel, C., Hostert, P., 2019a. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* 220, 135–151. <https://doi.org/10.1016/j.rse.2018.10.031>.
- Griffiths, P., Nendel, C., Pickert, J., Hostert, P., 2019b. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2019.03.017>.
- Henrys, P.A., Jarvis, S.G., 2019. Integration of ground survey and remote sensing derived data: producing robust indicators of habitat extent and condition. *Ecol. Evol.* 9, 8104–8112. <https://doi.org/10.1002/ece3.5376>.
- Hoekman, S.K., Broch, A., 2018. Environmental implications of higher ethanol production and use in the U.S.: a literature review. Part II – Biodiversity, land use change, GHG emissions, and sustainability. *Renew. Sustain. Energy Rev.* 81, 3159–3177. <https://doi.org/10.1016/j.rser.2017.05.052>.
- Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a circa 2000 landcover database for the United States. *Photogramm. Eng. Remote Sens.* 70, 829–840.
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* 81, 345–354. [https://doi.org/10.1016/S0099-1112\(15\)30100-2](https://doi.org/10.1016/S0099-1112(15)30100-2).
- Homer, C.G., Fry, J.A., Barnes, C.A., 2012. The national land cover database. *USGS Fact Sheet* 2012-3020. <https://doi.org/10.3133/fs20123020>.
- Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., Gong, P., Yang, J., Bai, Y., Zheng, Y., Zhu, Z., 2017. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sens. Environ.* 202, 166–176. <https://doi.org/10.1016/j.rse.2017.02.021>.
- International Civil Aviation Organization, 2019. *CORSIA Methodol. Calc. Actual Life Cycle Emiss. Values*.
- Johansson, R., Efland, A., Coble, K., 2017. *Falling Response Rates USDA Crop Surv.: Why It Matters*.
- Johnston, C.A., 2013. Wetland losses due to row crop expansion in the dakota prairie pothole region. *Wetlands* 33, 175–182. <https://doi.org/10.1007/s13157-012-0365-x>.
- Khatami, R., Mountrakis, G., Stehman, S.V., 2017. Mapping per-pixel predicted accuracy of classified remote sensing images. *Remote Sens. Environ.* 191, 156–167. <https://doi.org/10.1016/j.rse.2017.01.025>.
- Knight, E.J., Kvaran, G., 2014. Landsat-8 operational land imager design, characterization and performance. *Remote Sens.* 6, 10286–10305. <https://doi.org/10.3390/rs6110286>.
- Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P.H., 2018. Regional scale mapping of grassland mowing frequency with Sentinel-2 time series. *Remote Sens.* 10. <https://doi.org/10.3390/rs10081221>.
- Lamas, A.C., Abreu, D.A., Arroway, P., Lopiano, K.K., Young, L.J., 2007. *Model. Misclassification June Area Surv.*
- Lark, T.J., Meghan Salmon, J., Gibbs, H.K., 2015. Cropland expansion outpaces agricultural and biofuel policies in the United States. *Environ. Res. Lett.* 10, 44003. <https://doi.org/10.1088/1748-9326/10/4/044003>.
- Lark, T.J., Mueller, R.M., Johnson, D.M., Gibbs, H.K., 2017. Measuring land-use and land-cover change using the U.S. department of agriculture's cropland data layer: Cautions and recommendations. *Int. J. Appl. Earth Obs. Geoinf.* 62, 224–235. <https://doi.org/10.1016/j.jag.2017.06.007>.
- Larsen, A.E., Hendrickson, B.T., Dedeic, N., MacDonald, A.J., 2015. Taken as a given: evaluating the accuracy of remotely sensed crop data in the USA. *Agric. Syst.* 141, 121–125. <https://doi.org/10.1016/j.jagsy.2015.10.008>.
- Laso Bayas, J.C., Lesiv, M., Waldner, F., Schucknecht, A., Duerauer, M., See, L., Fritz, S., Fraisl, D., Moorthy, I., McCallum, I., Perger, C., Danylo, O., Defourny, P., Gallego, J., Gilliams, S., Akhtar, I.U.H., Baishya, S.J., Baruah, M., Bungnamei, K., Campos, A., Changkakati, T., Cipriani, A., Das, Krishna, Das, Keemee, Das, I., Davis, K.F., Hazarika, P., Johnson, B.A., Malek, Z., Molinari, M.E., Panging, K., Pawe, C.K., Pérez-Hoyos, A., Sahariah, P.K., Sahariah, D., Saikia, A., Saikia, M., Schlesinger, P., Seidacaru, E., Singha, K., Wilson, J.W., 2017. A global reference database of crowdsourced cropland data collected using the Geo-Wiki platform. *Sci. Data* 4. <https://doi.org/10.1038/sdata.2017.136>.
- Lepitzki, J., Axsen, J., 2018. The role of a low carbon fuel standard in achieving long-term GHG reduction targets. *Energy Policy* 119, 423–440. <https://doi.org/10.1016/j.enpol.2018.03.067>.
- Li, C., Fultz, L.M., Moore-Kucera, J., Acosta-Martínez, V., Horita, J., Strauss, R., Zak, J., Calderón, F., Weindorf, D., 2017. Soil carbon sequestration potential in semi-arid grasslands in the conservation reserve program. *Geoderma* 294, 80–90. <https://doi.org/10.1016/j.geoderma.2017.01.032>.
- Ludwig, C., Walli, A., Schleicher, C., Weichselbaum, J., Riffler, M., 2019. A highly automated algorithm for wetland detection using multi-temporal optical satellite data. *Remote Sens. Environ.* 224, 333–351. <https://doi.org/10.1016/j.rse.2019.01.017>.
- Malins, C., Plevin, R., Edwards, R., 2020. How robust are reductions in modeled estimates from GTAP-BIO of the indirect land use change induced by conventional biofuels? *J. Clean. Prod.* 258, 120716. <https://doi.org/10.1016/j.jclepro.2020.120716>.

- Maxwell, A.E., Warner, T.A., Vanderbilt, B.C., Ramezan, C.A., 2017. Land cover classification and feature extraction from national agriculture imagery program (naip) orthoimagery: a review. *Photogramm. Eng. Remote Sens.* 83, 737–747. <https://doi.org/10.14358/PERS.83.10.737>.
- Morefield, P.E., Leduc, S.D., Clark, C.M., Iovanna, R., 2016. Grasslands, wetlands, and agriculture: the fate of land expiring from the conservation reserve program in the midwestern United States. *Environ. Res. Lett.* 11. <https://doi.org/10.1088/1748-9326/11/9/094005>.
- Mueller, R., Boryan, C., Seffrin, R., 2009. August. Data partnership synergy: The cropland data layer. In 2009 17th International Conference on Geoinformatics (pp. 1–6). IEEE.
- Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* 129, 122–131. <https://doi.org/10.1016/j.rse.2012.10.031>.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2014.02.015>.
- Rapinel, S., Fabre, E., Dufour, S., Arvor, D., Mony, C., Hubert-Moy, L., 2019a. Mapping potential, existing and efficient wetlands using free remote sensing data. *J. Environ. Manag.* 247, 829–839. <https://doi.org/10.1016/j.jenvman.2019.06.098>.
- Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A., Hubert-Moy, L., 2019b. Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. *Remote Sens. Environ.* 223, 115–129. <https://doi.org/10.1016/j.rse.2019.01.018>.
- Rogers, J.N., Incorporated, E., Stokes, B., Science, A., De, W., Dunn, J., Cai, H., Wu, M., 2017. Assess. Potential Prod. Econ. Environ. Impacts Resulting a Billion ton bioeconomy 110–128. <https://doi.org/10.1002/bbb>.
- Sanderson, M., Wedin, D., Tracy, B., 2009. Grassland: definition, origins, extent, and future. *Geography*.
- Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., Yu, T.H., 2008. Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. *Sci.* (80-.). 319, 1238–1240. <https://doi.org/10.1126/science.1151861>.
- Souza, G.M., Ballester, M.V.R., de Brito Cruz, C.H., Chum, H., Dale, B., Dale, V.H., Fernandes, E.C.M., Foust, T., Karp, A., Lynd, L., Maciel Filho, R., Milanez, A., Nigro, F., Osseweijer, P., Verdade, L.M., Victoria, R.L., Van der Wielen, L., 2017. The role of bioenergy in a climate-changing world. *Environ. Dev.* <https://doi.org/10.1016/j.envdev.2017.02.008>.
- Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. *Remote Sens. Environ.* 132, 202–211. <https://doi.org/10.1016/j.rse.2013.01.016>.
- 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. *J. Appl. Remote Sens.* 6, 063590 <https://doi.org/10.1117/1.jrs.6.063590>.
- Stout, D.J., 2007. Wetlands layer - national spatial data infrastructure: points of contact. *Tech. Revis.*
- Tarantino, C., Adamo, M., Pasquariello, G., Lovergine, F., Blonda, P., Tomaselli, V., 2012. 8-band image data processing of the worldview-2 satellite in a wide area of applications. In: Rustamov, R. (Ed.), *Earth Obs.*
- Tarantino, C., Adamo, M., Lucas, R., Blonda, P., 2016. Detection of changes in semi-natural grasslands by cross correlation analysis with WorldView-2 images and new Landsat 8 data. *Remote Sens. Environ.* 175, 65–72. <https://doi.org/10.1016/j.rse.2015.12.031>.
- U.S. Environmental Protection Agency. 2010. Regulation of Fuels and Fuel Additives: Changes to Renewable Fuel Standard Program. 40 CFR Part 80 [EPA-HQ-OAR-2005-0161; FRL-9112-3]. RIN 2060-A081.
- Updike, T., Comp, C., 2010. Radiometric Use of WorldView-2 Imagery. *Technical note. DigitalGlobe* 1–17.
- USDA-NASS, 2020. Guide to NASS Surveys [WWW Document]. URL https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/index.php.
- USDA-NASS, 2019a. June Area Surv. June 2019 Rep.
- USDA-NASS, 2019b. A-2 Append. A 2017 Census Agric. Methodol.
- USDA-NRCS, 2020. A statistical survey of land use and natural resource conditions and trends on U.S. non-Federal lands [WWW Document]. URL <https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/nra/nri/>.
- USDA-NRCS, 2018, 2015. *Natl. Resour. Inventory Summ. Report. Wash., Dc.*
- Walter, T., Zink, R., Laaha, G., Zaller, J.G., Heigl, F., 2018. Fox sightings in a city are related to certain land use classes and sociodemographics: Results from a citizen science project. *BMC Ecol.* 18, 1–11. <https://doi.org/10.1186/s12898-018-0207-7>.
- Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sens. Environ.* 204, 31–42. <https://doi.org/10.1016/j.rse.2017.10.046>.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. *Remote Sens. Environ.* 236, 111402 <https://doi.org/10.1016/j.rse.2019.111402>.
- Wickham, J., Stehman, S.V., Gass, L., Dewitz, J.A., Sorenson, D.G., Granneman, B.J., Poss, R.V., Baer, L.A., 2017. Thematic accuracy assessment of the 2011 national land cover database (NLCD). *Remote Sens. Environ.* 191, 328–341. <https://doi.org/10.1016/j.rse.2016.12.026>.
- Wickham, J., Stehman, S.V., Homer, C.G., 2018. Spatial patterns of the United States national land cover dataset (NLCD) land-cover change thematic accuracy (2001–2011). *Int. J. Remote Sens.* 39, 1729–1743. <https://doi.org/10.1080/01431161.2017.1410298>.
- Wickham, J., Stehman, S.V., Sorenson, D.G., Gass, L., Dewitz, J.A., 2021. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. *Remote Sens. Environ.* 257, 112357 <https://doi.org/10.1016/j.rse.2021.112357>.
- Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A., Wade, T.G., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens. Environ.* 130, 294–304. <https://doi.org/10.1016/j.rse.2012.12.001>.
- Wright, C.K., Wimberly, M.C., 2013. Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proc. Natl. Acad. Sci.* 110, 4134–4139. <https://doi.org/10.1073/pnas.1215404110>.
- Wright, C.K., Larson, B., Lark, T.J., Gibbs, H.K., 2017. Recent grassland losses are concentrated around U.S. ethanol refineries. *Environ. Res. Lett.* 12. <https://doi.org/10.1088/1748-9326/aa6446>.
- Wu, C., Du, B., Cui, X., Zhang, L., 2017. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. *Remote Sens. Environ.* 199, 241–255. <https://doi.org/10.1016/j.rse.2017.07.009>.
- Xie, Z., Phinn, S.R., Game, E.T., Pannell, D.J., Hobbs, R.J., Briggs, P.R., McDonald-Madden, E., 2019. Using Landsat observations (1988–2017) and Google Earth Engine to detect vegetation cover changes in rangelands - A first step towards identifying degraded lands for conservation. *Remote Sens. Environ.* 232, 111317 <https://doi.org/10.1016/j.rse.2019.111317>.