

Integration of remote sensing, county-level census, and machine learning for century-long regional cropland distribution data reconstruction

Jia Yang^{a,*}, Bo Tao^b, Hao Shi^c, Ying Ouyang^d, Shufen Pan^c, Wei Ren^b, Chaoqun Lu^e

^a Department of Forestry, Mississippi State University, Mississippi State, MS 39762, United States

^b Department of Plant and Soil Sciences, College of Agricultural, Food and Environment, University of Kentucky, Lexington, KY 40546, United States

^c International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36830, United States

^d USDA Forest Service, Center for Bottomland Hardwoods Research, Mississippi State, MS 39762, United States

^e Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, United States

ARTICLE INFO

Keywords:

Land use change
Cropland expansion
Afforestation
Ecosystem services
Logistic regression
Marginal croplands

ABSTRACT

The Lower Mississippi Alluvial Valley (LMAV) was home to about ten million hectare bottomland hardwood (BLH) forests in the Southern U.S. It experienced over 80 % area loss of the BLH forests in the past centuries and large-scale afforestation in recent decades. Due to the lack of a high-resolution cropland dataset, impacts of land use change (LUC) on the LMAV ecosystem services have not been fully understood. In this study, we developed a novel framework by integrating the machine learning algorithm, county-level agricultural census, and satellite-based cropland products to reconstruct the LMAV cropland distribution during 1850–2018 at a 30-m resolution. Results showed that the LMAV cropland area increased from $0.78 \times 10^4 \text{ km}^2$ in 1850 to $6.64 \times 10^4 \text{ km}^2$ in 1980 and then decreased to $6.16 \times 10^4 \text{ km}^2$ in 2018. Cropland expansion rate was the largest in the 1960s ($749 \text{ km}^2 \text{ yr}^{-1}$) but decreased rapidly thereafter, whereas cropland abandonment rate increased substantially in recent decades with the largest rate of $514 \text{ km}^2 \text{ yr}^{-1}$ in the 2010s. Our dataset has three notable features: (1) the depiction of fine spatial details, (2) the integration of the county-level census, and (3) the inclusion of a machine-learning algorithm trained by satellite-based land cover product. Most importantly, our dataset well captured the continuous increasing trend in cropland area from 1930–1960, which was misrepresented by other cropland datasets reconstructed from the state-level census. Our dataset would be important to accurately evaluate the impacts of historical deforestation and recent afforestation efforts on regional ecosystem services, attribute the observed hydrological changes to anthropogenic and natural driving factors, and investigate how the socio-economic factors control regional LUC pattern. Our framework and dataset are crucial to developing managerial and policy strategies for conserving natural resources and enhancing ecosystem services in the LMAV.

1. Introduction

Human activities have been recognized as the dominating factor altering land cover types on the earth's surface (Foley et al., 2005). Driven by increased demands of food and fiber, the global cropland area experienced a five-fold increase from ~ 3 million km^2 (2% of land area) to ~ 15 million km^2 (11 % of land area) in the past three centuries (Klein Goldewijk et al., 2011). Presently, over 30 % of the global natural vegetation has been cleared for agricultural use (crop cultivation and grazing) (Vitousek et al., 1997), most of which are distributed in the temperate and tropical regions in the Northern Hemisphere. These changes have significantly disturbed the terrestrial ecological, biogeochemical, and hydrological processes (Foley et al., 2005). Although food and fiber production increased substantially along with

cropland expansion, many other ecosystem services experienced considerable degradations, such as soil erosion, water pollution, aquifer decline, and enhanced greenhouse gas (GHG) emissions (Foley et al., 2005). To maintain and improve ecosystem services, there is a critical need to understand the long-term patterns of land use change (LUC) as well as its impacts on natural resources.

Several global-scale historical cropland datasets have been developed through combining remote sensing observations, historical agricultural census data, and socioeconomic records (e.g. Klein Goldewijk et al., 2011; Pongratz et al., 2008; Ramankutty and Foley, 1999a). These datasets are valuable to understand the long-term cropland variations and spatial distribution. They have been widely used to examine the global impacts of the LUC on GHG emissions (DeFries et al., 1999; Houghton et al., 2012; Tian et al., 2016), terrestrial hydrological

* Corresponding author.

E-mail address: jy576@msstate.edu (J. Yang).

<https://doi.org/10.1016/j.jag.2020.102151>

Received 3 January 2020; Received in revised form 11 April 2020; Accepted 30 April 2020

Available online 15 May 2020

0303-2434/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

processes (Claussen et al., 2001; Sterling et al., 2013), and climate conditions (Betts et al., 2007; Gibbard et al., 2005).

However, global LUC datasets usually have a relatively coarse spatial resolution (e.g., half-degree for the dataset in Ramankutty and Foley (1999a)). Applications of global LUC datasets in evaluating national-level and regional carbon emissions (such as US and China) led to substantial uncertainties due to the coarse delineation of land conversion (Yu et al., 2019) and the inconsistency with satellite data in temporal trend of croplands (Lu and Tian, 2017). Therefore, much effort has been made to develop regional cropland datasets for countries that experienced significant cropland expansion, such as China, India, and the Continental U.S. Comparing to global-scale datasets, regional cropland datasets usually utilize remote sensing data with a higher spatial resolution and/or a more detailed agricultural census. For instance, Liu and Tian (2010) reconstructed China's LUC pattern during 1700–2005 using the 30-m Landsat-based land cover data and provincial-level census data. Yu and Lu (2018) reconstructed cropland patterns in the Continental US from 1850 to 2016 using the 30-m remote sensing data and state-level census data. Tian et al. (2014) developed India's LUC data during 1880–2010 using the 56-m Resourcesat-1 remote sensing data and census data at the district and state levels. In contrast, the global cropland pattern in the Historical Database of the Global Environment (HYDE) 3.1 (Klein Goldewijk et al., 2011) was reconstructed from the DISCover data and the Global Land Cover 2000 at a 5-min resolution (~ 9 km at the equator) and state-level agricultural census data. Because the regional datasets provide more accurate temporal and spatial information of cropland distribution, its application could result in a more accurate evaluation of the LUC impacts on the national carbon budget (Sohl et al., 2012; Yu et al., 2019).

A series of approaches and models have been developed to reconstruct the spatially-explicit cropland distribution. At the local and landscape scales, land cover distribution and temporal changes have been simulated by the Multi-Agent Systems of Land-Use/Cover Change (MAS/LUCC) models through combining agent-based representation of decision-making and cellular landscape modeling (e.g. Le et al., 2008; Li and Yeh, 2002; Parker et al., 2003; Veldkamp and Verburg, 2004). LUC agent parameters include farm size and number, rural population and household, landowners, etc. (Parker et al., 2003), while the cellular model integrates spatially-explicit parameters, such as physical properties, neighbor functions, and distance variables, to simulate the spatial pattern of cropland expansion and abandonment (Li and Yeh, 2002). At the national, continental, and global scales, historical cropland patterns have been reconstructed through a combination of the satellite observations in the contemporary period and historical agricultural census data (e.g. Klein Goldewijk et al., 2011; Liu and Tian, 2010; Ramankutty and Foley, 1999a; Tian et al., 2014; Yu and Lu, 2018). The total cropland area in one administrative unit is allocated to grid cells according to a specific spatialization rule. For instance, Ramankutty and Foley (1999b) and Zumkehr and Campbell (2013) assumed that the cropland spatial pattern in each political unit in the historical period was the same as that observed by satellite in the contemporary period. Klein Goldewijk et al. (2011) allocated the total cropland area to grid cells by considering cropland distribution in the contemporary period as well as the influences of topography, climate, and soil condition on cropland distribution. Sohl et al. (2016) used the Forecasting Scenarios of Land use Change (FORE-SCE) model to define the relationship between satellite-observed land cover and various environmental factors and identify the locations of historical land use change patches.

The Lower Mississippi Alluvial Valley (LMAV) was once the largest floodplain in North America extending from Illinois in the north to the Gulf of Mexico in the south. In the pre-settlement era, the LMAV was home to 8–10 million ha highly productive Bottomland Hardwood (BLH) forests (Stanturf et al., 2000). Since the 19th century, large-scale deforestation has been widely implemented in the LMAV, which cleared

more than 80 % of the natural BLH forests (King et al., 2006). The LMAV has become one of the major agricultural regions in the U.S. for soybean and corn productions owing to its flat landscape, fertile soil, and favorite climate conditions (Yang et al., 2019). Nevertheless, in recent decades, the LMAV forest area has increased considerably due to afforestation efforts on the marginal croplands under the financial support of conservation and wildlife habitat easement programs (Faulkner et al., 2011; King et al., 2006; Oswalt, 2013). The historical cropland expansion and the recent afforestation activities in the LMAV likely have exerted strong influences on natural resources and environment, but the spatial pattern and long-term variations have not been well quantified yet. The LMAV is an ideal area to investigate the LUC impacts on the regional environment and ecosystem services. The significant changes in the LMAV cropland also provide a testbed for the development of novel algorithms to reconstruct high-resolution cropland data.

A fine-resolution cropland dataset (< 100 m) is needed in the LMAV. Previous global and national land use datasets are associated with coarse resolutions (such as half-degree and 5-min) (Klein Goldewijk et al., 2011) and are difficult to be used at the landscape and watershed scale. For example, the Big Sunflower watershed in the LMAV has a total area of 10,488 km² (Ouyang et al., 2018), which contains only 3 half-degree grids or ~ 116 5-min grids. Such coarse resolutions limit their application as input data to drive ecosystem or hydrological models to evaluate the long-term impacts of land use change. A fine-resolution cropland dataset could be aggregated to various resolutions and meet the needs of landscape-level and watershed-level research. In the LMAV, land use presents significant spatial variations strongly controlled by local flooding regimes and topography (Kaminski et al., 1989). A long-term cropland data with a fine resolution can reveal the pattern of cropland location shift over the historical period and be important to identify the underlying driving factors.

Specific objectives of this study include: (1) Develop a machine learning algorithm trained by contemporary cropland pattern to retrieve the historical cropland distribution; (2) Reconstruct a high-resolution (30 m) cropland dataset from 1850 to 2018 constrained by the county-level agricultural census; and (3) Reveal the historical patterns of cropland expansion and abandonment in the LMAV.

2. Data and methods

2.1. Study domain

The LMAV is the major agricultural region in the Southern U.S. residing along the Lower Mississippi River (Fig. 1). In this study, we extracted its boundary from the United States Environmental Protection Agency (EPA) Level III ecoregion data (Omernik, 1987). There are 115 counties (or parishes) completely within or partially included in the LMAV. According to the USDA National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) in 2010, the LMAV has a total land area of $\sim 11.7 \times 10^4$ km², consisting of cropland (52.8 %), wetland (35.5 %), developed area (5.1 %), grassland (3.4 %), and forest (3.2 %). The landscape is relatively flat with an average elevation of 41 m. During 1991–2010, the average temperature was 17.6 °C, and the average precipitation was 1363 mm yr⁻¹ (Yang et al., 2019).

2.2. Input datasets for cropland reconstruction

In this study, the cropland reconstruction algorithm (Section 2.3) needs five types of input datasets, namely, county-level agricultural census, satellite-based cropland distribution, topography, rivers and streams, and population (Table 1).

2.2.1. County-level agricultural census

The definition of cropland diverges in the existing literature (e.g. Yu

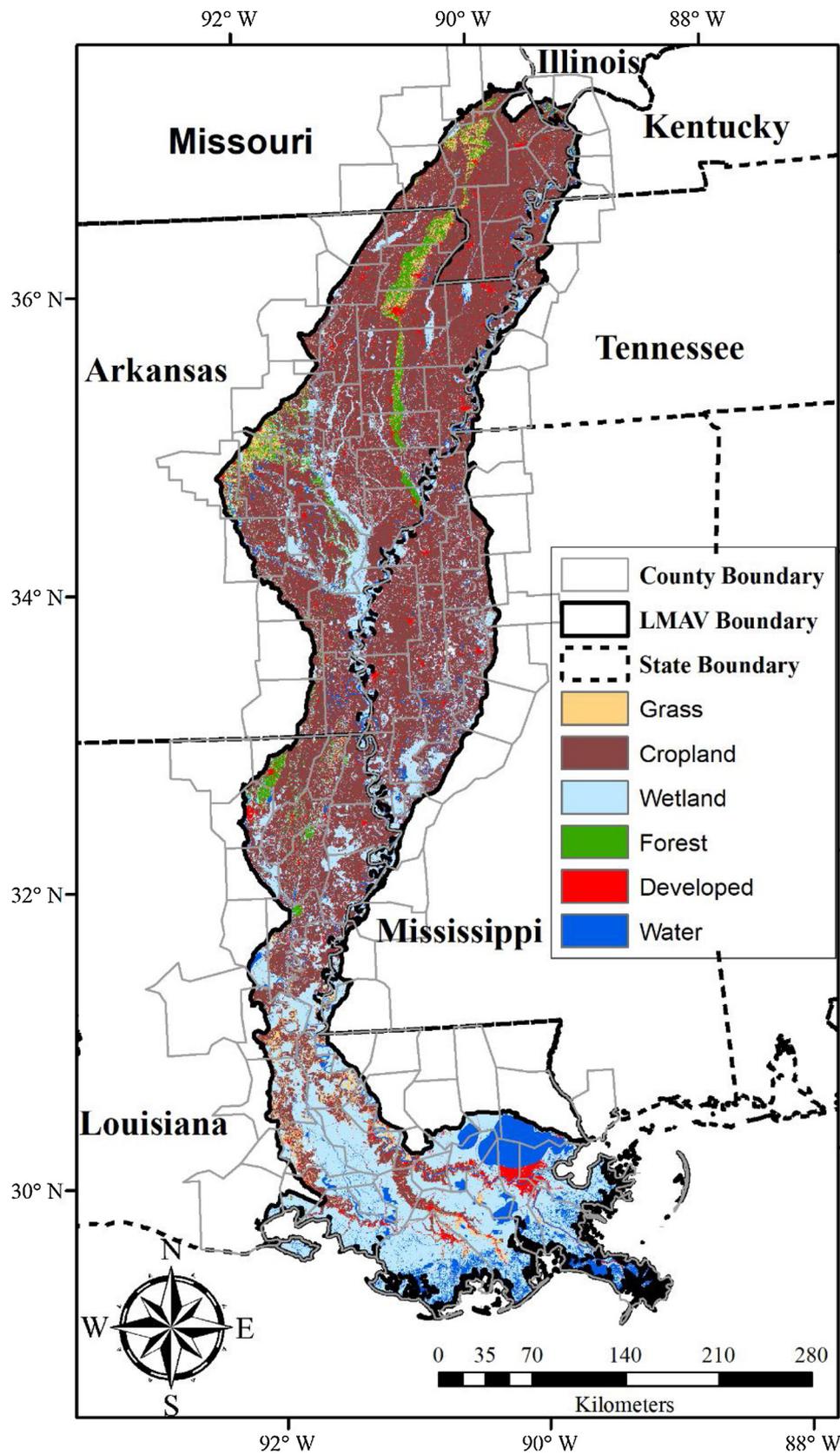


Fig. 1. The boundary of the Lower Mississippi Alluvial Valley and the cover type retrieved from USDA National Agricultural Statistics Service Cropland Data Layer 2010.

and Lu, 2018; Zumkehr and Campbell, 2013). In this study, we used the term defined by USDA (<https://www.ers.usda.gov/data-products/major-land-uses/glossary/>), in which cropland consists of five

sections: harvested cropland, crop failure, summer fallow, cropland pasture, and idle cropland. The county-level cropland areas from 1850 to 2017 were collected from three census datasets (Table 1). Cropland

Table 1
Input datasets for cropland reconstruction.

Data name	Period	Time interval	Data type	Website
USDA Census of Agriculture	2007 – 2017	5 years	County-level cropland area	https://www.nass.usda.gov/AgCensus/
USDA Census of Agriculture Historical Archive	1925 – 2002	4 to 5 years	County-level cropland area	http://agcensus.mannlib.cornell.edu/AgCensus/homepage.do
Historical, Demographic, Economic, and Social Data: The United States, 1790 – 2002	1850 – 1920	10 years	County-level improved farmland area	https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/02896
USDA National Agricultural Statistics Service Cropland Data Layer	2010–2018	Annual	Distribution of cropland and other land types	https://nassgeodata.gmu.edu/CropScape/
Shuttle Radar Topography Mission	2000	One time	Topography	https://dwtkns.com/srtm30m/
USA Rivers and Streams dataset	2006	One time	Rivers and Streams	http://geodata.utc.edu/datasets/0baca6c9f9d6499fb8e5fad50174c4e0_0
North America Rivers and Lakes	2006	One time	Rivers and Streams	https://www.sciencebase.gov/catalog/item/4fb55df0e4b04cb937751e02
Population density maps for the conterminous U.S.	1790 – 2010	Decadal	Population	http://doi.org/10.6084/m9.figshare.c.3890191

areas in 2007, 2012, and 2017 were from the USDA Census of Agriculture (CA), and cropland areas between 1925 and 2002 were from the USDA Census of Agriculture Historical Archive (CAHA). These data were available in every four or five years. From 1850 to 1920, the county-level improved farmland areas were available in every ten years, obtained from the Historical, Demographic, Economic, and Social Data: The United States, 1790 – 2002 (HDES-US) (Haines, 2010). Improved farmland, as defined by the USDA Census of Agriculture (http://usda.mannlib.cornell.edu/usda/AgCensusImages/1920/Farms_and_Property.pdf), includes all tilled or mowed land, tilled pastureland, fallow land, land in garden, nurseries, orchards, vineyards, and land occupied by farm buildings. As the land area occupied by farm buildings was much smaller than the other farmland, we assumed that the area of the improved farmland was comparable to the cropland area defined by USDA. Historical cropland census data from the three sources were merged, and the annual cropland area from 1850 to 2017 was linearly interpolated between two census periods. Fig. S1 shows the inter-annual variations of the total cropland area in the 115 counties from 1850 to 2017. Fig. 2 shows the spatial pattern of the county-level cropland area in 1850, 1890, 1930, 1970, and 2010.

2.2.2. Satellite-based cropland distribution

The USDA NASS CDL product is a geo-referenced and crop-specific land cover map (Boryan et al., 2011). The CDL was developed based on the 30-m Landsat satellite imageries in conjunction with various ground truth datasets. It represents the best cropland distribution product in the U.S. and provides the most detailed spatial information for more than 100 crop categories at a spatial resolution of 30 m (Johnson, 2013). In this study, the LMAV cropland distribution during 2010–2018 was directly retrieved from the CDL product. The CDL 2010 cropland distribution was used as the reference data to train the machine learning algorithm.

2.2.3. Ancillary datasets

Elevation, slope, distance to rivers and streams, and population were important parameters for the machine learning algorithm to estimate cropland probability at the pixel level (Sohl and Sayler, 2008). The 30-m Digital Elevation Model (DEM) was obtained from NASA's Shuttle Radar Topography Mission (SRTM) (Farr et al., 2007). Slope was computed from the SRTM DEM using the spatial analysis tools in ESRI ArcGIS 10.6. Rivers and streams were extracted from the USA Rivers and Streams dataset (USA-RS) and the North America Rivers and Lakes (NARL). For each 30-m pixel, distance to the nearest stream or river channel was calculated through the spatial analysis tools in ESRI ArcGIS 10.6. Additionally, water body pixels were extracted from the NASS CDL 2010 dataset. Fig. S2 shows the maps of DEM, slope, distance to the nearest river and stream, and water body in the LMAV.

Fang and Jawitz (2018) developed a 1-km population density dataset for the Continental U.S. from 1790 to 2010 using five population models. In this study, we used the population data produced using the most sophisticated model (M5), which considered natural suitability, socioeconomic desirability, and inhabitability parameters. The decadal population data (Fang and Jawitz, 2018) was downscaled into an annual scale through a temporal linear interpolation and then to a 30-m resolution through the bilinear spatial interpolation.

2.3. Workflow and algorithms

Fig. 3 illustrates the workflow and three major algorithms to develop the LMAV cropland dataset from 1850 to 2018. The three algorithms are: (1) a reconciliation algorithm to modify the county-level agricultural census to be consistent with the CDL cropland area; (2) a logistic regression-based machine learning algorithm to estimate annual cropland probability; and (3) a land conversion algorithm using cropland probability and the reconciled cropland area to guide the placement of cropland expansion and abandonment.

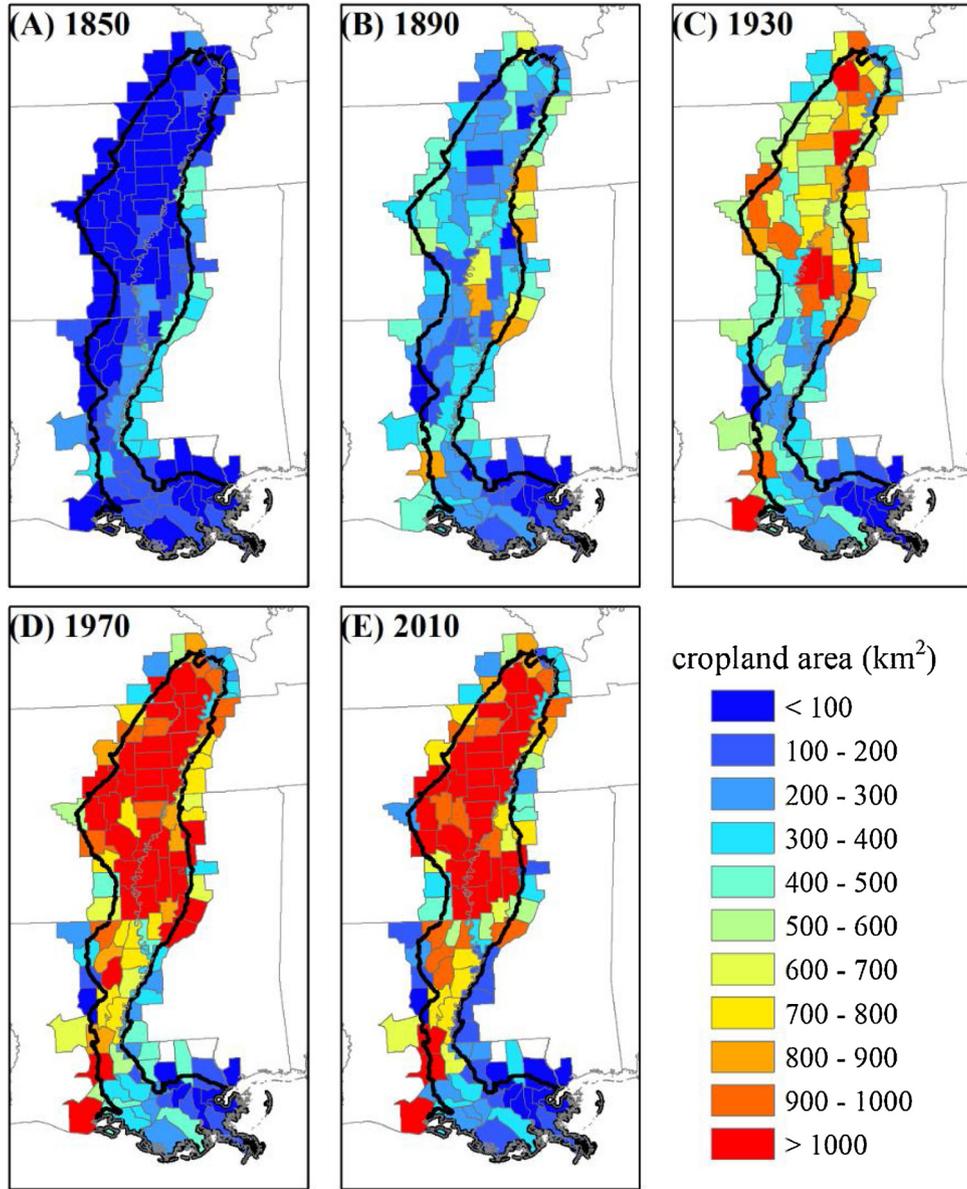


Fig. 2. Cropland area (km²) in 1850, 1890, 1930, 1970, and 2010 from the county-level census datasets.

2.3.1. Reconciliation between census data and remote sensing-based data

A gap existed between the county-level cropland area from the Census of Agriculture and the cropland area estimated from the CDL data. For example, for the 115 counties in 2012, the cropland area was 7.35×10^4 km² in the census data, but 6.79×10^4 km² in the CDL data. To fill this gap, we used the reconciliation algorithm in Ramankutty and Foley (1999b) to modify the county-level census. The reconciled cropland area was computed as a weighted average of the estimates based on relative anomalies and absolute anomalies of the inventory data. The reconciliation algorithm is,

$$A_{reconciled}^{t2}(k) = \alpha(k) \left[A_{reconciled}^{t1}(k) \frac{A_{census}^{t2}(k)}{A_{census}^{t1}(k)} \right] + (1 - \alpha(k)) [A_{reconciled}^{t1}(k) - (A_{census}^{t2}(k) - A_{census}^{t1}(k))] \quad (1)$$

where $t1$ is the current year; $t2$ is the previous year ($t2 = t1 - 1$); $A_{census}^{t1}(k)$ is the census-based cropland area in year $t1$ for the k th county; $A_{census}^{t2}(k)$ is the census-based cropland area in year $t2$ for the k th county; $A_{reconciled}^{t1}(k)$ is the reconciled cropland area in year $t1$ for the k th county; and $A_{reconciled}^{t2}(k)$ is the reconciled cropland area in year $t2$ for

the k th county. k is the county index in the range from 1 to 115. $\alpha(k)$ is the weight of the relative anomaly ($\frac{A_{census}^{t2}(k)}{A_{census}^{t1}(k)}$), while $1 - \alpha(k)$ is the weight of the absolute anomaly ($A_{census}^{t2}(k) - A_{census}^{t1}(k)$). $\alpha(k)$ is parameterized as

$$\alpha(k) = \min(1, \exp(-0.5(A_{census}^{t2}(k)/A_{census}^{t1}(k) - 1.1))) \quad (2)$$

The calculation started from 2010 ($t1 = 2010$). In 2010, the reconciled county-level cropland area $A_{reconciled}^{t1}(k)$ was retrieved from the CDL 2010. The reconciled county-level cropland area from 2009 to 1850, $A_{reconciled}^{t2}(k)$, was computed iteratively by solving Eqs. (1) and (2) with $t1$ moving from 2010 to 1851 by one year.

2.3.2. Reconstruction of historical cropland distribution

Cropland distribution from 1850–2009 was reconstructed according to the reconciled county-level cropland area and pixel cropland probability-of-occurrence (cropland probability for short). The cropland probability was estimated through logistic regression (Sohl and Saylor, 2008). In the LMAV, farmers tended to reclaim the forests and wetlands with low economic and labor costs, and abandon the agricultural lands with higher flooding frequency and lower crop production (i.e.,

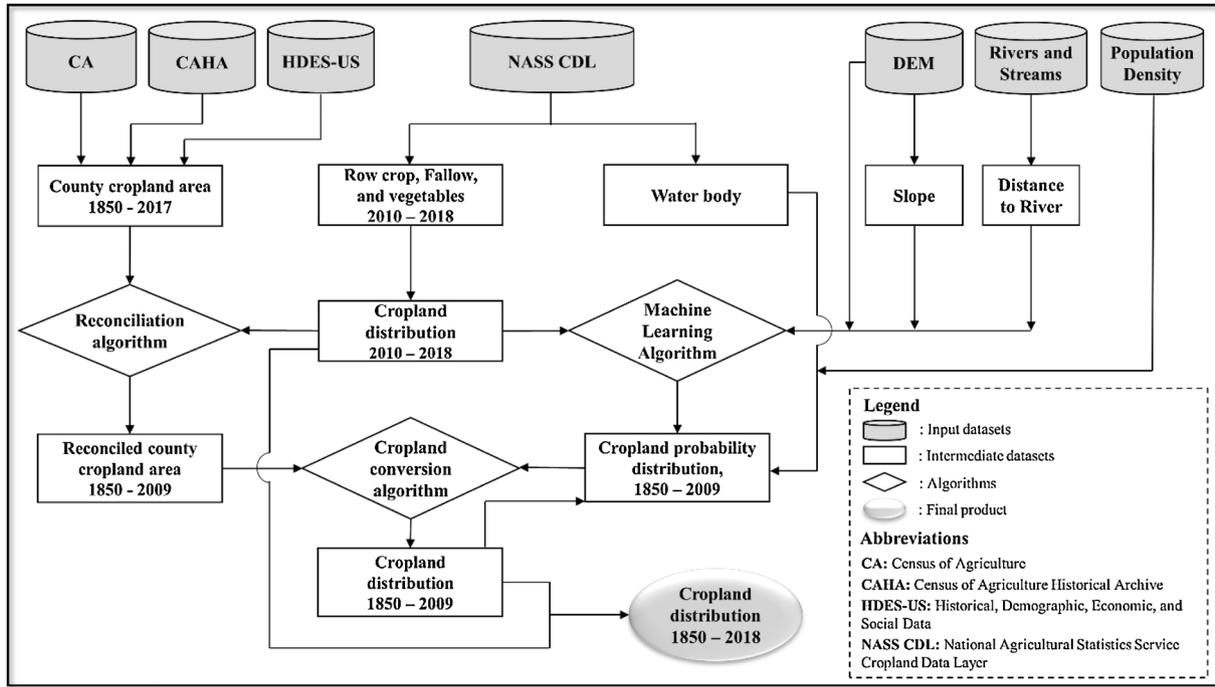


Fig. 3. Input datasets and workflow for the development of annual cropland dataset in the Lower Mississippi Alluvial Valley from 1850 to 2018.

marginal croplands) (Schoenholtz et al., 2001; Stanturf et al., 1998). Based on these situations, we made the following three basic assumptions:

- (1) Cropland probability in each pixel was determined by topography, flooding regimes, distance to cropland edge, and population density.
- (2) During cropland area increase years, the natural vegetation pixels with higher cropland probability (e.g., flat topography, low flooding frequency, and close to existing cropland) were converted to cropland earlier than the natural vegetation pixels with lower cropland probability (e.g., steep topography, high flooding frequency, and far from existing cropland).
- (3) During cropland area decrease years, the cropland pixels with high cropland probability (e.g., flat topography, low flooding frequency, and far from cropland edge) were abandoned later than the cropland pixels with lower cropland probability (e.g., steep topography, high flooding frequency, and close to cropland edge).

Reconstructing the LMAV cropland consisted of the following five steps.

Step 1. Develop and train a county-wise machine learning algorithm in the form of logistic regression. The logistic regression model is expressed as

$$\log_{10} \frac{Prob_{k,i}^{t1}}{1 - Prob_{k,i}^{t1}} = \beta_{0,k} + \beta_{1,k} DEM_{k,i} + \beta_{2,k} SLO_{k,i} + \beta_{2,k} DIS_{k,i} + \beta_{3,k} NCP_{k,i}^{t1} \quad (3)$$

where $Prob_{k,i}^{t1}$ is the cropland probability between 0 and 1 for pixel i in the k th county ($k = 1, 2, \dots, 115$); $t1$ equals 2010 in this step; $DEM_{k,i}$, $SLO_{k,i}$, and $DIS_{k,i}$ are the DEM, slope, and distance to the nearest river and streams for pixel i in the k th county; $NCP_{k,i}^{t1}$ is the number of cropland pixels in a pixel i centered circle with a radius of 90 m in year $t1$; and $\beta_{0,k}$, $\beta_{1,k}$, $\beta_{2,k}$, and $\beta_{3,k}$ are four county-specific parameters in the regression model. $DEM_{k,i}$ and $SLO_{k,i}$ are topography variables, and $DIS_{k,i}$ is a proxy for flooding frequency. $NCP_{k,i}^{t1}$ is an indicator of cropland edge, which was computed using the neighborhood operation in ESRI ArcGIS 10.6.

We used NASS CDL cropland distribution in 2010 as training data to fit the parameters. Additionally, we also used NASS CDL 2011, 2012, and 2013 to fit the parameters in each county, which were similar to that based on NASS CDL 2010. Therefore, we simply used the parameters derived from NASS CDL 2010. Cropland probability equaled 1 for cropland pixels ($Prob_{k,i}^{t1} = 1$) and 0 for non-cropland pixels ($Prob_{k,i}^{t1} = 0$). The General Linear Model (GLM) function in the Python statsmodels library (Seabold and Perktold, 2010) was adopted to estimate the parameters of $\beta_{0,k}$, $\beta_{1,k}$, $\beta_{2,k}$, and $\beta_{3,k}$ in the logistic regression. The statistics had a relatively low variance inflation factor (VIF), indicating the low collinearity among the four selected factors in the logistic regression.

Step 2. Estimate the pixel-based cropland probability in year $t1$ ($t1$ was 2010 at the first iteration).

$$Prob_{k,i}^{t1} = 1 - 1 / (1 + 10^{\beta_{0,k} + \beta_{1,k} DEM_{k,i} + \beta_{2,k} SLO_{k,i} + \beta_{2,k} DIS_{k,i} + \beta_{3,k} NCP_{k,i}^{t1}}) \quad (4)$$

Cropland probability ($Prob_{k,i}^{t1}$) was estimated using the four parameters ($\beta_{0,k}$, $\beta_{1,k}$, $\beta_{2,k}$, and $\beta_{3,k}$) fitted in Step 1. $NCP_{k,i}^{t1}$ was updated annually based on the cropland distribution map in year $t1$. $Prob_{k,i}^{t1}$ was set to 0 for the water body pixels. We assumed the pixels with a dense population (usually urban area) had no space for crop cultivation, and $Prob_{k,i}^{t1}$ was set to 0 for pixels with population density over 1000 people / km^2 .

Step 3. Determine the placement of cropland area changes between the previous year ($t2 = t1 - 1$) and the current year ($t1$, $t1$ was 2010 at the first iteration). County-level cropland area difference between year $t2$ and year $t1$ in the k th county ($\Delta A^{t2 \rightarrow t1}(k)$) was calculated using the reconciled cropland area,

$$\Delta A^{t2 \rightarrow t1}(k) = A_{reconciled}^{t2}(k) - A_{reconciled}^{t1}(k) \quad (5)$$

Positive $\Delta A^{t2 \rightarrow t1}(k)$ indicates cropland area decrease in the k th county from year $t2$ to year $t1$, while negative value indicates cropland area increase. The number of pixels in the k th county that experienced land conversion between cropland and non-cropland ($NP_k^{t2 \rightarrow t1}$) was calculated as,

$$NP_k^{t2 \rightarrow t1} = round\left(\frac{abs(\Delta A^{t2 \rightarrow t1}(k))}{Ap}\right) \quad (6)$$

where A_p is the pixel area ($= 900 \text{ m}^2$), and the $NP_k^{t2 \rightarrow t1}$ is the nearest integer to $\frac{abs(\Delta A^{t2 \rightarrow t1}(k))}{A_p}$.

For the k^{th} county in year $t1$, two sequences of cropland probability were extracted from the $Prob_{k,i}^{t1}$ map calculated in Step 2. The first sequence was the cropland probability for all the cropland pixels, and the second sequence was the cropland probability for all the non-cropland pixels. Both sequences were sorted in ascending order.

We used the pixel value of 1 to indicate cropland ($PV_{k,i}^{t1} = 1$), and 0 to indicate non-cropland ($PV_{k,i}^{t1} = 0$). When $\Delta A^{t2 \rightarrow t1}(k) > 0$ (cropland area in $t2$ was larger than that in $t1$), $NP_k^{t2 \rightarrow t1}$ non-cropland pixels with the highest cropland probability in the non-cropland sequence were selected and changed to cropland pixels in year $t2$ ($PV_{k,i}^{t2} = 1$); the non-cropland pixels that were not selected still had a pixel value of 0 in year $t2$ ($PV_{k,i}^{t2} = 0$); and all the cropland pixels in year $t1$ were still cropland pixels in year $t2$ ($PV_{k,i}^{t2} = 1$).

When $\Delta A^{t2 \rightarrow t1}(k) < 0$ (cropland area in $t2$ was smaller than that in $t1$), $NP_k^{t2 \rightarrow t1}$ cropland pixels with the lowest cropland probability in the cropland sequence were selected and changed to non-cropland pixels in year $t2$ ($PV_{k,i}^{t2} = 0$); the cropland pixels that were not selected still had a pixel value of 1 in year $t2$ ($PV_{k,i}^{t2} = 1$); and all the non-cropland pixels in year $t1$ were still non-cropland pixels in year $t2$ ($PV_{k,i}^{t2} = 0$). Through this cropland conversion algorithm, locations of cropland changes were determined and county-level cropland area in year $t2$ was distributed to pixels.

Finally, cropland spatial pattern in the 115 counties in year $t2$ was merged into one dataset to represent cropland spatial pattern in the LMAV.

Step 4. Iterate the processes in Step 2 and Step 3 with $t1$ decreased by one at each iteration until $t1 = 1851$ and $t2 = 1850$.

Step 5. Combine the reconstructed cropland data (1850–2009) with the NASS CDL data (2018–2018). The final product is the LMAV cropland distribution from 1850 to 2018 at a 30-m spatial resolution.

2.4. Tipping point detection

Piece-wise regression has been used to detect ecological and climatological breakpoints (Toms and Lesperance, 2003; Wang et al., 2011). In this study, we used this approach to identify the tipping point in the time series of cropland area in each county and the entire LMAV.

$$y = \begin{cases} \gamma_0 + \gamma_1 x, & x \leq \alpha \\ \gamma_0 + \gamma_1 x + \gamma_2 (x - \alpha), & x > \alpha \end{cases} \quad (7)$$

where y is the annual cropland area, x is the year from 1850–2018, α is the breaking point year, γ_0 , γ_1 , and γ_2 are the parameters in the fitted lines. The time series of cropland area was divided into two periods at the breaking point, period I: 1850 to year α and period II: year α to 2018. The parameters γ_0 and γ_1 were estimated through the ordinary least squares (OLS) regression in the period I, and γ_2 was estimated through the OLS regression in the period II. γ_1 was the linear trend of cropland area changes before the breaking point, and $(\gamma_1 + \gamma_2)$ was the linear trend after it. The two fitted lines were joined at the breaking point.

Table 2
Global- and national-scale cropland datasets for comparison.

Data name	Period	Time interval	Extent	Resolution	Website
RF dataset	1700 – 1992	Annual	Global	0.5 arc degree	https://nelson.wisc.edu/sage/data-and-models/datasets.php
HYDE 3.2	10,000 BC to 2017 AD	Millennial between 10,000 BC and 0 AD Centennial between 0 and 1000 Decadal between 1000 and 2000 Annual between 2000 and 2017	Global	5 arc min	ftp://ftp.pbl.nl/hyde/
ZC dataset	1850 – 2000	Annual	Continental U.S.	5 arc min	https://portal.nersc.gov/project/m2319/
YL dataset	1850 – 2016	Annual	Continental U.S.	1 km	https://doi.pangaea.de/10.1594/PANGAEA.881801
This study	1850–2018	Annual	The LMAV	30 m	Will be uploaded to public platform for download

α moved from 1855 to 2013 by one in each step. The range of 1855–2013 for α was defined to avoid regression with too few data points in the period I or period II (Wang et al., 2011). For each, we estimated a set of γ_0 , γ_1 , and γ_2 , and computed the sum of squared residuals (SSR),

$$SSR = \sum_{i=1850}^{2018} (y_i - \hat{y}_i)^2 \quad (8)$$

where \hat{y}_i is the fitted annual cropland area by the piece-wise linear regression. In total, we calculated 159 SSR with α moving from 1855 to 2013. The tipping point was the breaking point with the least R . Additionally, we used the student t -test to examine if the tipping point was significant, i.e., if the changing trend before the tipping point (γ_1) was significantly different from the changing trend after it (i.e., $\gamma_1 + \gamma_2$) at the 95 % confidence interval. The null hypothesis (H0) was $\gamma_2 = 0$.

According to the changing trends before and after the tipping point, we defined six types of tipping point, which were Type 1: $\gamma_1 > 0$, $\gamma_1 + \gamma_2 > 0$, and $\gamma_1 > \gamma_1 + \gamma_2$; Type 2: $\gamma_1 > 0$, $\gamma_1 + \gamma_2 > 0$, and $\gamma_1 < \gamma_1 + \gamma_2$; Type 3: $\gamma_1 > 0$ and $\gamma_1 + \gamma_2 < 0$; Type 4: $\gamma_1 < 0$, $\gamma_1 + \gamma_2 < 0$, and $\gamma_1 < \gamma_1 + \gamma_2$; Type 5: $\gamma_1 < 0$, $\gamma_1 + \gamma_2 < 0$, and $\gamma_1 > \gamma_1 + \gamma_2$; and Type 6: $\gamma_1 < 0$ and $\gamma_1 + \gamma_2 > 0$.

In this study, we simply used one tipping point to separate the time-series of cropland area into two periods. From 1850 to 2018, most of the LMAV counties experienced fast cropland expansion in the early period followed by the decelerated cropland expansion or even cropland area shrinkage (King et al., 2006). We acknowledge that some counties may have multiple tipping points in cropland area time-series, and the SSR could include multiple local maxima. Nonetheless, to keep it consistent, we used the global maximum to identify only one most significant tipping point and break the time-series into two most contrasting periods for all the counties and the LMAV.

2.5. Cropland datasets for comparison

Four sets of global- or national-level cropland datasets developed under the constraint of agricultural census (Table 2) were chosen to compare with the LMAV cropland dataset in this study. The four datasets were (1) global historic cropland dataset by Ramankutty and Foley (1999a) (named RF dataset), (2) the History Database of the Global Environment (HYDE 3.2) (Klein Goldewijk et al., 2017), (3) historical U.S. cropland area by Zumkehr and Campbell (2013) (named ZC dataset), and (4) another historical U.S. cropland area by Yu and Lu (2018) (named YL dataset).

The RF and HYDE3.2 datasets provided global cropland fractions at a resolution of 0.5° and 5-minute latitude/longitude, respectively. Both datasets used an AVHRR-based cropland distribution (Loveland and Belward, 1997) as the reference map and state-level census data to constrain the state total cropland area. For the comparison purpose, we downscaled the decadal HYDE data between 1850 and 2000 to an annual scale through the temporal linear interpolation. ZC and YL datasets provided historic cropland fraction in the Continental U.S. at a spatial resolution of 5-minute latitude/longitude and 1 km, respectively. The

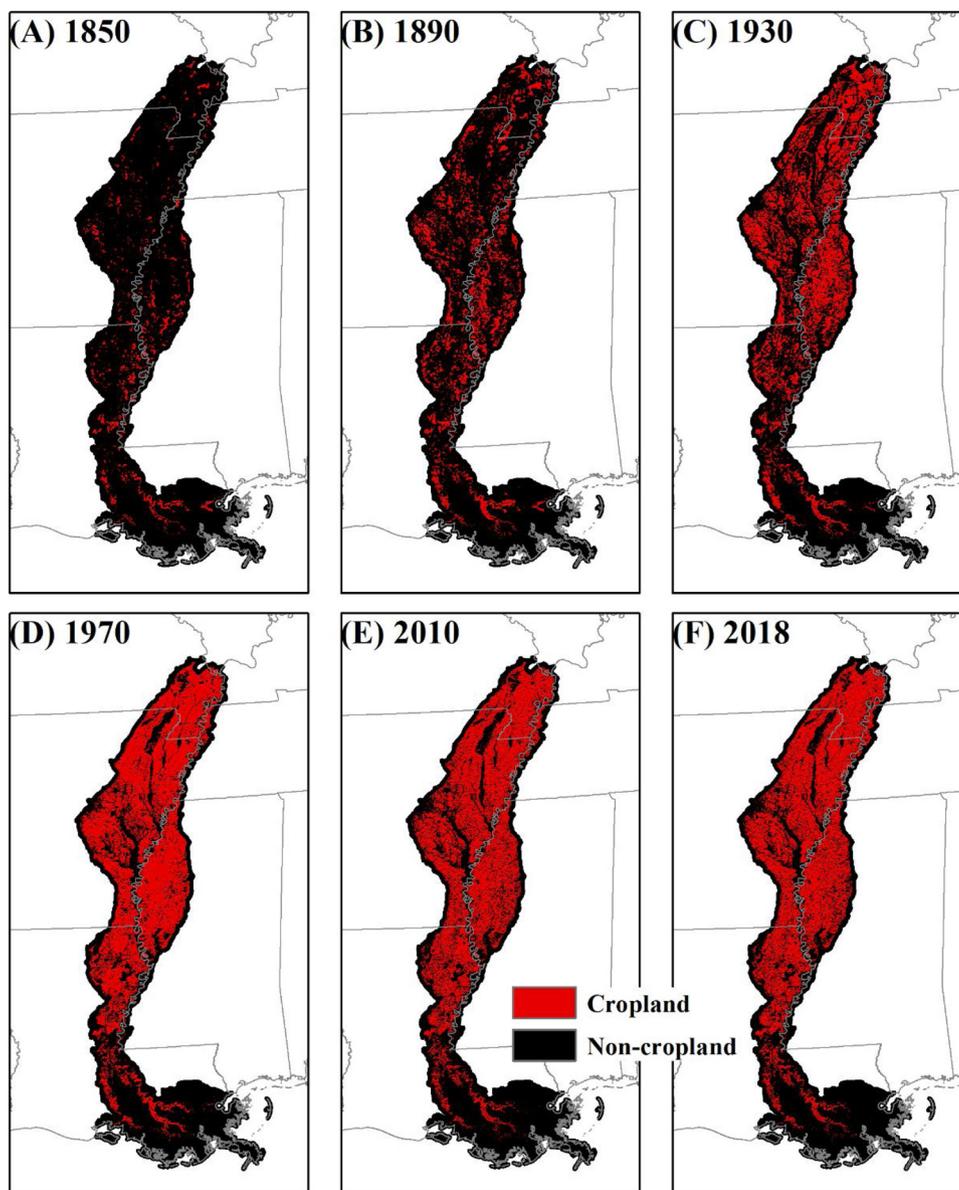


Fig. 4. Reconstructed cropland distribution across the Lower Mississippi Alluvial Valley in 1850, 1890, 1930, 1970, 2010, and 2018.

ZC dataset was reconstructed from the DISCover cropland distribution and county-level cropland area (Waisanen and Bliss, 2002), while the YL dataset was developed from the NASS CDL 2010 data and state-level census data.

2.6. Latitude vs. time plot

Latitude vs. Time plot of the cropland area was produced to show temporal variations of cropland area across latitudinal bands. First, the 30-m cropland maps were aggregated to cropland fraction pixels at a 5000-m resolution and converted to the Geographic Coordinate System at a resolution of 0.05° latitude/longitude. Next, cropland fraction in each 0.05° latitudinal band was averaged to obtain the zonal-mean cropland fraction. During this process, the three-dimensional cropland dataset (time × latitude × longitude) was compressed into a two-dimensional cropland dataset (time × latitude). Last, the time series of the zonal-mean cropland fraction in each latitudinal band was normalized to the range between 0 and 1 by the minimum and maximum annual cropland fractions during 1850–2018.

2.7. Definition of cropland expansion and abandonment

In previous literatures, abandoned cropland is defined as the cropland difference between the maximum cropland area in the historical period and the cropland area in the most recent period (Zumkehr and Campbell, 2013); while cropland expansion is defined as the cropland difference between the beginning and end of the study period (Yu and Lu, 2016). Following these definitions, we defines the pixel-level cropland expansion and abandonment in this study: cropland expansion over the entire period refers to the pixels that were non-cropland in 1850 but cropland in 2018, while cropland abandonment over the entire period referred to the pixels that were cropland in any year during 1850–2017 but non-cropland in 2018. Expansion pixels in each decade were a part of the expansion pixels over the entire period that experienced a land conversion from non-cropland to cropland in that decade. Abandonment pixels in each decade were a part of the abandonment pixels over the entire period that experienced land conversion from cropland to non-cropland in that decade.

It is noteworthy that the croplands, which were once abandoned but converted back to croplands later, were not counted as cropland

abandonment in this study. Therefore, the difference between cropland expansion and abandonment over a certain period (for example, the 1860s) does not necessarily equal the net cropland area change during that period estimated through the agricultural census data. The magnitudes, occurrence periods, and locations of cropland abandonment and expansion were retrieved from the reconstructed LMAV cropland dataset.

2.8. Validation method

The National Land Cover Database (NLCD) provides a national-wide land use and land cover dataset at a 30-m spatial resolution developed based on the Landsat imageries (Homer et al., 2015). The latest NLCD version 2016 provides a multi-temporal LUC dataset in the Continental U.S. in every two to three years from 2001 to 2016 (<https://www.mrlc.gov/data/nlcd-2016-land-cover-conus>). In this study, we used the land cover map in 2001 from the NLCD version 2016 to validate the reconstructed cropland dataset. For validation purposes, the 15 NLCD land cover types were regrouped into cropland type and non-cropland type. Then, we generated 5000 points randomly distributed in the LMAV (Fig. S3) using ESRI ArcGIS 10.6 sampling tool. For these 5000 points, we extracted land cover types from the reconstructed raster cropland dataset and compared them with those extracted from the NLCD dataset. A confusion matrix was built to estimate the overall accuracy, user's accuracy, and producer's accuracy for the reconstructed dataset in the LMAV.

We also validated the reconstructed cropland distribution with the NASS CDL in 2017. This year was selected because it is the last year of the current USDA Census of Agricultural. The same 5000 points were used to extract land cover types from our reconstructed data and NASS CDL, and a confusion matrix was created to quantify data accuracy.

3. Results

3.1. Long-term variations in cropland area

Historical patterns of the LMAV croplands (Fig. 4) were reconstructed through the new algorithms and county-level census data. Our results suggest that the LMAV had a total cropland area of $0.78 \times 10^4 \text{ km}^2$ in 1850 with more croplands distributed in the south part of LMAV. Louisiana (referring to the part of Louisiana included in the LMAV, the same for the other states) made the largest contribution (44.8 %) to the LMAV cropland area, followed by Mississippi (28.2 %) and Arkansas (19.8 %) (Fig. 5). Notably, the cropland area in 1870 was less than that in 1860 possibly due to the influence of the American Civil War in the 1860s. After the 1870s, the cropland area showed a continuous increase until the 1970s. Over the entire study period, the maximum cropland area was $6.71 \times 10^4 \text{ km}^2$ in 1978. In this year, Arkansas, Mississippi, and Louisiana made the largest contribution to the total cropland area, which was 42.4 %, 22.3 %, and 20.8 %, respectively (Fig. 5). Between 1850 and 1978, the cropland area increased by ~ 17 times in Arkansas, ~ 6 times in Mississippi, and ~ 3 times in Louisiana.

Over the recent four decades, the LMAV cropland area presented a declining trend. Cropland area in 2018 was $6.16 \times 10^4 \text{ km}^2$, which was $\sim 8\%$ less than that in 1978. Cropland area decreased by 13.9 % in Mississippi, 10.5 % in Louisiana, and 6.6 % in Arkansas. Over the 169 years from 1850–2018, Louisiana's contribution to the total LMAV cropland area decreased from 44.8%–20.3%, and Mississippi's contribution decreased from 28.2%–20.9% (Fig. 5), whereas Arkansas' contribution increased from 19.8%–43.1% and Missouri's contribution increased from 5.4%–13.3%.

The Latitude vs. Time plot shows that the cropland area presented distinct temporal patterns across the latitudinal gradient (Fig. 6). In southern LMAV ($< 31^\circ \text{N}$), the peaked cropland area occurred before the 1960s. In central LMAV (between 31°N and 36°N), the peaked

cropland area was in the 1970s, the 1980s, and the 1990s. In northern LMAV ($> 36^\circ \text{N}$), the peaked cropland area occurred in the 1990s and the 2000s. The distinct temporal pattern along the latitudinal gradient revealed that cropland expansion in central and northern LMAV lasted a longer time until the recent decades, while cropland abandonment started earlier in southern LMAV.

3.2. Tipping points

We fitted the piece-wise linear regression to the time series of the LMAV cropland area (Fig. 7). The SSR reached its lowest value of $6.14 \times 10^4 \text{ km}^2$ when $\alpha = 1980$, indicating the tipping point (Type-3) was in 1980. The LMAV cropland area in this year was $6.64 \times 10^4 \text{ km}^2$. Cropland area increased at a rate of $0.048 \times 10^4 \text{ km}^2 \text{ yr}^{-1}$ ($\gamma_1 = 0.048$) before the tipping point (i.e., 1850–1980), and decreased at a rate of $0.015 \times 10^4 \text{ km}^2 \text{ yr}^{-1}$ ($\gamma_1 + \gamma_2 = -0.015$) after it (i.e., 1980–2018). Of the seven states in the LMAV, five states had the Type-3 tipping point, which were Louisiana ($\alpha = 1992$), Mississippi ($\alpha = 1973$), Arkansas ($\alpha = 1980$), Illinois ($\alpha = 1915$), and Tennessee ($\alpha = 1969$) (Fig. S4). However, Missouri ($\alpha = 1969$) and Kentucky ($\alpha = 1917$) had the Type-1 tipping point, i.e., the cropland expansion slowed down after the tipping point comparing to the periods before the tipping point.

Additionally, the tipping point detection algorithm was implemented on the time-series of cropland area in the 115 counties (Fig. 8). We found that 80 counties had the type-3 tipping point, and 20 counties had the type-1 tipping point. For these 100 counties, the initial increasing trend of cropland area reversed or slowed down after the tipping points. For the 80 counties with the Type-3 tipping point, 27 counties had a tipping point during 1971–1990, 15 counties had a tipping point during 1951–1970, and 13 counties had a tipping point during 1931 – 1950.

3.3. Cropland expansion and abandonment

From 1850 to 2018, the LMAV cropland area increased by ~ 7 times. Cropland expansion primarily occurred in central and northern LMAV (Fig. 9). Cropland expansion rate was $516 \text{ km}^2 \text{ yr}^{-1}$ in the 1850s but decreased to only $60 \text{ km}^2 \text{ yr}^{-1}$ in the 1860s. Between the 1880s and the 1940s, cropland expansion kept at a relatively high rate with an average of $459 \text{ km}^2 \text{ yr}^{-1}$. The most rapid expansion happened in the 1960s at a rate of $749 \text{ km}^2 \text{ yr}^{-1}$. However, cropland expansion slowed down in the 1970s and the 1980s. The slowest expansion was in the 1980s at $38 \text{ km}^2 \text{ yr}^{-1}$. In the 1990s, the 2000s, and the 2010s (refers to 2011–2018 in this study), cropland expansion kept at a relatively low rate with an average of $90 \text{ km}^2 \text{ yr}^{-1}$. Over the entire period, 45.6 % of the expanded cropland area were in Arkansas, 20.3 % in Mississippi, 17.6 % in Louisiana, and 14.1 % in Missouri (Figure S5).

Cropland abandonment in the LMAV remained at a relatively low level from the 1850s to the 1930s ($< 15 \text{ km}^2 \text{ yr}^{-1}$). The abandonment rate increased slowly from $21.5 \text{ km}^2 \text{ yr}^{-1}$ in the 1940s to $46.5 \text{ km}^2 \text{ yr}^{-1}$ in the 1960s and increased rapidly in the following five decades. The most rapid cropland abandonment occurred in the 2010s when cropland was converted to other land cover types at a rate of $514 \text{ km}^2 \text{ yr}^{-1}$. From the 1850s to the 1960s, Louisiana contributed 63.6 % of abandoned croplands in the LMAV (Figure S6), indicating cropland abandonment occurred earlier in Louisiana than the other states. Over the entire period, $1.5 \times 10^4 \text{ km}^2$ croplands in the LMAV were abandoned, of which 33.4 % were in Louisiana, 32.6 % in Arkansas, and 24.6 % in Mississippi.

3.4. Validation and comparison

We used cropland maps from the NLCD in 2001 and NASS CDL in 2017 to validate the reconstructed cropland dataset (Fig. 10). The comparison with NLCD shows that the overall accuracy of our reconstructed in 2001 reached 92.6 %, and the user's accuracy and the

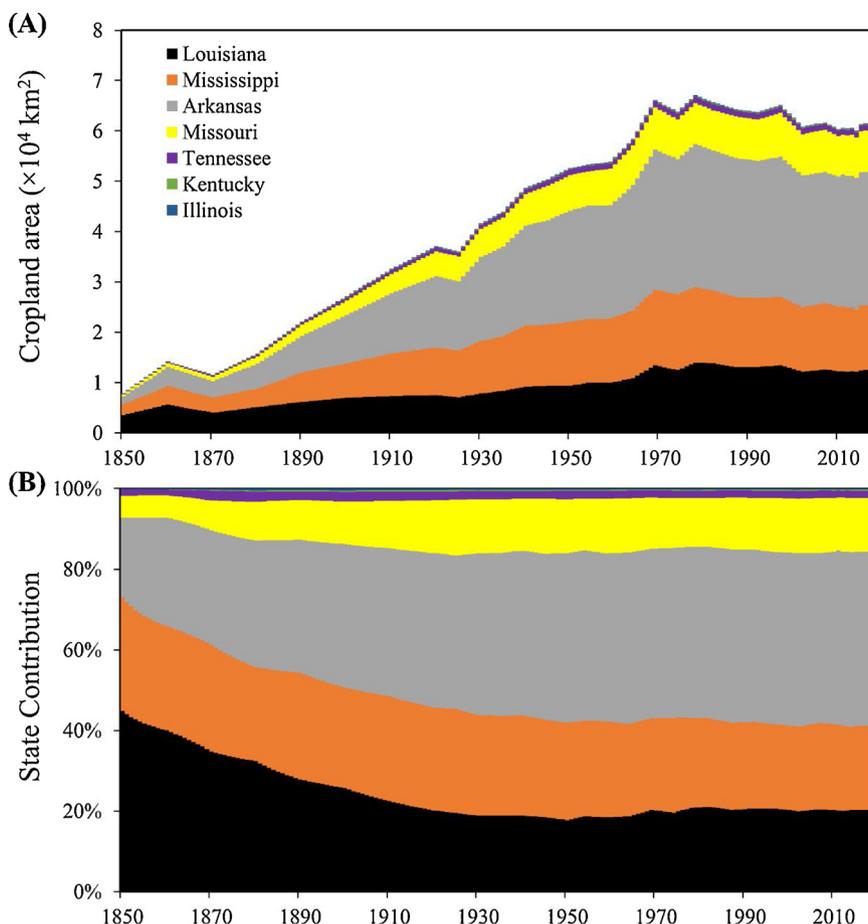


Fig. 5. Inter-annual variations of the LMAV cropland area (A) and the contributions of seven states to the total area (B).

producer’s accuracy for the cropland category were both 92.7 % (Table S1). The comparison with NASS CDL shows that the overall accuracy in 2017 was 94.2 %, the user’s accuracy was 95.5 %, and the producer’s accuracy was 92.9 % (Table S2). The high accuracies indicate that our reconstructed dataset captured the cropland distribution. However, due to the lack of reliable reference data sources with a high spatial resolution before 2001, we did not validate data in the earlier period.

Compared to the other four datasets (RF, HYDE3.2, ZC, and YL), one

distinct feature of our LMAV cropland dataset is the high spatial resolution (30 m). The LMAV dataset distinguished crop and non-crop pixels, while the other four datasets provided cropland fraction in a larger grid (see Fig. 11 for the cropland distribution in 1992). It is apparent that the YL dataset (Fig. 11D) and our LMAV dataset (Fig. 11E) provide more detailed spatial information than the other three datasets and, therefore, are more appropriate for regional and landscape-level applications. When aggregating to the same spatial

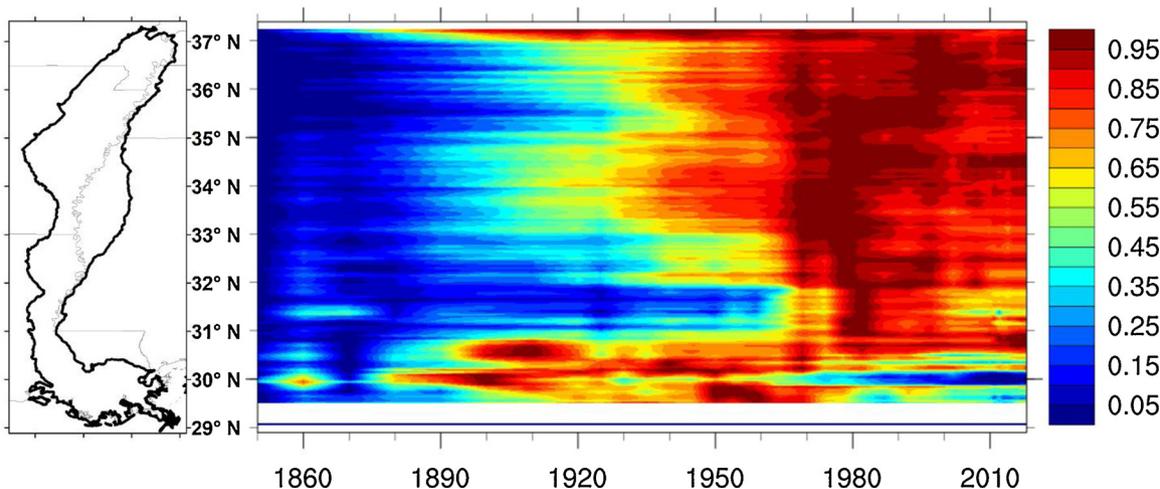


Fig. 6. Latitude vs. Time plot of the normalized cropland area (cropland fraction was normalized to the range between 0 and 1 by the minimum and maximum annual cropland fractions over the entire study period). The numbers of zero and one denote the least and the largest cropland area during 1850 – 2018 in each latitudinal band.

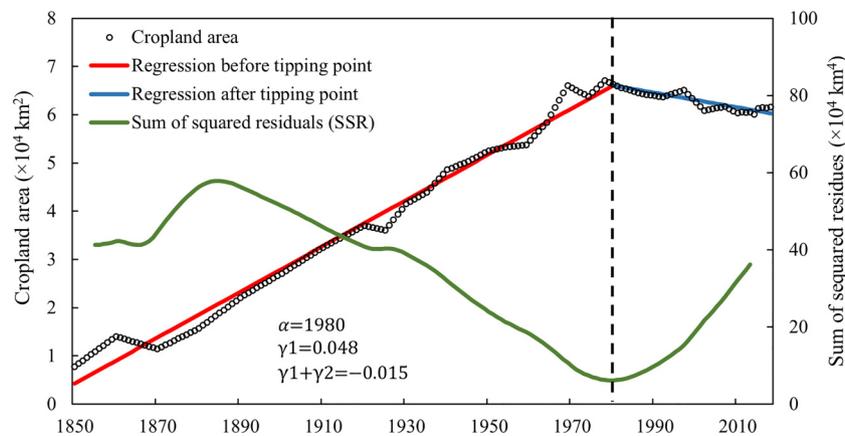


Fig. 7. Time series of the cropland area (black circle) in the Lower Mississippi Alluvial Valley from 1850 to 2018 and the detected tipping point (dash line). The green curve is the sum of squared residuals (SSR) for the fitted piece-wise linear regression with α moving from 1855 to 2013. The fitted lines before and after the tipping point are shown in red and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

resolution as the other datasets (Figure S7), we found that cropland distribution in our dataset is similar to that in HYDE, ZC, and YL. However, the RF dataset has a significantly higher cropland area in the Southern LMAV than our dataset and the other three.

In terms of temporal pattern, all the five datasets showed that the largest LMAV cropland area occurred in the period around 1980 (Fig. 12). Specifically, it was 1978 for RF, ZC, and the LMAV dataset in this study, 1980 for HYDE 3.2, and 1981 for YL. The RF dataset had a larger cropland area than the others in most of the study period. For instance, in the 1980s, the cropland area was $6.9 \times 10^4 \text{ km}^2$ for RF, $6.5 \times 10^4 \text{ km}^2$ for the LMAV dataset, $6.3 \times 10^4 \text{ km}^2$ for YL and HYDE, and $6.2 \times 10^4 \text{ km}^2$ for ZC. The most significant divergence in the changing trend among these datasets occurred during the period between 1930 and 1960 (shaded area and the inset in Fig. 12). The three datasets based on the state-level census (RF, HYDE 3.2, and YL) presented a declined cropland area during these 30 years, which was $0.8 \times 10^4 \text{ km}^2$ for RF, $0.5 \times 10^4 \text{ km}^2$ for HYDE 3.2, and $0.7 \times 10^4 \text{ km}^2$ for YL. In contrast, the two datasets based on the county-level census (ZC and the LMAV dataset) showed an increase in cropland area, which was $0.9 \times 10^4 \text{ km}^2$ for ZC and $1.3 \times 10^4 \text{ km}^2$ for the LMAV dataset. The divergences in the 30 years led to substantial discrepancies in cropland area magnitude for the period before the 1950s. In general, RF, HYDE 3.2, and YL had a larger area than the other two datasets. For instance, in 1910, regional cropland area was $6.0 \times 10^4 \text{ km}^2$ for RH, $5.1 \times 10^4 \text{ km}^2$ for HYDE 3.2, and $4.6 \times 10^4 \text{ km}^2$ for YL, but only $3.3 \times 10^4 \text{ km}^2$ for our LMAV dataset and $3.2 \times 10^4 \text{ km}^2$ for ZC.

4. Discussion

4.1. Driving forces of land use change in the LMAV

At the global level, human's reaction to economic opportunities, as regulated by institutional factors, was recognized as the major driving factors for LUC (Lambin et al., 2001). Historical changes of the LMAV cropland area could be potentially explained by the government policies and regional and national markets. We found that cropland expansion kept at a high level in the late 19th century and the early 20th century. The Swamp Land Acts of 1849–1850 granted federal swamplands to the states to encourage the land reclamation for agricultural use, which was considered as a critical driver for the fast agricultural expansion from the early 1800s to the 1930s (Stanturf et al., 2000). In the 1960s, soybean, wheat, and corn prices reached an all-time high and caused another round of fast cropland expansion in the LMAV (Oswalt, 2013; Stanturf et al., 2000; Sternitzke, 1976).

In recent decades, cropland abandonment overweighed cropland

expansion in the LMAV. The significant cropland abandonment started in the 1980s. Since then, afforestation on the marginal croplands has been widely implemented under the support of multiple conservation and wildlife habitat easement programs (Faulkner et al., 2011; King et al., 2006). For instance, the Conservation Reserve Program (CRP) and the Agricultural Conservation Easement Program (ACEP, formerly Wetlands Reserve Program (WRP)) offer rental payments or cost-share assistance to farmers who revert or convert agricultural land into forests or wetlands (Hamdar, 1999; Oswalt, 2013). These programs supported forest plantation on more than 270,000 ha of marginal farmland in the LMAV between 1990 and 2013 (De Steven et al., 2015). Through the administration of conservation programs, the government adjusted cropland enrollment acreage according to crop prices and exerted strong influences on the changes in cropland area at the regional and national levels (Hendricks and Er, 2018).

Conservation easement programs are voluntary, in which landowners retire land from agriculture in exchange for rental payments (<https://www.fsa.usda.gov/programs-and-services/conservation-programs/index>). Although landowners participating in easement programs could be largely grouped into the categories of financial orientation, ecological orientation, and residential orientation (Farmer et al., 2017), the primary driving force is the declining agricultural income relative to forest production (Napton et al., 2010). Therefore, agricultural lands with low production and high flooding frequency tend to be abandoned earlier than those in the upland area with a shorter inundation period (Stanturf et al., 2000).

4.2. Advantages of the LMAV cropland dataset

As described in Section 3.4, cropland areas in the five datasets showed significant discrepancies before the 1960s largely due to the difference in the incorporated agricultural census data. Between 1930 and 1960, datasets using state-level census showed declining cropland area, but datasets using county-level census showed the opposite. In terms of the changing trend, we have more confidence in ZC and our dataset as they included more detailed county-level census data. The county-level forest survey data from the Forestry Inventory and Analysis (FIA) program showed that forest area in the LMAV decreased by $\sim 1.5 \times 10^4 \text{ km}^2$ from 1930 to 1960 (identified from Fig. 5 in Oswalt (2013)), which supported the increased cropland area in our dataset by $1.3 \times 10^4 \text{ km}^2$ over the same period. Therefore, we argue that using state-level census data could introduce unrealistic changing trend and temporal variations into the reconstructed regional cropland area.

Comparing to the ZC dataset (5 min), our dataset (30 m resolution) give finer spatial information. The coarse-resolution dataset has been

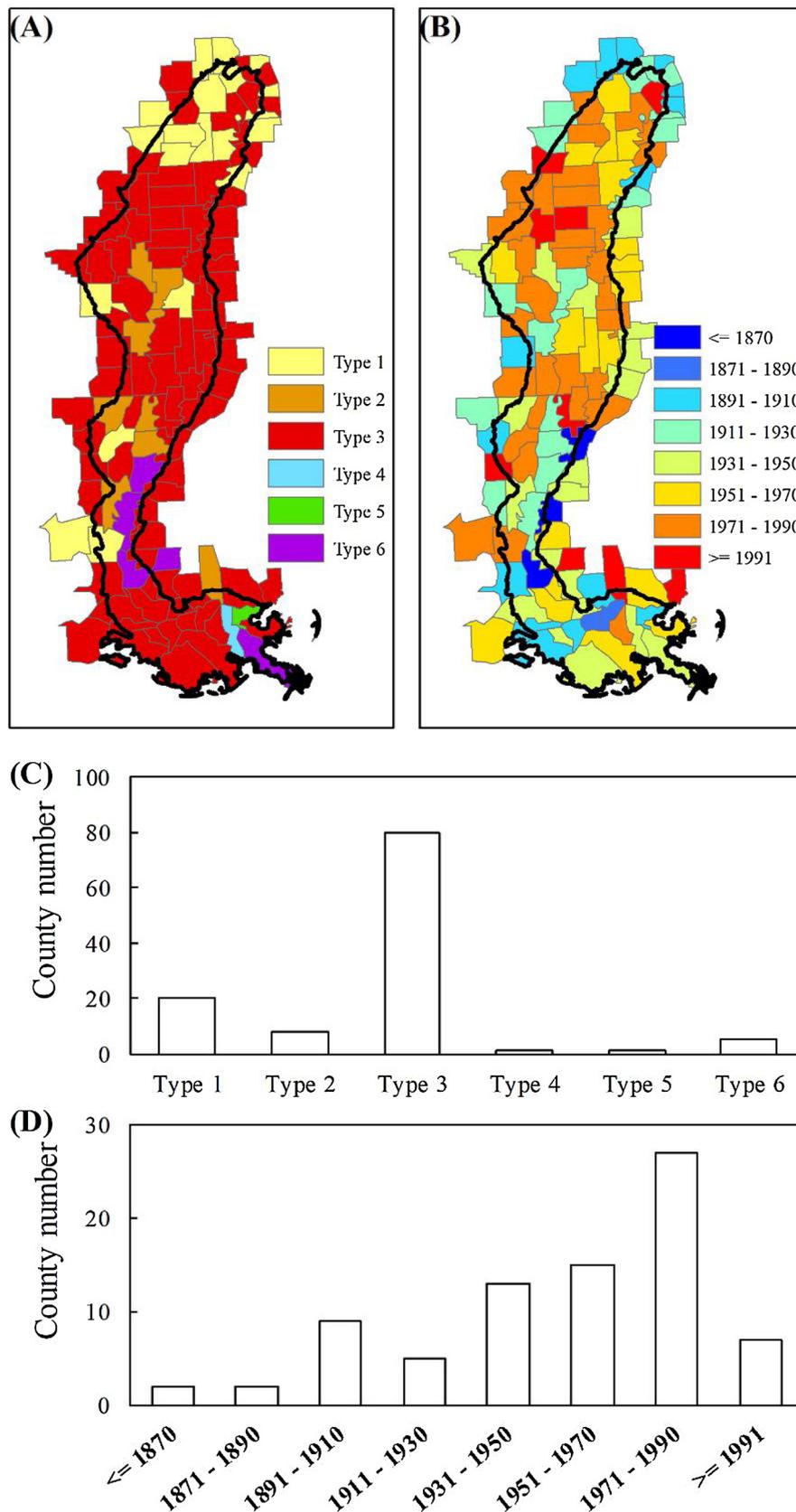


Fig. 8. Tipping point for the county-level cropland area. (A) Types of the tipping points, (B) occurrence period of the tipping points, (C) number of counties for each of the six types of tipping point, and (D) number of counties with the type-3 tipping point in eight periods. For the tipping points, changing trend of cropland area before the tipping point is γ_1 , and changing trend after it is $\gamma_1 + \gamma_2$. Type 1: $\gamma_1 > 0$, $\gamma_1 + \gamma_2 > 0$, and $\gamma_1 > \gamma_1 + \gamma_2$; Type 2: $\gamma_1 > 0$, $\gamma_1 + \gamma_2 > 0$, and $\gamma_1 < \gamma_1 + \gamma_2$; Type 3: $\gamma_1 > 0$ and $\gamma_1 + \gamma_2 < 0$; Type 4: $\gamma_1 < 0$, $\gamma_1 + \gamma_2 < 0$, and $\gamma_1 < \gamma_1 + \gamma_2$; Type 5: $\gamma_1 < 0$, $\gamma_1 + \gamma_2 < 0$, and $\gamma_1 > \gamma_1 + \gamma_2$; and Type 6: $\gamma_1 < 0$ and $\gamma_1 + \gamma_2 > 0$.

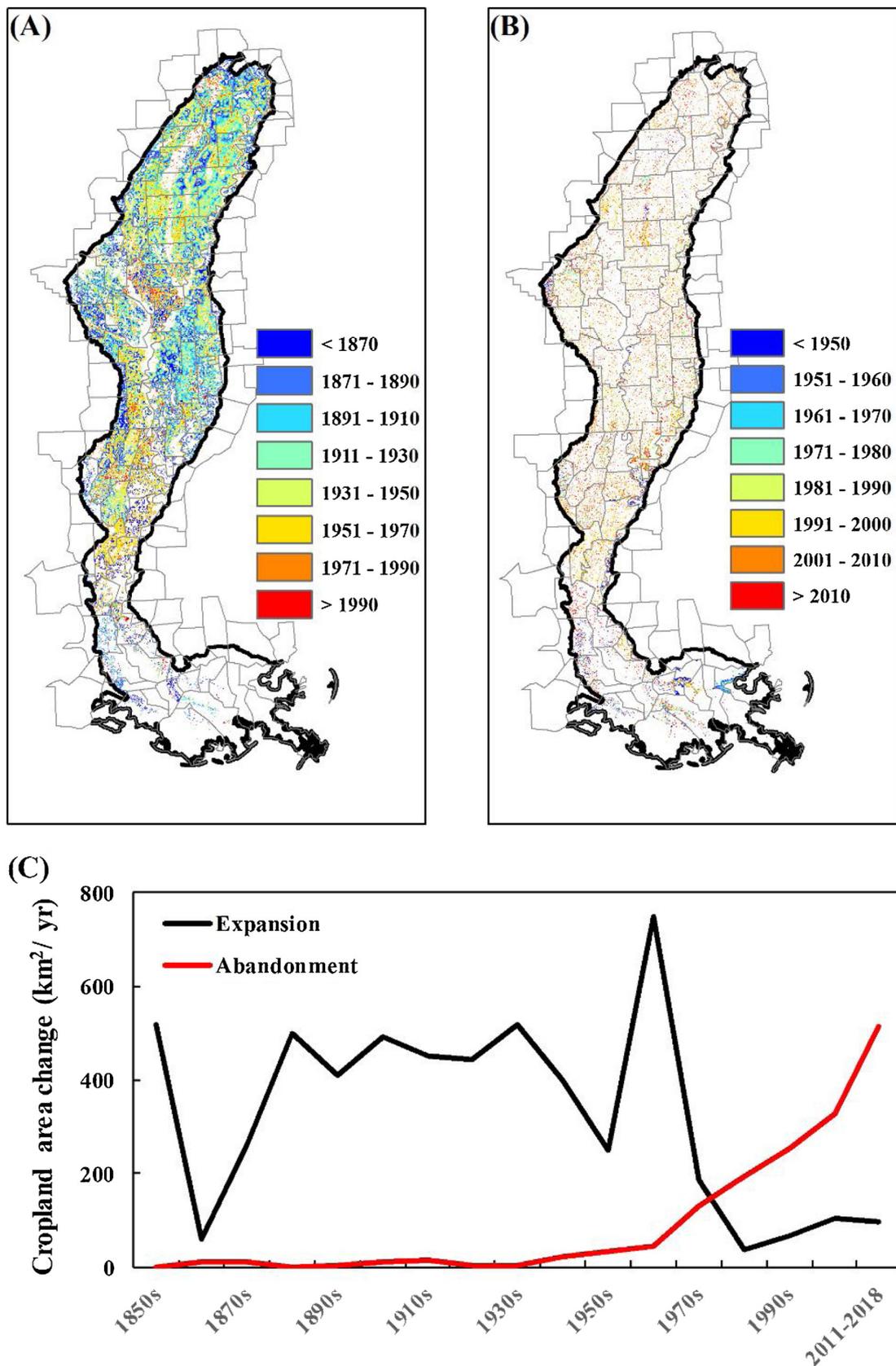


Fig. 9. Cropland expansion and abandonment in the Lower Mississippi Alluvial Valley. Spatial distribution and the periods of cropland expansion (A) and abandonment (B), and (C) decadal variations in the rates of cropland expansion and abandonment.

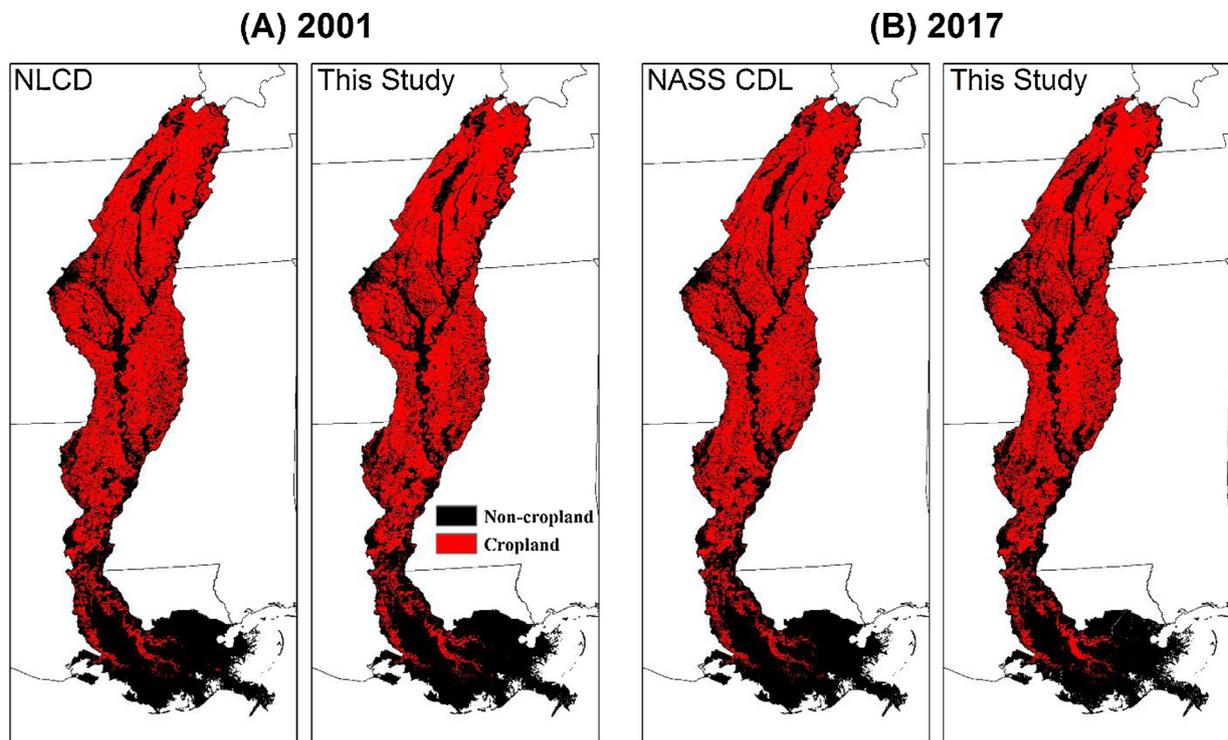


Fig. 10. Validation of the reconstructed cropland distribution in 2001 (A) and 2017 (B) against the National Land Cover Database dataset (NLCD) and the Cropland Data Layer (CDL).

found unable to represent land cover changes within the grid and tends to underestimate the LUC impacts on ecosystem dynamics (Yu et al., 2019). Therefore, our dataset would show better performance in evaluating the LUC impacts at the watershed, landscape, and county scales in the LMAV.

Croplands in our dataset consisted of five major components (see Section 2.2.1), while Yu and Lu (2018) defined the cropland as the lands that crops are planted on, including harvested cropland, cover crops, and crop failure. Cropland pasture and fallow were not considered as cropland in the YL dataset. The difference in cropland definition explained the slightly larger cropland area in our dataset than that in the YL data from 2010 to 2016 (Fig. 12), although both incorporated NASS CDL over this period. We also noticed a drop in cropland area from 2009 to 2010 in the YL dataset ($5.9 \times 10^4 \text{ km}^2$ in 2009 and $5.4 \times 10^4 \text{ km}^2$ in 2010) likely caused by the inconsistent cropland record in the NASS CDL (2010–2016) and the census data (Crop Production Historical Report, 1909–2016) used by Yu and Lu (2018). This artificial gap was removed in our dataset according to a reconciliation algorithm (section 2.3.1) through modifying the census data, implying the necessity of area reconciliation when merging two or more cropland datasets.

In the RF and ZC datasets, the total cropland area in one political unit (county or state) was allocated to grids using the spatial pattern observed by satellite in the contemporary period. HYDE 3.2 spatialized the total cropland area through a more sophisticated algorithm by arbitrarily setting the weights to satellite observations and other geographic parameters. In this study, we developed a machine learning algorithm and used the CDL cropland pattern in 2010 to train this algorithm. Through this approach, the most likely places for cropland area change were automatically determined. Our study introduced the combination of machine learning and county-level census to the long-term cropland reconstruction at a high spatial resolution, which could potentially improve the assessment accuracy of LUC impacts. This methodology can also be transplanted to other regions or basins for the long-term cropland reconstruction.

4.3. Potential applications of this dataset

In the LMAV, the current evaluation of the LUC impacts on ecosystem services was primarily conducted at the site level through in-situ observations and measurements (e.g. Brye et al., 2016; Shoch et al., 2009). We acknowledge that these site-level observations are instrumental in understanding ecosystem services offered by natural ecosystems and agricultural lands. However, there is still a knowledge gap regarding how the LUC affected the LMAV ecosystem services at the regional and landscape scales over a long period. Process-based land ecosystem modeling and hydrological modeling could potentially fill this knowledge gap (e.g. Francesconi et al., 2016; Logsdon and Chaubey, 2013; Schröter et al., 2005). Cropland dataset developed in this study could be processed in conjunction with the potential vegetation map to produce a long-term land cover dataset (Liu et al., 2013; Ramankutty and Foley, 1999a). Simulations of the process-based models, driven by the LUC dataset in this study, could provide spatially-explicit information regarding the impacts of cropland expansion and afforestation on regional ecosystem services, such as ecosystem productivity, carbon sequestration, land ecosystem GHG emissions, water quality, groundwater storage, and soil erosion (e.g. Tao et al., 2014; Tian et al., 2016; Zhang et al., 2017).

Additionally, our dataset could help to attribute the long-term river discharge and water quality records (Green et al., 2014). USGS provided the hydrological observations for Mississippi and other states (<https://waterdata.usgs.gov/ms/nwis/qw>). Data records in many sites could date back to the early 20th century. For instance, river discharge data for the site on the Big Sunflower River (USGS 07288500) in Mississippi started in 1936, and the water quality data started in 1964. Our dataset could be employed to quantify the contributions of land conversion and agricultural management to the time series of water quality and river flow observation. The information on the magnitude of cropland area and cropland distribution could be instrumental for social economists to investigate how socioeconomic, geographical, and environmental factors controlled the regional land use patterns. This dataset is expected to aid in developing better managerial and policy

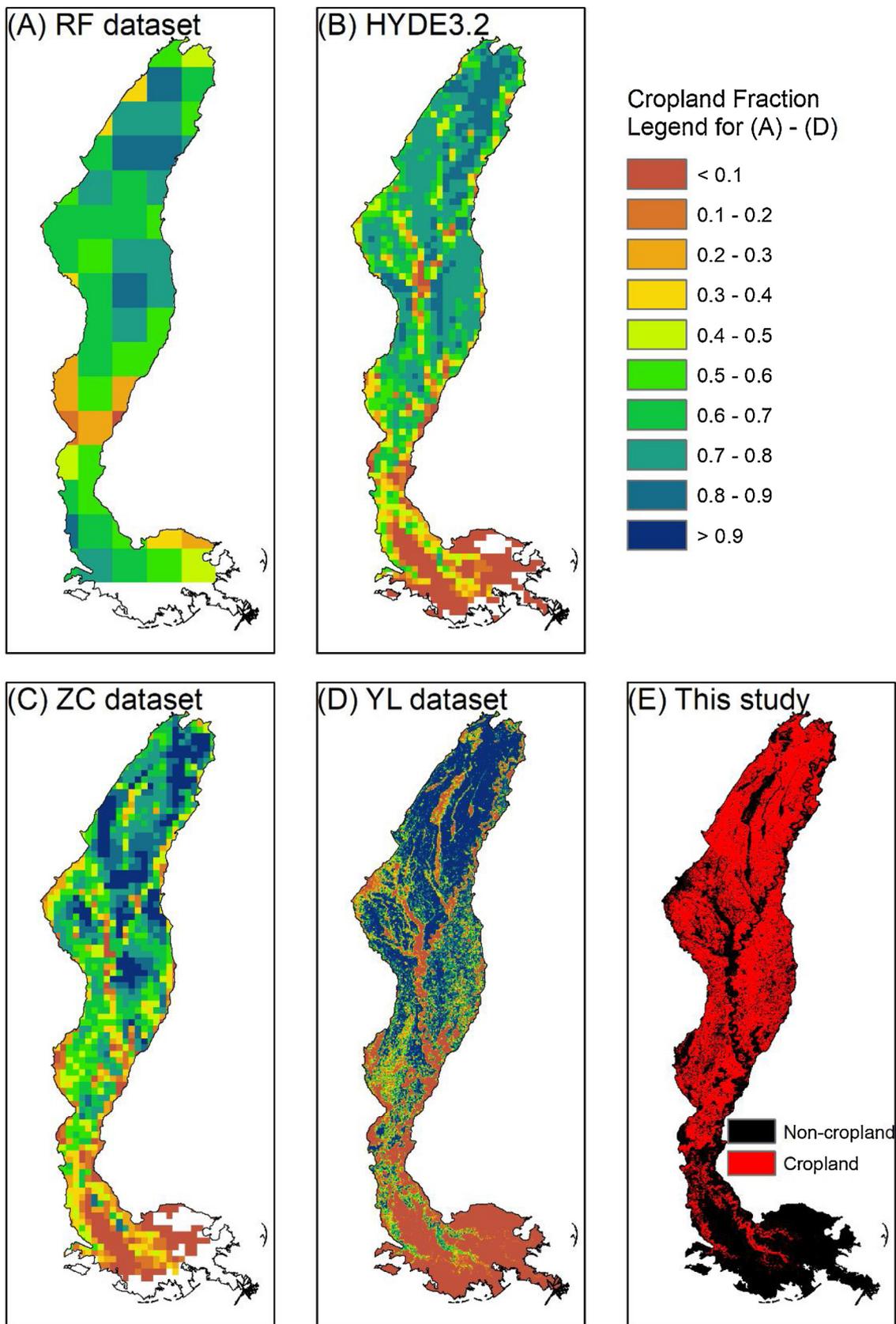


Fig. 11. Spatial distribution of cropland in 1992 in the Lower Mississippi Alluvial Valley in five cropland datasets. All the datasets are projected into an Albers Equal Area coordinate system.

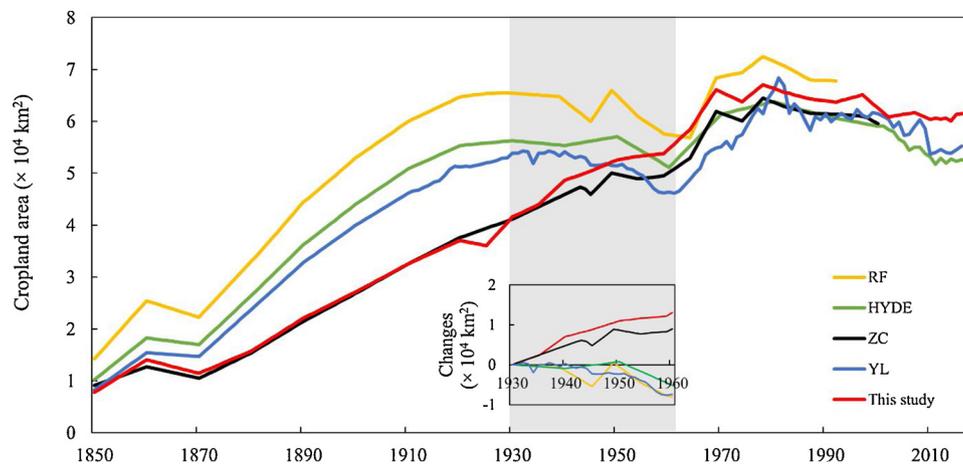


Fig. 12. Inter-annual variations of the total cropland area from 1850 to 2018 in the Lower Mississippi Alluvial Valley in the five datasets. The inset shows the changed cropland area between 1930 and 1960.

strategies for conserving natural resources and enhancing ecosystem services.

4.4. Implications for stakeholders

To acquire financial support from national and state-level conservation programs and restore natural resources, many farmers in the LMAV have a strong willingness to plant bottomland hardwoods on the marginal agricultural lands (Faulkner et al., 2011; King et al., 2006). The delineation of marginal agricultural land could provide useful information for landowners to determine the locations to retire croplands and enroll their lands in easement programs. Eq. 4 considers topographical and hydrological factors to estimate cropland probability. In conjunction with the spatially-explicit crop yield census data and satellite-based vegetation productivity, a machine learning approach could be developed to estimate crop productivity (Chlingaryan et al., 2018) and support landowners to make decisions on agricultural land retirement and hardwood forest plantation.

The declined cropland area in recent decades (Fig. 5) indicates the effectiveness of conservation easement programs in restoring natural vegetation. With the emerging environmental issues in the LMAV, such as groundwater overdraft, soil erosion, and water quality degradation (Ullah and Faulkner, 2006; Yang et al., 2019), national and state easement programs are expected to play a more important role in mitigating the negative consequences caused by crop cultivation. The cropland dataset developed in this study could be helpful to understand to what extent the governmental policies and easement programs have changed land use pattern in the LMAV.

4.5. Uncertainties and future needs

In this study, a logistic regression algorithm was used to estimate cropland probability and determine the placement of LUC occurrence. Although this method has been proved to be capable of reconstructing the historical land use pattern (Sohl et al., 2016, 2012; Sohl and Sayler, 2008), our approach and dataset are subject to some uncertainties. One source of uncertainty is the fitted parameters based on the satellite-observed cropland distribution in 2010. The parameters were assumed to be constant over the entire study period. Nonetheless, this assumption may not hold in the early period, particularly the 19th century, when the decisions on reclamation location might be different from the contemporary period. We acknowledge that data accuracy would decrease with time back to 1850. Uncertainty may also come from the annual cropland area changes. As the county-level census data could only provide information on the net changes in the annual cropland area, we assumed that, in one specific county, either cropland

expansion or abandonment equaled the net changes in the cropland area. This assumption ignored the simultaneous occurrence of cropland expansion and abandonment in one county between two census periods and could lead to the underestimated rates in cropland expansion and abandonment. According to the county chronologies in the Atlas of Historical County Boundary (<https://publications.newberry.org/ahcbp/index.html>), county boundaries in the LMAV remained roughly stable after the 1850s. Therefore, we expected that the application of contemporary county boundaries would not lead to a significant uncertainty in the data reconstruction.

Besides the logistic regression, other machine learning algorithms, such as Random Forest (RF) and Artificial Neural Networks (ANN), have also been implemented to model local and regional land use change (e.g. Li and Yeh, 2002; Samardžić-Petrović et al., 2017). These algorithms have the potential to be installed into the framework in this study to determine the appropriate locations of land use change. In the future work, we suggest reconstructing historical cropland patterns through multiple machine learning algorithms and compare data accuracy.

5. Conclusions

Human activities have substantially changed global land surfaces and ecosystem services. However, applications of the coarse-resolution cropland dataset led to significant uncertainties when evaluating the LUC impacts at the regional and landscape levels. To this end, we developed a novel framework to reconstruct a regional cropland dataset in the LMAV through integrating remote sensing observations, county-level census, and a machine learning algorithm. The major advantages of the method and dataset reside in (1) the relatively high spatial resolution (30 m) allowing its applications in investigating basin and watershed-level land use change impacts on the long-term hydrological and ecological processes; and (2) the integration of county-level census data enhancing data accuracy of the long-term trend in cropland area, particularly during the period before the 1960s. This dataset is expected to reduce the uncertainties in evaluating the LUC impacts and provide a sound scientific basis for the government and policy makers to develop land use and management plans to enhance ecosystem services. The framework developed in this study can also be used in other regions for long-term high-resolution cropland reconstruction.

Data availability

Availability of all input data is described in section 2 “Data and Methods”. The reconstructed land use dataset in the LMAV could be obtained upon reasonable request from the corresponding author.

CRedit authorship contribution statement

Jia Yang: Conceptualization, Methodology, Software, Writing - original draft. **Bo Tao:** Conceptualization, Methodology, Writing - review & editing. **Hao Shi:** Methodology, Writing - review & editing. **Ying Ouyang:** Conceptualization, Writing - review & editing. **Shufen Pan:** Writing - review & editing. **Wei Ren:** Data curation, Writing - review & editing. **Chaoqun Lu:** Data curation, Writing - review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgement

This work was supported by the U.S. Department of Agriculture McIntire-Stennis project under accession number 1019116, National Science Foundation grants (1903772 and 1940696). We appreciate the three anonymous reviewers for their helpful comments and suggestions.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jag.2020.102151>.

References

- Betts, R.A., Falloon, P.D., Goldewijk, K.K., Ramankutty, N., 2007. Biogeophysical effects of land use on climate: model simulations of radiative forcing and large-scale temperature change. *Agricultural and Forest Meteorology, The Contribution of Agriculture to the State of Climate* Vol. 142. pp. 216–233. <https://doi.org/10.1016/j.agrformet.2006.08.021>.
- 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>.
- Brye, K.R., Nalley, L.L., Tack, J.B., Dixon, B.L., Barkley, A.P., Rogers, C.W., Smartt, A.D., Norman, R.J., Jagadish, K.S.V., 2016. Factors affecting methane emissions from rice production in the Lower Mississippi river valley, USA. *Geoderma Reg.* 7, 223–229. <https://doi.org/10.1016/j.geodrs.2016.04.005>.
- Chlingaryan, A., Sukkarieh, S., Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agric.* 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>.
- Claussen, M., Brovkin, V., Ganopolski, A., 2001. Biogeophysical versus biogeochemical feedbacks of large-scale land cover change. *Geophys. Res. Lett.* 28, 1011–1014. <https://doi.org/10.1029/2000GL012471>.
- De Steven, D., Faulkner, S.P., Keeland, B.D., Baldwin, M.J., McCoy, J.W., Hughes, S.C., 2015. Understory vegetation as an indicator for floodplain forest restoration in the Mississippi River Alluvial Valley, USA. *Restor. Ecol.* 23, 402–412.
- DeFries, R.S., Field, C.B., Fung, I., Collatz, G.J., Bounoua, L., 1999. Combining satellite data and biogeochemical models to estimate global effects of human-induced land cover change on carbon emissions and primary productivity. *Glob. Biogeochem. Cycles* 13, 803–815. <https://doi.org/10.1029/1999GB900037>.
- Fang, Y., Jawitz, J.W., 2018. High-resolution reconstruction of the United States human population distribution, 1790 to 2010. *Sci. Data* 5, 180067. <https://doi.org/10.1038/sdata.2018.67>.
- Farmer, J.R., Ma, Z., Drescher, M., Knackmuhs, E.G., Dickinson, S.L., 2017. Private landowners, voluntary conservation programs, and implementation of conservation friendly land management practices. *Conserv. Lett.* 10, 58–66. <https://doi.org/10.1111/conl.12241>.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. *Rev. Geophys.* 45. <https://doi.org/10.1029/2005RG000183>.
- Faulkner, S., Barrow, W., Keeland, B., Walls, S., Telesco, D., 2011. Effects of conservation practices on wetland ecosystem services in the Mississippi Alluvial Valley. *Ecol. Appl.* 21, S31–S48. <https://doi.org/10.1890/10-0592.1>.
- Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. *Science* 309, 570–574. <https://doi.org/10.1126/science.1111772>.
- Francesconi, W., Srinivasan, R., Pérez-Miñana, E., Willcock, S.P., Quintero, M., 2016. Using the Soil and Water Assessment Tool (SWAT) to model ecosystem services: a systematic review. *J. Hydrol.* 535, 625–636. <https://doi.org/10.1016/j.jhydrol.2016.01.034>.
- Gibbard, S., Caldeira, K., Bala, G., Phillips, T.J., Wickett, M., 2005. Climate effects of global land cover change. *Geophys. Res. Lett.* 32. <https://doi.org/10.1029/2005GL024550>.
- Green, C.T., Bekins, B.A., Kalkhoff, S.J., Hirsch, R.M., Liao, L., Barnes, K.K., 2014. Decadal surface water quality trends under variable climate, land use, and hydrogeological setting in Iowa, USA. *Water Resour. Res.* 50, 2425–2443. <https://doi.org/10.1002/2013WR014829>.
- Haines, M.R., 2010. Historical, Demographic, Economic, and Social Data: The United States, 1790–2002. <https://doi.org/10.3886/ICPSR02896.v3>.
- Hamdar, B., 1999. An efficiency approach to managing Mississippi's marginal land based on the conservation reserve program (CRP). *Resour. Conserv. Recycl.* 26, 15–24. [https://doi.org/10.1016/S0921-3449\(98\)00067-6](https://doi.org/10.1016/S0921-3449(98)00067-6).
- Hendricks, N.P., Er, E., 2018. Changes in cropland area in the United States and the role of CRP. *Food Policy* 75, 15–23. <https://doi.org/10.1016/j.foodpol.2018.02.001>.
- 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 Sensing* 81, 345–354.
- Houghton, R.A., House, J.L., Pongratz, J., Werf, G.R., van der DeFries, R.S., Hansen, M.C., Quére, C.L., Ramankutty, N., 2012. Carbon emissions from land use and land-cover change. *Biogeosciences* 9, 5125–5142. <https://doi.org/10.5194/bg-9-5125-2012>.
- Johnson, D.M., 2013. A 2010 map estimate of annually tilled cropland within the conterminous United States. *Agric. Syst.* 114, 95–105. <https://doi.org/10.1016/j.agsy.2012.08.004>.
- Kaminski, R.M., Moorhead, D.J., Hodges, J.D., Nassar, J.R., 1989. *Mississippi Alluvial Valley. Habitat Management for Migrating and Wintering Water-fowl in North America*. Texas Tech Univ. Press, Lubbock, pp. 203–247.
- King, S.L., Twedt, D.J., Wilson, R.R., 2006. The role of the wetland reserve program in conservation efforts in the Mississippi River Alluvial Valley. *Wildl. Soc. Bull.* 34, 914–920.
- Klein Goldewijk, K., Beusen, A., van Drecht, G., de Vos, M., 2011. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Glob. Ecol. Biogeogr.* 20, 73–86. <https://doi.org/10.1111/j.1466-8238.2010.00587.x>.
- Klein Goldewijk, K., Beusen, A., Doelman, J., Stehfest, E., 2017. Anthropogenic Land Use Estimates for the Holocene – HYDE 3.2. *Earth System Science Data* 9. pp. 927–953. <https://doi.org/10.5194/essd-9-927-2017>.
- Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skånes, H., Steffen, W., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C., Xu, J., 2001. The causes of land-use and land-cover change: moving beyond the myths. *Glob. Environ. Change* 11, 261–269. [https://doi.org/10.1016/S0959-3780\(01\)00007-3](https://doi.org/10.1016/S0959-3780(01)00007-3).
- Le, Q.B., Park, S.J., Vlek, P.L.G., Cremers, A.B., 2008. Land-Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human–landscape system. I. Structure and theoretical specification. *Ecol. Inform.* 3, 135–153. <https://doi.org/10.1016/j.ecoinf.2008.04.003>.
- Li, X., Yeh, A.G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int. J. Geogr. Inf. Sci.* 16, 323–343. <https://doi.org/10.1080/13658810210137004>.
- Liu, M., Tian, H., 2010. China's land cover and land use change from 1700 to 2005: estimations from high-resolution satellite data and historical archives. *Glob. Biogeochem. Cycles* 24. <https://doi.org/10.1029/2009GB003687>.
- Liu, M., Tian, H., Yang, Q., Yang, J., Song, X., Lohrenz, S.E., Cai, W.-J., 2013. Long-term trends in evapotranspiration and runoff over the drainage basins of the Gulf of Mexico during 1901–2008. *Water Resour. Res.* 49, 1988–2012. <https://doi.org/10.1002/wrcr.20180>.
- Logsdon, R.A., Chaubey, I., 2013. A quantitative approach to evaluating ecosystem services. *Ecol. Modell.* 257, 57–65. <https://doi.org/10.1016/j.ecolmodel.2013.02.009>.
- Loveland, T.R., Belward, A.S., 1997. The IGBP-DIS global 1km land cover data set, DISCover: first results. *Int. J. Remote Sens.* 18, 3289–3295. <https://doi.org/10.1080/01431697217099>.
- Lu, C., Tian, H., 2017. Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. *Earth Syst. Sci. Data* 9, 181–192. <https://doi.org/10.5194/essd-9-181-2017>.
- Napton, D.E., Auch, R.F., Headley, R., Taylor, J.L., 2010. Land changes and their driving forces in the Southeastern United States. *Reg. Environ. Change* 10, 37–53. <https://doi.org/10.1007/s10113-009-0084-x>.
- Omernik, J.M., 1987. Ecoregions of the Conterminous United States. *Ann. Assoc. Am. Geogr.* 77, 118–125. <https://doi.org/10.1111/j.1467-8306.1987.tb00149.x>.
- Oswalt, S.N., 2013. *Forest Resources of the Lower Mississippi Alluvial Valley*. Gen. Tech. Rep. SRS-GTR-177. USDA-Forest Service, Southern Research Station, Asheville, NC, pp. 1–29. 29 p. 177.
- Ouyang, Y., Feng, G., Parajuli, P., Leininger, T., Wan, Y., Jenkins, J.N., 2018. Assessment of surface water quality in the big sunflower river watershed of Mississippi Delta using nonparametric analysis. *Water Air Soil Pollut.* 229, 373. <https://doi.org/10.1007/s11270-018-4022-8>.
- Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J., Deadman, P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Ann. Assoc. Am. Geogr.* 93, 314–337. <https://doi.org/10.1111/1467-8306.9302004>.
- Pongratz, J., Reick, C., Raddatz, T., Claussen, M., 2008. A reconstruction of global agricultural areas and land cover for the last millennium. *Glob. Biogeochem. Cycles* 22. <https://doi.org/10.1029/2007GB003153>.
- Ramankutty, N., Foley, J.A., 1999a. Estimating historical changes in global land cover: croplands from 1700 to 1992. *Glob. Biogeochem. Cycles* 13, 997–1027. <https://doi.org/10.1029/1999GB900046>.

- Ramankutty, N., Foley, J.A., 1999b. Estimating historical changes in land cover: North American croplands from 1850 to 1992. *Glob. Ecol. Biogeogr.* 8, 381–396. <https://doi.org/10.1046/j.1365-2699.1999.00141.x>.
- Samardžić-Petrović, M., Kovačević, M., Bajat, B., Dragičević, S., 2017. Machine learning techniques for modelling short term land-use change. *ISPRS Int. J. Geoinf.* 6, 387. <https://doi.org/10.3390/ijgi6120387>.
- Schoenholtz, S.H., James, J.P., Kaminski, R.M., Leopold, B.D., Ezell, A.W., 2001. Afforestation of bottomland hardwoods in the Lower Mississippi Alluvial Valley: status and trends. *Wetlands* 21, 602–613.
- Schröter, D., Cramer, W., Leemans, R., Prentice, I.C., Araújo, M.B., Arnell, N.W., Bondeau, A., Bugmann, H., Carter, T.R., Gracia, C.A., de la Vega-Leinert, A.C., Erhard, M., Ewert, F., Glendining, M., House, J.I., Kankaanpää, S., Klein, R.J.T., Lavorel, S., Lindner, M., Metzger, M.J., Meyer, J., Mitchell, T.D., Reginster, I., Rounsevell, M., Sabaté, S., Sitch, S., Smith, B., Smith, J., Smith, P., Sykes, M.T., Thonicke, K., Thuiller, W., Tuck, G., Zaehle, S., Zierl, B., 2005. Ecosystem service supply and vulnerability to global change in Europe. *Science* 310, 1333–1337. <https://doi.org/10.1126/science.1115233>.
- Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python. In: *Proceedings of the 9th Python in Science Conference*. Scipy, P. pp. 61.
- Shoch, D.T., Kaster, G., Hohl, A., Souter, R., 2009. Carbon storage of bottomland hardwood afforestation in the Lower Mississippi Valley, USA. *Wetlands* 29, 535. <https://doi.org/10.1672/08-110.1>.
- Sohl, T., Saylor, K., 2008. Using the FORE-SCE model to project land-cover change in the southeastern United States. *Ecol. Modell.* 219, 49–65. <https://doi.org/10.1016/j.ecolmodel.2008.08.003>.
- Sohl, T., Sleeter, B.M., Zhu, Z., Saylor, K.L., Bennett, S., Bouchard, M., Reker, R., Hawbaker, T., Wein, A., Liu, S., Kanengietter, R., Acevedo, W., 2012. A land-use and land-cover modeling strategy to support a national assessment of carbon stocks and fluxes. *Appl. Geogr.* 34, 111–124. <https://doi.org/10.1016/j.apgeog.2011.10.019>.
- Sohl, T., Reker, R., Bouchard, M., Saylor, K., Dornbierer, J., Wika, S., Quenzer, R., Friesz, A., 2016. Modeled historical land use and land cover for the conterminous United States. *J. Land Use Sci.* 11, 476–499. <https://doi.org/10.1080/1747423X.2016.1147619>.
- Stanturf, J.A., Schweitzer, C.J., Gardiner, E.S., 1998. Afforestation of marginal agricultural land in the lower Mississippi River Alluvial Valley, U.S.A. *Silva Fenn.* 32 (3), 281–297.
- Stanturf, J.A., Gardiner, E.S., Hamel, P.B., Devall, M.S., Leininger, T.D., Warren, M.E., 2000. Restoring bottomland hardwood ecosystems in the Lower Mississippi Alluvial Valley. *J. For.* 98, 10–16. <https://doi.org/10.1093/jof/98.8.10>.
- Sterling, S.M., Ducharne, A., Polcher, J., 2013. The impact of global land-cover change on the terrestrial water cycle. *Nat. Clim. Chang.* 3, 385–390. <https://doi.org/10.1038/nclimate1690>.
- Sternitzke, H.S., 1976. Impact of changing land use on Delta hardwood forests. *J. For.* 74, 25–27. <https://doi.org/10.1093/jof/74.1.25>.
- Tao, B., Tian, H., Ren, W., Yang, J., Yang, Q., He, R., Cai, W., Lohrenz, S., 2014. Increasing Mississippi river discharge throughout the 21st century influenced by changes in climate, land use, and atmospheric CO₂. *Geophys. Res. Lett.* 41, 4978–4986. <https://doi.org/10.1002/2014GL060361>.
- Tian, H., Banger, K., Tao, B., Dadhwal, V.K., 2014. History of land use in India during 1880–2010: large-scale land transformations reconstructed from satellite data and historical archives. *Glob. Planet. Change* 121, 78–88. <https://doi.org/10.1016/j.gloplacha.2014.07.005>.
- Tian, H., Lu, C., Ciais, P., Michalak, A.M., Canadell, J.G., Saikawa, E., Huntzinger, D.N., Gurney, K.R., Sitch, S., Zhang, B., Yang, J., Bousquet, P., Bruhwiler, L., Chen, G., Dlugokencky, E., Friedlingstein, P., Melillo, J., Pan, S., Poulter, B., Prinn, R., Saunio, M., Schwalm, C.R., Wofsy, S.C., 2016. The terrestrial biosphere as a net source of greenhouse gases to the atmosphere. *Nature* 531, 225–228. <https://doi.org/10.1038/nature16946>.
- Toms, J.D., Lesperance, M.L., 2003. Piecewise regression: a tool for identifying ecological thresholds. *Ecology* 84, 2034–2041. <https://doi.org/10.1890/02-0472>.
- Ullah, S., Faulkner, S.P., 2006. Denitrification potential of different land-use types in an agricultural watershed, lower Mississippi valley. *Ecol. Eng.* 28, 131–140. <https://doi.org/10.1016/j.ecoleng.2006.05.007>.
- Veldkamp, A., Verburg, P.H., 2004. Modelling land use change and environmental impact. *J. Environ. Manage.* 72, 1–3. <https://doi.org/10.1016/j.jenvman.2004.04.004>.
- Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human domination of Earth's ecosystems. *Science* 277, 494–499.
- Waisanen, P.J., Bliss, N.B., 2002. Changes in population and agricultural land in conterminous United States counties, 1790 to 1997. *Glob. Biogeochem. Cycles* 16, 84–1–84-19. <https://doi.org/10.1029/2001GB001843>.
- Wang, X., Piao, S., Ciais, P., Li, J., Friedlingstein, P., Koven, C., Chen, A., 2011. Spring temperature change and its implication in the change of vegetation growth in North America from 1982 to 2006. *PNAS* 108, 1240–1245. <https://doi.org/10.1073/pnas.1014425108>.
- Yang, J., Ren, W., Ouyang, Y., Feng, G., Tao, B., Granger, J.J., Poudel, K.P., 2019. Projection of 21st century irrigation water requirement across the Lower Mississippi Alluvial Valley. *Agric. Water Manag.* 217, 60–72. <https://doi.org/10.1016/j.agwat.2019.02.033>.
- Yu, Z., Lu, C., 2018. Historical cropland expansion and abandonment in the continental U.S. During 1850 to 2016. *Glob. Ecol. Biogeogr.* 27, 322–333. <https://doi.org/10.1111/geb.12697>.
- Yu, Z., Lu, C., Tian, H., Canadell, J.G., 2019. Largely underestimated carbon emission from land use and land cover change in the conterminous United States. *Glob. Chang. Biol.* 25, 3741–3752. <https://doi.org/10.1111/gcb.14768>.
- Zhang, S., Fan, W., Li, Y., Yi, Y., 2017. The influence of changes in land use and landscape patterns on soil erosion in a watershed. *Sci. Total Environ.* 574, 34–45. <https://doi.org/10.1016/j.scitotenv.2016.09.024>.
- Zumkehr, A., Campbell, J.E., 2013. Historical U.S. cropland areas and the potential for bioenergy production on abandoned croplands. *Environ. Sci. Technol.* 47, 3840–3847. <https://doi.org/10.1021/es3033132>.