Earth System
Science
Data

Open Access
Discussions

Four-century history of land transformation by humans in the United States: 1630-2020

Xiaoyong Li[1,2,3], Hanqin Tian[2], Shufen Pan[2], Chaoqun Lu[4]

[1]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-environmental Sciences, Chinese Academy
5 of Sciences, Beijing 100085, China.
[2]International Center for Climate and Global Change Research, College of Forestry and Wildlife Sciences, Auburn University,
Auburn, AL 36849, USA.
[3]University of Chinese Academy of Sciences, Beijing 100049, China.
[4]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50010, USA.

10 *Correspondence to*: Hanqin Tian (tianhan@auburn.edu)

Abstract. The land of the conterminous United States (CONUS) has been transformed dramatically by humans over the last four centuries through land clearing, agricultural land expansion and intensification, and urban sprawl. Spatial-temporal data on long-term historical changes in land use and land cover (LULC) across the CONUS is essential for understanding and predicting the dynamics of coupled natural-human systems. A few efforts have focused on reconstructing historical databases 15 to characterize changes in cropland and urban extent in the CONUS. However, the high-resolution and long-term trajectories of multiple land use types remain unclear. By integrating multi-source data, such as high-resolution remote sensing image-based LULC data, model-based LULC products, and historical census data, we reconstructed LULC history at an annual time scale and 1 km x 1 km spatial resolution for the CONUS in the past 390 years (1630–2020). The results show widespread expansion of cropland and urban land associated with rapid loss of natural vegetation. Newly reclaimed croplands are mainly 20 converted from forest, shrubland, and grassland, especially in the Great Plains and North Central. Forest planting and regeneration accelerated the forest recovery in the Northeast and Southeast since the 1920s. The geospatial and long-term historical land use data from this study can be applied to assess the LULC impacts on regional climate, hydrology, carbon cycle, and greenhouse gas emissions. The datasets are available at https://doi.org/10.5281/zenodo.6469247 (Li et al., 2022).

1 Introduction

25 Land use and land cover (LULC) change is an essential component of global change, and humans have altered over one-third of the Earth's land surface (Foley et al., 2005; Winkler et al., 2021). The human-induced LULC changes, such as cropland expansion, deforestation, wood harvest, and tree planting, have profound impacts on climate change, carbon cycle, and biodiversity (Houghton et al., 1999; Dangal et al., 2014; Domke et al., 2020; Lark et al., 2020). In particular, agriculture and forest-related land use activities have been recognized as a critical pathway to achieve climate mitigation targets (Grassi et al., 30 2017; Griscom et al., 2017). Thus, a better understanding of historical LULC and its spatial-temporal dynamics is critical to quantify the effects of LULC change on the ecosystem.

In the past four centuries, the conterminous United States (CONUS) has experienced dramatic land use and land cover (LULC) changes associated with land clearing, cropland, and urban land expansion (Steyaert and Knox, 2008; Drummond and Loveland, 2010; Oswalt et al., 2014; Sohl et al., 2016). Before the arrival of Europeans, agriculture and crop planting existed

35   in the eastern woodlands, the Great Plains, and the southwestern US. Since the establishment of the first colony in Virginia, cropland and pasture land began to expand by land clearing, which mainly occurred in the eastern United States, and agriculture was the primary livelihood for 90% of the population during the colonial era (Steyaert and Knox, 2008). Driving by the westward movement in the 19th century, land clearing, agriculture expansion, and deforestation expanded across the Appalachian Mountains into Ohio, the upper Mississippi River basins, and the Great Lakes region (Cole et al., 1998; Billington

40   et al., 2001; Steyaert and Knox, 2008; Yu and Lu, 2018). In the Mississippi River Valley and Alabama, hardwood forests were cleared for cotton and grain production (Hanberry et al., 2012). Cropland and pasture land in New England and the Atlantic coast were abandoned, and the forest grew again in the late 19th century (Foster, 1992; Hall et al., 2002; Jeon et al., 2014). Though the environmental protection movement originated in the 1880s, tree planting began to increase until the 1930s (Stanturf et al., 2014). In the following 90 years, the national total plantation forest area increased to 27 Mha (Oswalt et al.,

45   2014; Chen et al., 2017). However, it still lacks a long-term land use dataset to characterize historical LULC trajectories for the CONUS.

Several efforts have produced LULC data for the CONUS in the past several decades. For example, multiple contemporary and spatially explicit LULC products with a resolution from 30 m to 1 km are available, including Global Land Cover (GLC) 2000 (Bartholome and Belward, 2005), MODIS land cover (Friedl et al., 2010), GlobeLand30 (Chen et al., 2015), National

50   Land Cover Database (NLCD) (Yang et al., 2018; Homer et al., 2020), and Cropland Data Layer (CDL) (Boryan et al., 2011; Lark et al., 2017, 2021). However, these datasets were generated using remote sensing images and cannot be used to characterize the century-long land use dynamics. Global-scale and long-term coverage land use datasets (e.g., Land and Use Harmonization (LUH2), the History Database of Global Environment (HYDE)) are widely used in global climate simulations and carbon budget projects (Goldewijk et al., 2017a, 2017b; Hurtt et al., 2006, 2020). However, these datasets have a coarse

55   resolution (from 5 arcmins to 0.5 degrees) and substantial uncertainties, which cannot present regional-scale details (Li et al., 2016; Yu and Lu, 2018). Moreover, the data uncertainties will significantly impact the quantification of LULC effects on the ecosystem (Peng et al., 2017; Yu et al., 2019). Some studies focused on reconstructing historical single-type land use datasets (e.g., settlement and cropland) for the US (Zumkehr and Campbell, 2013; Yu and Lu, 2018; Lerk et al., 2020). Nevertheless, the dynamics of pasture, forest, shrubland, and grassland also profoundly impact the ecosystem (Chen et al., 2006; Tian et al.,

60   2012). Therefore, developing a long-term and high-resolution land use dataset with multiple land use classes for the CONUS is essential for understanding the LULC change history and LULC impact on ecosystem regional climate, hydrology, carbon cycle, and greenhouse gas emissions.

In this study, we aim to reconstruct and analyze the spatial and temporal pattern of LULC in the CONUS during 1630–2020 by integrating high-resolution satellite data, reliable inventory data, and model-based LULC data. This study consists of three

65   parts: a description of input data and methods, an analysis of spatiotemporal characteristics of LULC in the past four centuries,

and a comparison between our results and other studies. We also discussed the driving forces of LULC changes and the uncertainties of the newly developed dataset.

2 Materials and Method

This study reconstructed the land use history (1630–2020) at annual time step and 1 km x 1 km spatial resolution for the
70 CONUS (48 states) using remote sensing-based LULC data, model-based land use data, and historical census data. In addition, we aggregated the state-level data to eight subregions to analyze the regional divergence of land use changes. These subregions include Northeast, Northeast, North Central, Southeast, South Central, Great Plains, Intermountain, Pacific Northwest, and Pacific Southwest (Oswalt et al., 2014, 2019) (Figure 1).

The reconstruction process of historical land use and land cover data mainly included two parts: (1) reconstructing the historical
75 urban land, cropland, pasture, and forest area at the state level (Section 2.2), (2) generating 1 km x 1 km spatial resolution gridded land use and land cover data (Section 2.3). Figure 2 shows the workflow for generating historical land use and land cover data. The following sections provide a detailed description of the input data and how we process the data.
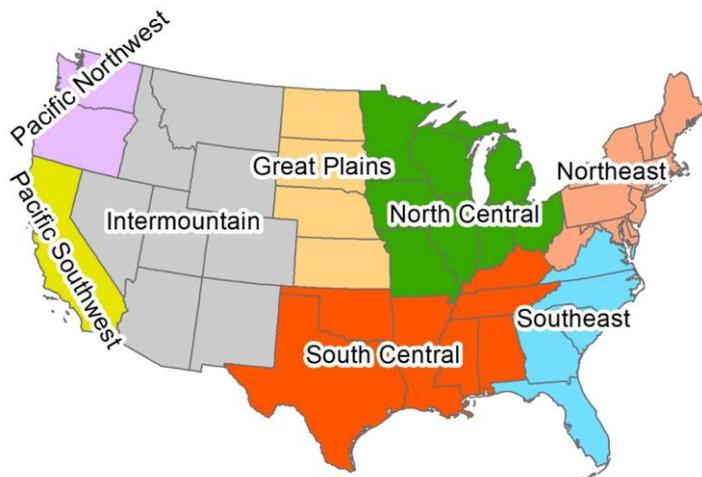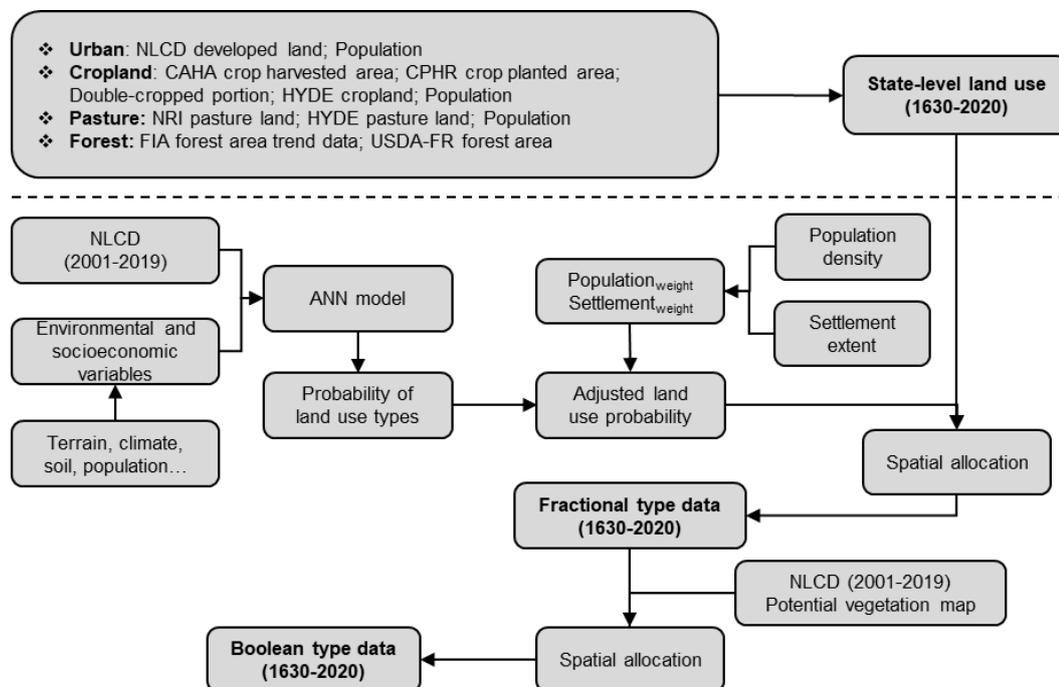

Figure 1: The division of the conterminous United States into eight subregions for data synthesis and analysis in this study.

Figure 2: Workflow for generating fractional and Boolean historical land use and land cover data. NLCD: National Land Cover Database; CAHA: Census of Agriculture Historical Archive; CPHR: the Crop Production Historical Report; HYDE: History Database of the Global Environment; NRI: National Resource Inventory; FIA: Forest Inventory and Analysis; USDA-FR: USDA Forest Resources of the United States, 2017; ANN: Artificial Neural Network.

2.1 Input datasets for land use and land cover reconstruction

The input datasets included satellite-based land use and land cover data (National Land Cover Database, NLCD), model-based land use datasets (i.e., HYDE), land use census data, and auxiliary data. We also collected some other land use datasets to validate our results, including Historical Settlement Data Compilation (HISDAC) (Leyk and Uhl, 2018; Uhl et al., 2021), Yu and Lu (2018) cropland density, Zumkehr and Campbell (2013) cropland fraction, Economic Research Service (ERS) major land uses data, CONUS historical land use and land cover (Sohl et al., 2016), and Haines et al. (2018) hay area. All the spatial data were resampled to 1 km x 1 km resolution for further processing. Table 1 and Table A1 show a detailed description of the data used in this study.

Table 1: Summary of data sources

Data variables	Time period	Resolution	Data sources
National Land Cover Database	2001, 2003, 2006, 2008, 2011, 2013, 2016, 2019	30 m	Multi-Resolution Land Characteristics Consortium https://www.mrlc.gov/
Historical Land Use and Land Cover	1938-1992	250 m	https://www.sciencebase.gov/catalog/item/59d3c73de4b05fe04cc3d1d1
Major land uses (ERS)	1945-2012	State-level	https://www.ers.usda.gov/data-products/major-land-uses/
Settlement data	1810-2015	250 m	Historical Settlement Data Compilation https://dataverse.harvard.edu/dataverse/hisdacus
Cropland harvested area	1879-2002	National and State level	USDA National Agricultural Statistics Service https://www.nass.usda.gov/index.php
Cropland planted area	1910-2018	National and State level	USDA Crop Production Historical Report http://agcensus.mannlib.cornell.edu/AgCensus/homepage.do
Cropland density (YLmap)	1850-2016	1 km	https://doi.pangaea.de/10.1594/PANGAEA.881801
Cropland fraction (ZCmap)	1850-2000	5 arcmin	https://portal.nersc.gov/project/m2319/
Cropland and pasture fraction (HYDE)	1600-2016	5 arcmin	https://dataportaal.pbl.nl/downloads/HYDE/HYDE3.2/
Pasture area (NRI)	1982-2017	State-level	National Resource Inventory Summary Report 2017 https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/nra/nri/results/
Haine et al. (2018) hay area	1840-2012	County level	United States Agriculture Data, 1840-2012 https://www.icpsr.umich.edu/web/ICPSR/studies/35206
Forest area (USDA)	1630-2017	State level	Forest Resources of the United States, 2017. https://www.fs.usda.gov/treesearch/pubs/57903
Forest area (FIA)	1630-2000	State level	Forest Inventory and Analysis https://www.fia.fs.fed.us/slides/Trend-data/Web%20Historic%20Spreadsheets/1630_2000_US_pop_and_forestarea.xls
Total population	1630-1999	State level	Coulson and Joyce (2003). United States State-level Population Estimates: Colonization to 1999.
	2000-2020	State level	https://www.census.gov/en.html
Population density	1630-2010	1 km	Fang et al. (2018) https://doi.org/10.6084/m9.figshare.c.3890191
The extent of settled area	1630-1890		https://maps.lib.utexas.edu/maps/histus.html

95 **Note: ERS: Economic Research Service, U.S. Department of Agriculture; YLmap: Yu and Lu (2018) cropland density; ZCmap: Zumkehr and Campbell (2013) historical fractional cropland area; HYDE: History Database of the Global Environment; USDA: United States Department of Agriculture; FIA: Forest Inventory Analysis.**

2.2 Historical land use and land cover area reconstruction

2.2.1 Urban land

100 This study regarded the developed land in the NLCD dataset as urban land. The developed land area during 2001–2016 was set as the baseline to reconstruct the historical time-series total urban land area. We first calculated the urban land per capita at the state level over the period using NLCD developed land area and population. Following previous studies (Liu and Tian, 2010; Goldewijk et al., 2017a), we estimated the total urban land area during 1630–2020 by multiplying the urban land per capita and total population at the state level.

105 2.2.2 Cropland

Cropland is defined as the areas used for to produce crops, such as corn, soybeans, and cotton (Homer et al., 2020). In this study, we counted cropland area as the area of land on which crops are planted within a year, excluding crop failure, summer fallow, idle crop, and cropland pasture (Bigelow and Borchers, 2017; Yu and Lu, 2018). The United States Department of Agriculture (USDA) provided many agricultural census data, the most important reference to reconstruct the historical

110 cropland area. However, the USDA did not provide the physical cropland area. Therefore, we used the planted area to estimate the physical cropland area at the state level by subtracting the double-cropped area.

The Crop Production Historical Report (CPHR) provided state-level and annual cropland planted crop area data from 1975 to 2020 (Table 1). Meanwhile, the USDA Census of Agriculture Historical Archive (CAHA) recorded state-level cropland harvested areas during 1879–2020 at 5 to 10 years intervals (Table 1). Because the annual state-level crop planted area data

115 were only available in the past 45 years, we used the state-level crop harvested area for further processing to keep the consistency of historical data. First, we linearly interpolated the crop harvested area during 1879–2020 between the time points. Then, we used the CPHR annual national-level total crop harvested area during 1909–2020 to adjust the interpolated crop harvested area, making the inter-annual variations more reasonable (Figure A2). The adjustment was based on the ratio of the state crop harvested area accounting for the national total.

120 After getting the state-level cropland harvested area, we used a conversion factor to estimate planted area. The conversion factor was calculated using a linear fit method and state-level crop harvested area and planted area during 1978–2017 ($y = 1.0665x$, $R^2 = 0.99$, $p < 0.05$; Figure A3). We calculated the physical cropland area by subtracting the double-cropped area. The regional double-cropped portion was derived from Borchers et al. (2014) and then dis-aggregated to the state level based on the cropland planted area as the weight value (Figure A4). The state-level double-cropped portion and the ratio of planted

125 area and harvested area were assumed to be consistent. The reconstructed cropland area in 1879 showed a significant difference from that in 1889, so we only used the state-level cropland area since 1889 for further analysis.

There was no available agricultural census data at the state level between 1630 and 1889, so we used the HYDE3.2 data to reconstruct for this period. We first summarized the national total cropland area and population to calculate the national cropland per capita. Then, we estimated the state-level cropland per capita during 1630–1889 based on the trend of HYDE

Earth System
Science
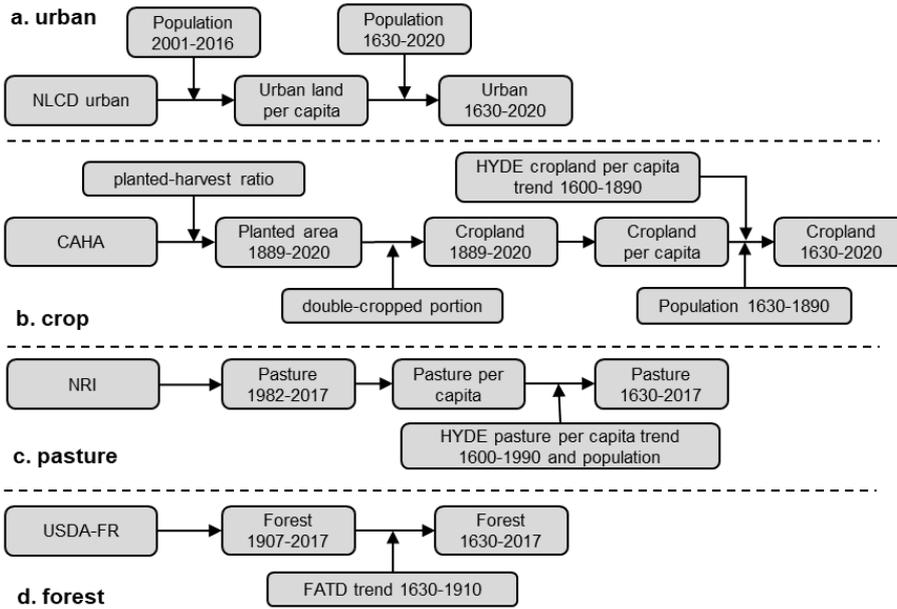Data

Open Access

Discussions

130   (Figure A5) and cropland per capita in 1889. In this step, we assumed that the changes in cropland per capita at the state level
        were consistent with that at the national level. Then, we calculated the national cropland area during 1630–1889 by multiplying
        state-level total population and cropland per capita. Finally, we combined the results in these two periods and got the cropland
        area at the state level during 1630–2020. Figure 3b shows all the cropland data used in different periods.

2.2.3 Pasture

135   Pasture is the land that has a vegetation cover of grasses, legumes, and forbs, regardless of whether it is being grazed by
        livestock, planted for livestock grazing, or the production of seed or hay crops (U.S. Department of Agriculture, 2020; Homer
        et al., 2020). In this study, we set the National Resource Inventory (NRI) pasture area from 1982 to 2017 as the baseline for
        its historical reconstruction. For the early period, though USDA provided the information on plowable pasture (1940 and
        before) or cropland used for pasture (1945 and after), farm woodland pasture, and other pasture (Wasianen and Bliss, 2002),
140   the pasture definitions changed several times, and the pasture and rangeland are not separated. Therefore, we used the HYDE
        pasture before 1982. First, the NRI data were interpolated linearly between 1982 and 2017. Then, we used the pasture per
        capita at the state level (NRI data-based) in 1982 and the national level (HYDE data-based) to estimate the pasture per capita
        during 1630–1982 (Figure A5). In this step, we assumed the trend of pasture per capita at the state level was consistent with
        that at the national level. Then, each state's pasture land area was calculated by multiplying the pasture per capita and
145   population. Additionally, we assumed that the area of pasture land during 2018–2020 was the same as that in 2017. Figure 3c
        shows all the pasture data used in different periods.

2.2.4 Forest

        Identical to the Forest Inventory and Analysis (FIA), the forest is defined as land at least 10 percent stocked by forest trees of
        any size, or formerly having such tree cover, with a minimum area classification of 1 acre. FIA's Forest area trend data (FATD)
150   provided state-level forest area from 1760 to 2000 at 10-year intervals and a snapshot in 1630. The data was rebuilt by
        integrating FIA field data and reports (1950–2000), field inventories (1910–1940), Bureau of the Census land clearing statistics
        (1850–1900), and clearing estimates proportional to population growth (1760–1840) and USDA forest report. The USDA
        Forest Resources (USDA-FR) of the United States 2017 (Oswalt et al., 2019) provided state-level forest areas from 1630 to
        2017, including twelve snapshots (i.e., 1907, 1920, 1938, 1953, 1963, 1977, 1987, 1997, 2007, 2012, 2017) and a shot in 1630.
155   We combined the two data sets and reconstructed a new historical forest inventory dataset. For the period 1907–2017 and
        1630, USDA-FR data was used. Before 1907, we calculated the ratio of forest area in 1760–1900 with 1630 (i.e.,
        $FATD_t/FATD_{1630}$, $1760 <= t <= 1900$) and then multiplied the forest area from USDA-FR to generate the forest area in 1760–
        1900. For 2018, 2019, and 2020, we first collected the latest forest area of each state. If one state did not publish the forest
        area of the latest year, we assumed that the area during these three years was the same as that in 2017. The latest forest area
160   data can be accessed at https://www.fia.fs.fed.us/tools-data/ (last accessed: April 18, 2022). The data used in different periods
        is shown in Figure 3d.

Earth System
Open Access
Science
Data
Discussions

Figure 3: Workflow for reconstructing historical land use area at state-level. NLCD: National Land Cover Database; CAHA: Census of Agriculture Historical Archive; NRI: National Resource Inventory; HYDE: History Database of the Global Environment; FIA: Forest Inventory and Analysis; USDA-FR: USDA Forest Resources of the United States, 2017.

2.2.5 Post-processing of historical urban, cropland, pasture and forest land area

Due to the difference in data sources in the reconstruction step, the total area of urban land, cropland, pasture, and forest may exceed the state's total land area (TLA). Therefore, we calibrated the reconstructed historical land use and land cover area using the following equations:

$$
\begin{cases}
A^t_{i,rc}(s) = A^t_{i,r}(s) & if \ TA^t_r(s) \leq TLA(s) \\
A^t_{i,rc}(s) = \frac{A^t_{i,r}(s)}{TA^t_r(s)} * TLA(s) & if \ TA^t_r(s) > TLA(s)
\end{cases}
\tag{1}
$$

$$
TA^t_{rc} = \sum_{i=1}^{n} A^t_{i,r}(s)
\tag{2}
$$

where t is the current year; $A^t_{i,rc}(s)$ and $A^t_{i,r}(s)$ are re-calibrated area and reconstructed area for the land use class i in the state s, respectively; TA^t_{rc} is the total area of urban, cropland, pasture and forest; n is total number of land use types; s is the state index in the range from 1 to 48.

2.3 Approach for generating gridded land use and land cover data

2.3.1 Calculating the land use probability

We reconstructed the historical land use and land cover data with 1 km x 1 km spatial resolution based on the state-level land use area and land use probability (Figure 2). Previous spatially explicit land use models, such as Conversion of Land Use and

its Effects (CLUE) model and Forecasting Scenarios of Land use Change (FORE-SCE) model, used the logistic regression

180 (LR) model to develop land use probability-of-occurrence (Verburg et al., 2009; Sohl et al., 2014, 2016; Yang et al., 2020; Li et al., 2016). However, it needs to train the LR model for the different units (e.g., county, grid) to calculate a good probability map due to the spatial heterogeneity of land conversion. In comparison, artificial neural networks (ANNs) can learn and fit complex relationships between input data and training targets and can be used to solve various non-linear geographical problems. Moreover, ANN has better performance than LR in land use simulation (Liu et al., 2016). Therefore, we estimated

185 the land use probability for urban, cropland, pasture, and forest using ANN and NLCD data. The variables for the ANN model training include elevation, slope, annual mean temperature, annual precipitation, annual maximum temperature (July), annual minimum temperature (January), crop productivity index, population density, distance to the city, distance to the road, distance to the railway, distance to the river, soil organic carbon, soil sand, soil clay. Table A1 shows the detailed information on the independent variables.

190 Over the past four centuries, the extent of the settled area in the CONUS expanded from the Northeast to the West coast, making the impacts of the natural environment and socioeconomic factors on LULC change gradually. We divided the study period into four sub-periods (p1: 1630–1790; p2: 1790–1850; p3: 1850–1920; p4: 1920–2020) to improve the ANN modeled probability. In the early period, the population density was an essential factor because the total area and spatial distribution of the human-dominated land use types (urban, crop, pasture) were always related to the population density. Thus, we use the

195 extent of the settled area and population density to restrict the land use change boundary (Figure A6). In the p4 period, the natural environment was not the decisive factor for human-dominated land use types due to the technology development. We further used the remote sensing-based land use map in the 2000s to constrain the land use probability (Goldewijk et al., 2017a). As a result, we calculate the final probability as follows:

$$\text{Prob}_{k,t} = (1 - w_t) * \text{Prob}_{k,\text{Pop},\text{SE},t} + w_t * \text{Frac}_{k,2000s} \tag{3}$$

200

$$\text{Prob}_{k,\text{Pop},\text{SE},t} = \text{Prob}_{k\_\text{ANN}} * \text{Pop}_{\text{weight},t} * \text{ES}_{\text{weight},t} \tag{4}$$

$$\text{Pop}_{\text{weight},t} = \frac{\text{Popd}_{i,t}}{\text{Popd}_{\text{mean},t}} \tag{5}$$

$$\text{SE}_{\text{weight},t} = \text{SE}_{t0} + w_t * \text{SE}_{t1} \tag{6}$$

$$w_t = \frac{t - t0}{t1 - t0} \tag{7}$$

where $\text{Prob}_{k,t}$ is the probability of land use type k in t year; $\text{Frac}_{k,2000s}$ is the fraction of land use type k; $\text{Prob}_{k\_\text{ANN}}$ is the

205 probability of land use type k determined by natural environmental conditions; $\text{Pop}_{\text{weight},t}$ is the population adjustment factor in t year, $\text{Popd}_{i,t}$ is population density at t year, $\text{Popd}_{\text{mean},t}$ is the mean population density at state level in t year; $\text{ES}_{\text{weight},t}$ is settlement weight in t year, which is calculated based on the settlement in t0 year and t1 year. For the p1 sub-period, we used the population weight in 1790 due to the lack of population density data. For the p1, p2 and p3 sub-period,

we assumed that the land use dynamics was mainly constrained by natural environmental conditions and population
210 density, and the weight of $Frac_{k,2000s}$ was set as 0.

2.3.2 Strategies to generate fractional and Boolean land use and land cover data

In order to generate the fractional grid data, we assumed that the fraction of each land use type at the grid level was determined
by the probability (Fuchs et al., 2013; Tian et al., 2014; West et al., 2014; He et al., 2015). A grid (land use type k) with a high
probability will have a high density. Based on this principle and the state-level land use area, we generated the fractional land
215 use data at 1 km x 1 km resolution and annual time scale.

We further generated the Boolean type land use and land cover data at 1 km x 1 km resolution using the fractional data for
four land use types (urban, crop, pasture, and forest). First, the total number of potential pixels or the land use demand was
determined based on the reconstruction results in Section 2.2. Then, the area difference of land use type k between the target
and current map was calculated. If the difference is negative, land use type k will lose. In that case, the pixels of type k with
220 the high fraction will keep the condition of the current LULC map, and the rest pixels with a low fraction will be converted to
other types. If the difference is positive, land use type k will expand, the pixels of type k in the current map will be assigned
the value of k, then the pixels (non-k type) ranging on top will be assigned as k to meet the rest demand. Once a pixel has been
assigned to more than one land type, we will compare their probability among different categories and assign the type with the
highest probability to the conflicted pixel. Only urban land, cropland, pasture, and forest can be allocated spatially. The pixels
225 not assigned value will be updated using the NLCD and LANDFIRE Biophysical Settings data. Finally, each state was iterated
annually from 1630 to 2020.

### 2.4 Data validation and uncertainties

To validated the newly developed land use dataset, we compared it with other land use datasets. Considering the time cover
period of land use datasets, we derived the average state-level statistics area for urban, cropland, pasture, and forest from 2000–
230 2020 for comparison. However, the land use datasets, except for NLCD, only include parts of four land use types. Therefore,
the comparison is as follows: urban (This study, NLCD, and HISDAC), cropland (This study, NLCD, HYDE, ERS, and
YLmap), pasture (This study, NLCD, and HYDE), and forest (This study, NLCD, and LUH2). In the discussion section, we
compared the national-level statistics area with NLCD, HYDE, LUH2, ZCmap, YLmap, Sohl et al. (2016), Haines et al. (2018),
ERS, and HISDAC data. We also analyzed the spatial consistency and differences between our data and other land use datasets.
235 The spatial comparison is as follows: urban (This study, HISDAC, and Sohl et al. (2016)), cropland (This study, HYDE,
YLmap, and ZCmap), pasture (This study, HYDE, and LUH2), and forest (This study and LUH2).

3 Results

3.1 Validation of the newly developed land use dataset

We compared the state-level urban, cropland, pasture, and forest areas using data derived from NLCD, HYDE, ERS, YLmap,
240 and HISDAC with our data (Figure 4). Generally, the new developed land use dataset match well with the data for comparison.
The urban land acreages from this study are close to NLCD data (Figure 4a; R^2=0.99, Slope = 0.93), whereas lower than
HISDAC urban land data (Figure 4a; R^2=0.86, Slope = 1.33). Our cropland acreages are consistent with NLCD (Figure 4b; R^2
= 0.99, Slope = 0.99) and YLmap (Figure 4b; R^2 = 0.99, Slope = 0.91). However, ERS and HYDE data tend to overestimate
the cropland (Figure 4b, $Slope_{ERS}$ = 1.26, $Slope_{HYDE}$ = 1.14). The coefficients of determination between our pasture acreages
245 and NLCD (Figure 4c; R^2 = 0.93, Slope = 1.02) and HYDE (Figure 4c; R^2 = 0.87, Slope = 0.99) are higher than 0.87. For the
forest, both NLCD and LUH2 data are lower than our data, especially in the Rocky Mountain states (Figure 4d; $Slope_{NLCD}$ =
0.72, $Slope_{LUH2}$ = 0.66).
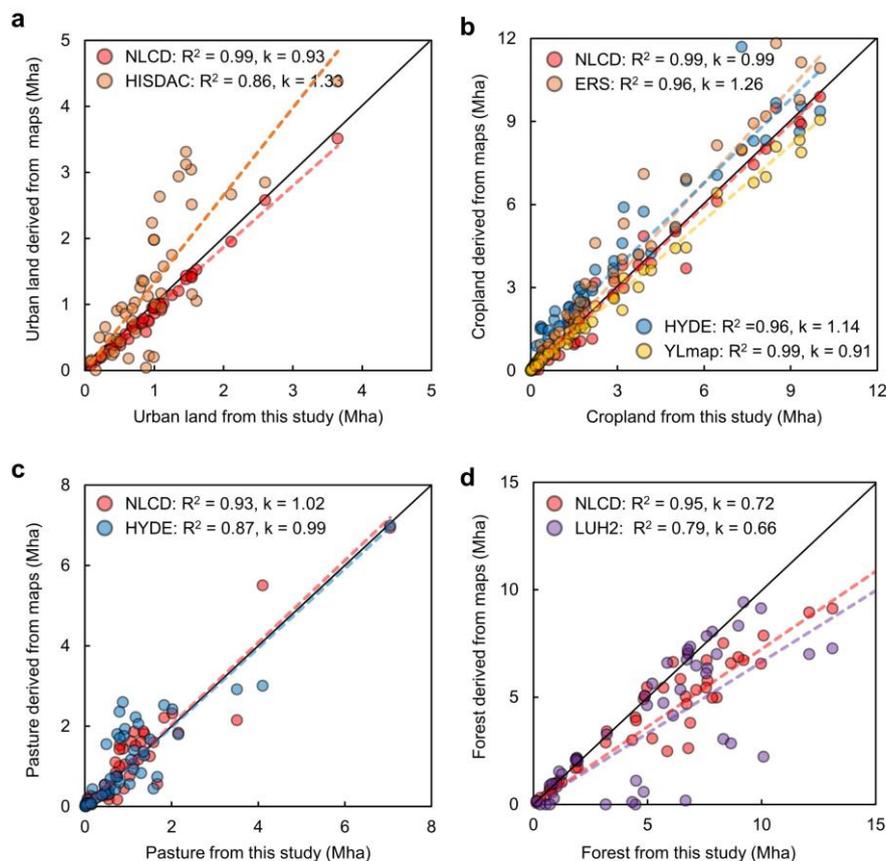
Figure 4. Comparison of the average urban (a), cropland (b), pasture (c), and forest (d) area in each state among National Land
250 **Cover Database (NLCD), Historical Settlement Data Compilation (HISDAC), Economic Research Service (ERS), Yu and Lu (2018)**
cropland (YLmap), History Database of the Global Environment (HYDE), Land Use Harmonization (LUH2) and this study. This

study: 2000–2020; NLCD: 2001, 2003, 2006, 2008, 2011, 2013, 2016, 2019; HISDAC: 2000, 2005, 2010, 2015; ERS: 2002, 2007, 2012; HYDE: 2000–2017; YLmap: 2000-2016; LUH2: 2000–2019.

3.2 Land use and land cover change during 1630–2020 in CONUS

255   The results showed that the land use and land cover change from 1630 to 2020 were the expansion of cropland and urban land and the shrinking of natural land cover (e.g., forest, grassland, and shrubland) (Figure 5, Figure A7-A10). In 1630, the primary landscape was forest in the eastern CONUS and Pacific Coastal region, grassland in the Great Plains, and shrubland in the Rocky Mountains (Figure 5). Urban land, cropland, and pasture were mainly distributed in the east of CONUS before 1850. Rapid cropland and pasture expansion occurred in the North Central region (e.g., Iowa, Illinois,

260   Minnesota), the Great Plains, and the Mississippi River Valley during 1850–1920 (Figure 5 and Figure A8). After 1920, the distribution of major land use classes became stable (Figure 6). The cropland in the Corn Belt regions, Central California, and Mississippi Alluvial Plain had the highest cropland density (Figure A8). The highest pasture density was found in the east of Texas, Oklahoma, Missouri, and Kentucky (Figure A9).
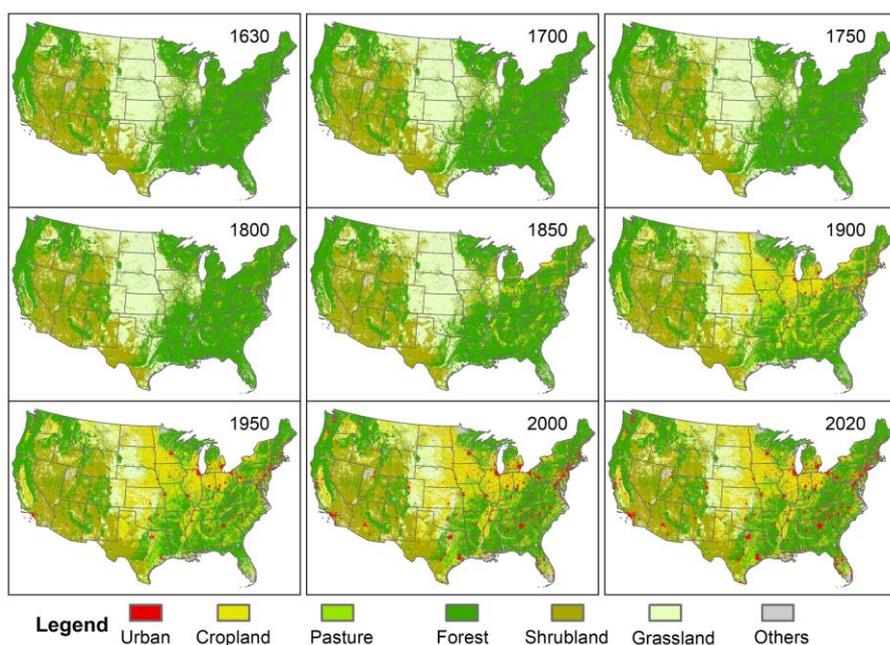
265   **Figure 5: Spatiotemporal patterns of land use and land cover in the conterminous United States during 1630-2020**
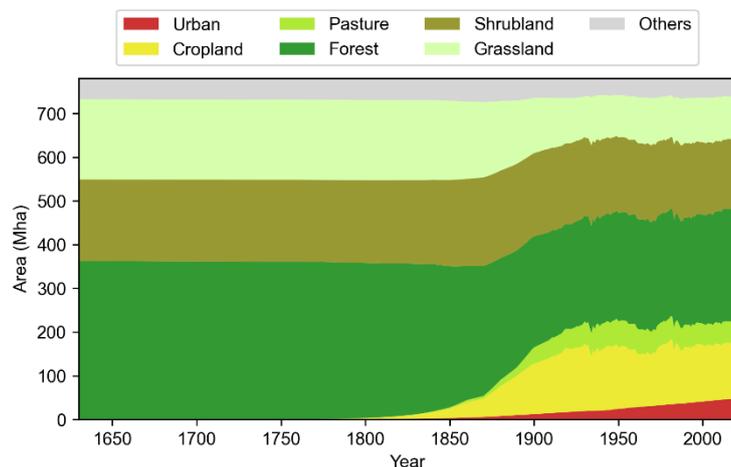
Figure 6: Changes in areas of land use in the conterminous United States from 1630 to 2020

The United States experienced the colonial era, the war of independence, and territorial expansion during 1630–1850. However, the total increase of urban land was only 3.23 Mha (Figure 6) and was mainly distributed in the Northeast (Figure 5

270 and Figure A10). In the middle of the 1800s, the cheap land and the industrial revolution growth prospect attracted many European and Mexican immigrants, which accelerated urban development. In 1880, the national total urban land area reached 7.64 Mha (Figure 6). Though the government limited the total amount of immigrants, the urban land area in the CONUS still increased by 21.85 Mha during 1880–1965. After that, the new immigration policy promoted the increase in population and urban sprawl. As a result, the national total urban land area increased to 47.81 Mha in 2020.

275 Cropland expanded slowly by 21.62 Mha from 1630 to 1850, and it increased substantially to 145.37 Mha in the following 70 years (Figure 6). Agriculture turned to be intensified after 1920, and the national total cropland area did not change significantly, with a peak area of 155.37 Mha in 1930 (Figure 6). Due to the competition of newly reclaimed cropland with high production in the Midwest, cropland abandonment occurred in the Northeast, South, and Southeast (Bigelow and Borchers, 2017; Yu and Lu, 2018). During 1950–1975, the rise of the manufacturing and service industry resulted in

280 agricultural labor and cropland reduction. As the demand for biofuel and bulk grain grew in the 2000s, cropland began to extend again, and the national area in 2020 was 127.32 Mha (Figure 6). Pasture showed an increasing trend with a slow rate during 1630–1850. It expanded more than 20 times from 1850 to 1950 and reached the maximum historical area (59.67 Mha) in 1950. The national total pasture area kept stable and decreased slowly in the following 70 years (Figure 6).

Forest was the dominant land use type in the CONUS before the colonial era, which accounted for about 47% of the total land

285 area. The trends in forest area were contrary to that of agricultural land in the past four centuries. During 1630–1850, the national total forest loss was 33.91 Mha (Figure 6). Over the second period (1850-1920), forest area decreased by 83.95 Mha because of agricultural land occupation, lumber cut, and fuelwood consumption (Steyaert and Knox, 2008). In the third period (1920–2020), forest area has been relatively stable through forest management and planting (Figure 6).

3.3 Land use and land cover transitions during 1630-2020

290   The changes in the land use area only reflected its quantitative changes. However, the land use transition map takes a further step to illustrate the spatial distribution of conversion between two land use types (Figure 7). Over the past 390 years, cropland expansion by occupying forest, shrubland, and grassland was the primary land use change characteristic (Figure 7). The natural land loss was mainly distributed in the North Central region (e.g., Ohio, Indiana) and Southern states such as Tennessee, Texas, Alabama, and Georgia (Figure 7d). New reclaimed cropland encroached 36.37 Mha (10.03%) of forest and 66.20 Mha

295   (18.12%) of grassland and shrubland. Meanwhile, 28.98 Mha of forest and 16.74 Mha of shrubland and grassland were converted to pasture. Moreover, urban land occupied more than 27 Mha of forest and 17 Mha of grassland and shrubland (Figure 7d). During the early period (1630–1850), forest converted to cropland was the dominant land use transition type, especially Eastern U.S. (Figure 7a). The U.S. experienced the most significant land use changes during 1850–1920, characterized by grassland converted to cropland in the Midwest and North Central and forest converted to pasture in Southern

300   (Figure 7b). Furthermore, abandoned cropland converted to the forest (17.36 Mha) distributed in the Northeast and Southern states was an essential feature of land use changes from 1920–2020 (Figure 7c).
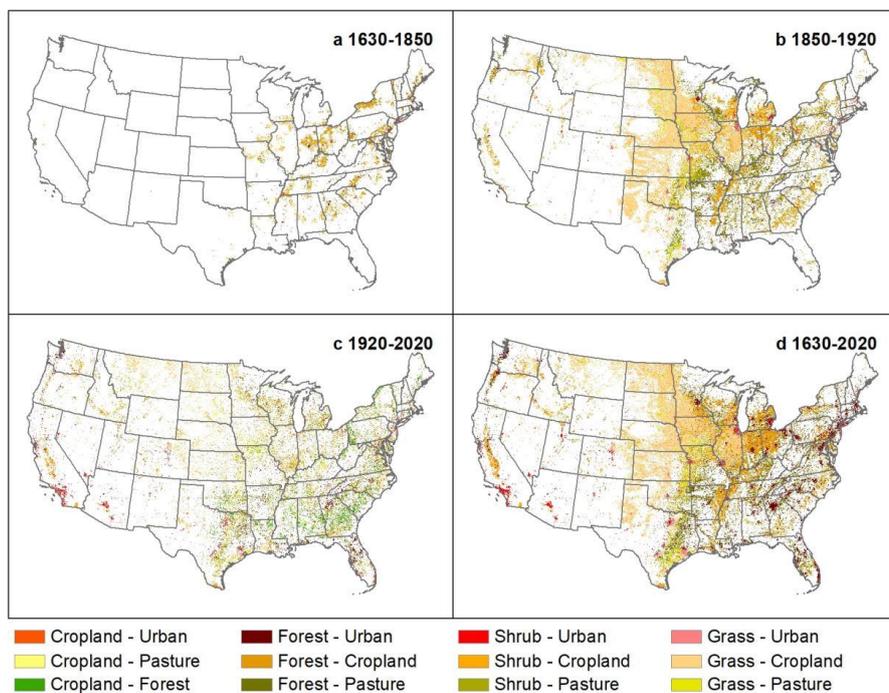

Figure 7: Land transition (1 km x 1km spatial resolution) between 1630 and 1850 (a), 1850 and 1920 (b), 1920–2020 (c), 1630 and 2020 (d) in the conterminous United States.

305 **3.4 Land use and land cover changes during 1630–2020 at regional level**

Given the differences in natural environmental conditions and social-economic development, land use and land cover changes showed significant spatial heterogeneity in the CONUS during 1630–2020. Since 1630, the South Central region experienced the most intensive urban land expansion (11.21 Mha), followed by the North Central (10.25 Mha), Southeast (7.80 Mha), and Northeast (6.32 Mha), respectively (Figure 8a). Rapid cropland expansion first occurred in the North Central, Northeast, South

310 Central, and Southeast in the 1830s. Cropland in the Intermountain and the Great Plains began to develop after 1860. The trends of cropland in eight regions except South Central and Southeast were consistent with the national total. Over the past four centuries, the North Central region had the largest cropland expansion area (47.89 Mha), followed by the Great Plains (33.03 Mha) and the South Central (20.38 Mha) (Figure 8b). The trends of cropland in eight regions were consistent with the national total. Cropland in the South Central and Southeast had decreased by 3.55 Mha and 14.50 Mha since the 1930s due to

315 the increasing urbanization pressures and low cropland profitability.

Similar to cropland, the Northeast region was the first to develop pasture. The pasture experienced a rapid expansion during 1790–1950 and finally reached the maximum historical (4.7 Mha) in the 1950s, and gradually decreased (Figure 8c). For a long period (1865–1980), the South Central region had the largest pasture area. The maximum historical area was 22.42 Mha in 1950 and accounted for 38% of the national total. However, the pasture area in the North Central region had decreased since

320 1960, and only 11.17 Mha of pasture was left by 2020 (Figure 8c).

Agricultural land encroachment, land clearing, and wood harvest resulted in forest loss in eight regions (Oswalt et al., 2014, 2019). The North Central region lost the most forest area (24.85 Mha), followed by the South Central region (36.12 Mha). During 1850–1920, the forest area decreased rapidly in the North Central (24.97 Mha), South Central (30.29 Mha), Southeast (14.03 Mha), and Northeast regions (6.5 Mha). Most of the lost forest converted to cropland and pasture (Figure 8d). Since the

325 1920s, the regional forest area has been relatively stable with small fluctuations. Notably, the forest land recovered gradually, especially in the Northeast. Compared with the 1920s, the total forest area in the Northeast increased by 6.86 Mha (Figure 8d).
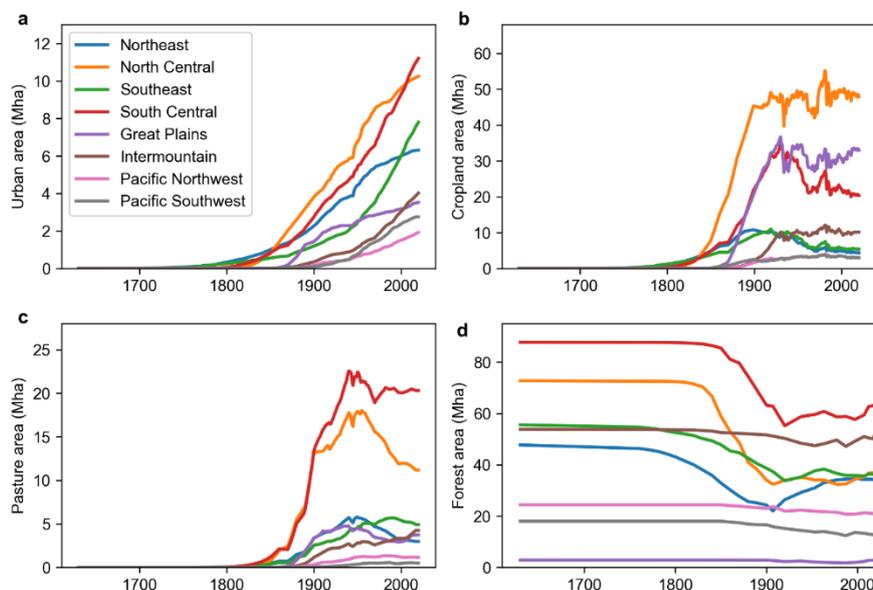
Figure 8: Changes in areas of urban land (a), cropland (b), pasture (c), forest (d) in different geographic regions during 1630–2020.

4 Discussion

4.1 Comparison with the previous dataset

This study reconstructed a gridded time series of LULC data for the CONUS from 1630 to 2020. Compared with the ERS and HYDE data, the reconstructed urban land was higher (Figure 9a), attributed to the differences in urban land definition. ERS urban areas include densely populated areas with at least 50,000 people and densely populated areas with 2,500 to 50,000 people, which have changed several times over the past 70 years (Bigelow and Borchers, 2017). In addition, the HISDAC settlement data was developed using the detailed address points data (Lerk et al., 2020), and remote sensing images cannot identify small-scale built-up land. As a result, the HISDAC settlement was higher than the reconstructed urban land in the recent four decades (Figure 9a). Moreover, the HISDAC settlement dataset may underestimate the total urban area due to the lack of address points or approximate location data in the early period (Lerk et al., 2020). We assumed that the developed land per capita was unchanged, and our results may overestimate the total urban land area in the early period. The spatial pattern of Boolean type urban land was consistent with the Sohl et al. (2016) data and was mainly distributed in the area near the city, road, and railway (Figure 10). The spatial allocation rule determined that the grid with a high probability of occurrence would be allocated first, which may underestimate the developed land in the rural area (Verburg et al., 2009; Yang et al., 2020). Though the urban data had some uncertainties, we provided a long-term description of urban land with higher resolution and consistency for the CONUS.

Earth System
Science
Data

Open Access

Discussions
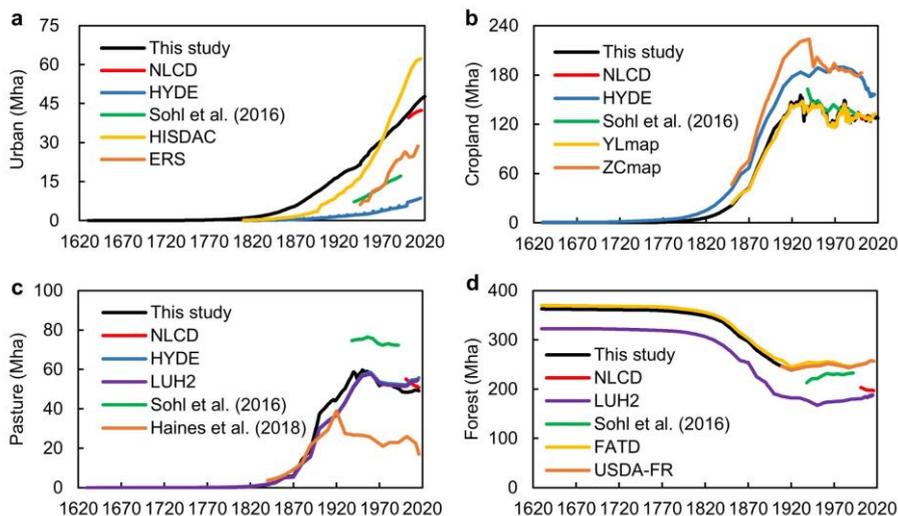
Figure 9: Comparison with other datasets for the conterminous United States: urban land (a); cropland (b); pasture (c); forest (d). NLCD: National Land Cover Database; HYDE: History Database of the Global Environment; HISDAC: Historical Settlement Data Compilation; ERS: Economic Research Service; YLmap: Yu and Lu (2018) cropland density; ZCmap: Zumkehr and Campbell (2013) cropland fraction; LUH2: Land Use Harmonization; FATD: Forest Area Trend Data; USDA-FR: USDA Forest Resources of the United States of 2017.
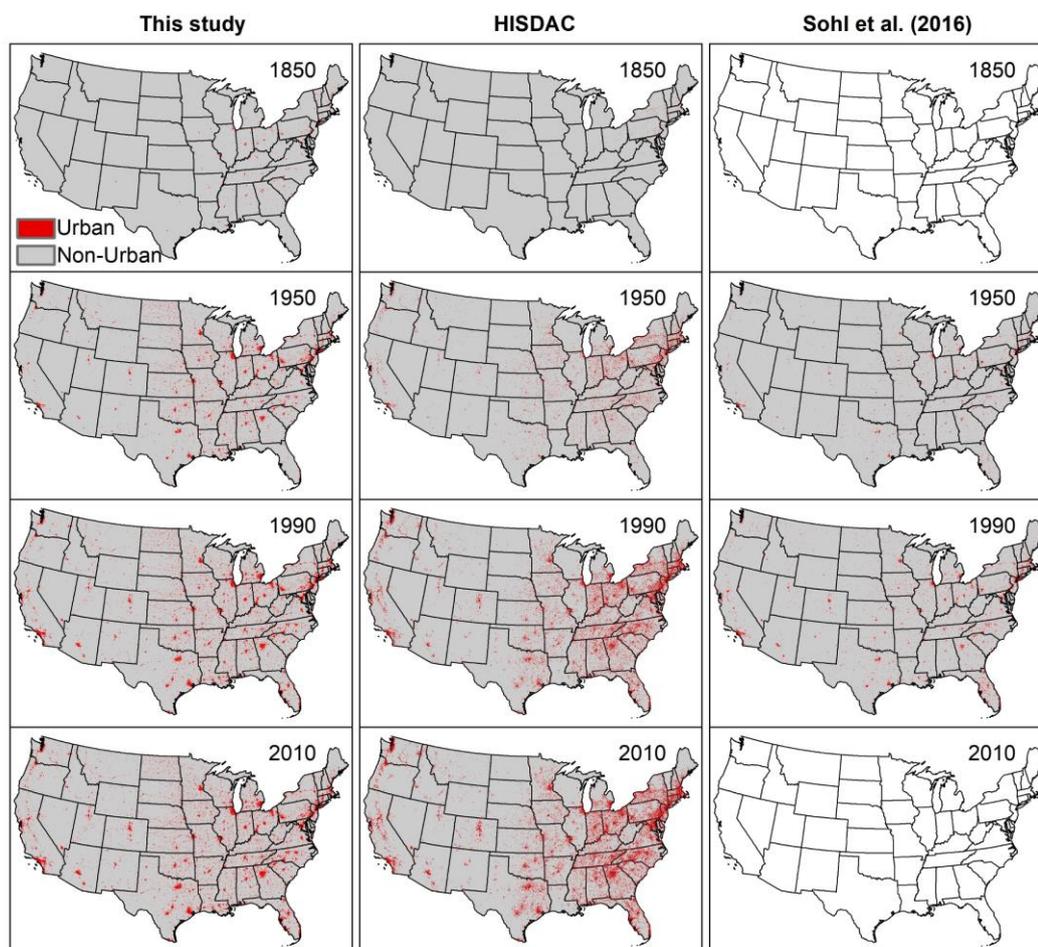
Figure 10: Comparison of urban land maps among three data sets for the conterminous United States: this study (left column), Historical Settlement Data Compilation (HISDAC) map (central column), and Sohl et al. (2016) map (right column).

For the cropland, our result was close to NLCD, YLmap, and ZCmap during 2001–2016 (Figure 9b). Because we used the crop planted area to estimate the physical cropland area, our data were consistent with YLmap during 1850–2016 (Figure 9b). However, the ZCmap and HYDE cropland were higher than our data over the research period (Figure 9b), which could be explained by the fall, idle, and pasture land area counted in ZCmap (Zumkehr and Campbell, 2013). Applying the HYDE cropland historical trend made our result close to it during 1630–1889 (Figure 9b). Spatially, four fractional cropland maps all showed rapid cropland expansion in the Midwest and the Great Plain during 1850–1920 and cropland abandonment in the Northeast and Southeast during 1920–2010 (Figure 11, Figure A8). But our results can reflect the cropland abandonment in New England, the South, and the Southeast (Reuss et al., 1948; Land, 1974; Foster, 1992). Compared with other LULC datasets, our product has higher spatial resolution and more extended temporal coverage, a better understanding of cropland dynamics in CONUS over the past four centuries.

18
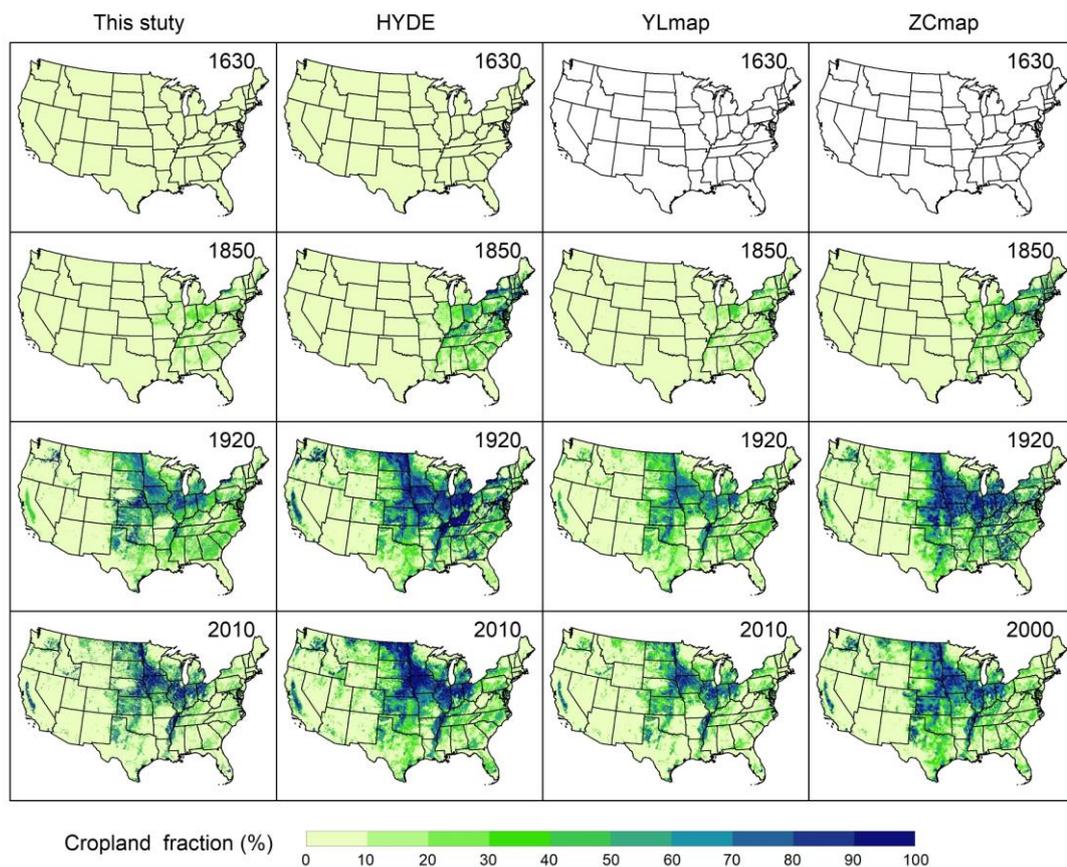
Figure 11: Comparison of cropland maps among four datasets for the conterminous United States: this study (first column), the History Database of Global Environment (HYDE), Yu and Lu (2018) cropland density (YLmap), and Zumkehr and Campbell (2013) cropland fraction (ZCmap).

To the best of our knowledge, accurate temporal and spatially explicit data are still lacking to describe the pasture dynamics for the CONUS. This study set the state-level pasture area from the National Resource Inventory (NRI) as the baseline data for historical pasture reconstruction, which made our data more reliable than HYDE. During 2001–2020, the total national area of pasture located in non-federal land ranged from 48 to 53 Mha, which was close to the NLCD (53 Mha) and HYDE (52 Mha) (Figure 9c). Moreover, Haines et al. (2018) pasture only include hay, making it significantly lower than our result (Figure 9c). The ERS land use data also provided grazing land area, but the rangeland and pasture were not separated (Bigelow and Borchers, 2017). The application of the HYDE pasture historical trend made our result close to it and reached the maximum historical value in the 1950s (Figure 9c). The three maps all show the highest pasture density in eastern Texas, Oklahoma, and Missouri on three maps (Figure 12). Our results characterized the historical changes of pasture with higher spatial resolution than current LULC products.
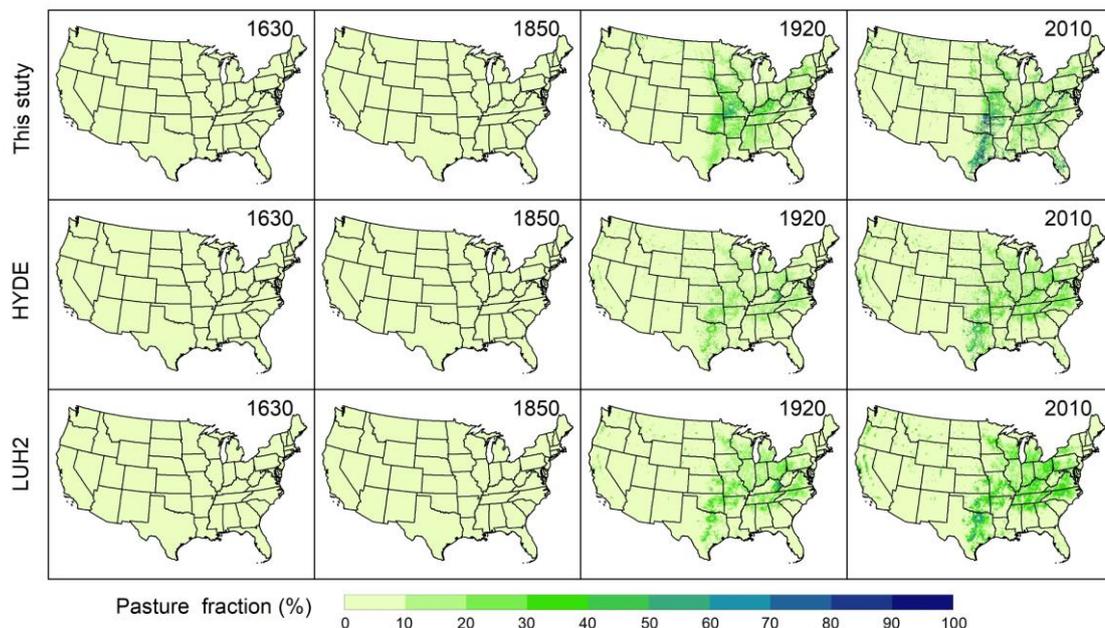
Figure 12: Comparison of pasture patterns and changes among three data products for the conterminous United States: this study
380 **(upper panel), the History Database of Global Environment (HYDE) (middle panel), and Land Use Harmonization (LUH2) (lower panel).**

We used the inventory-based data to reconstruct historical forests, which was more reliable than the satellited-based forest (NLCD) and biomass density-based forest (LUH2). Because NLCD and Sohl et al. (2016) data define forest as the areas dominated by trees generally greater than 5 meters tall and greater than 20% of total vegetation cover, higher than that in our
385 forest definition (forest cover greater than 10%) (Sohl et al., 2016; Homer et al., 2020). Thus, the forest area in this study was higher than the NLCD, Sohl et al. (2016) data (Figure 9d). Moreover, the forest in LUH2 is determined by the vegetation biomass density and country-level forest area (Hurtt et al., 2020), underestimating the forest land in the western US. Previous studies reported deforestation in southern Michigan and forest cutting for agriculture and fuel in Virginia during the early settlement period (Carl, 2012; Mergener et al., 2014), also shown in our maps during the 19th century (Figure A10). Forest
390 loss during the westward expansion period can be captured in the Northeast, Midwest, and Great Plains (Figure 13, Figure A10). We also found the forest regrowth on much cutover and abandoned land in the late 20th century (Foster et al., 1998; MacCleery, 2011) (Figure A10, Figure 8). In addition, our data overcome the underestimation in the Rock Mountain region and Texas in LUH2 forest (Figure 13).
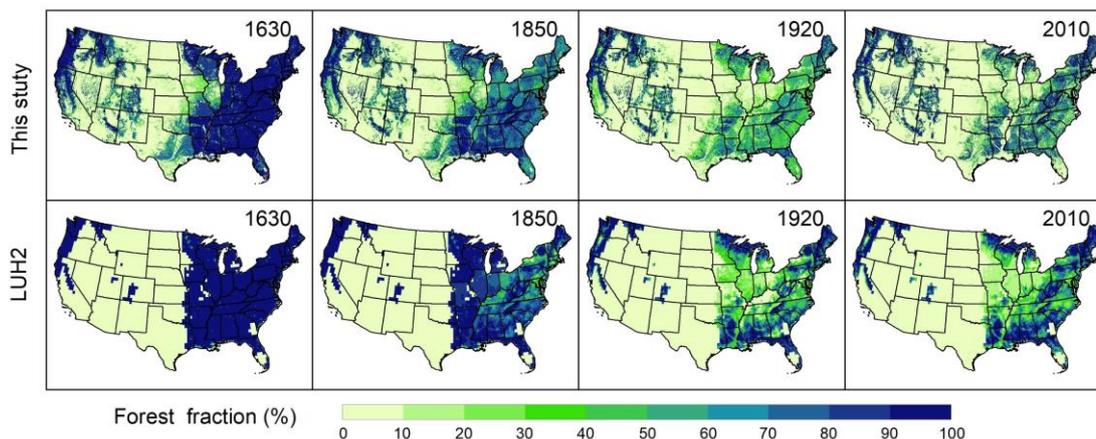
Figure 13: Comparison of forest distribution between this study (upper panel) and Land Use Harmonization (LUH2) (lower panel) for the conterminous United States.

4.2 Drivers of land use and land cover changes

Human activities are the main drivers of land use and land cover changes. In the CONUS, agricultural land changes were influenced by climate, soil productivity, population size, economy, and technological improvement (Waisanen and Bliss, 2002), and agricultural land expansion resulted in forest, shrubland, and grassland loss (Oswalt et al., 2014). The population increased to 3 million from 1630–1775, but more than 97% of people lived in the east of the Appalachian Mountains. Though agriculture had developed in Virginia and Maryland, colonists in the Northeast or Mid-Atlantic region worked in small-scale farming. Therefore, urban land, cropland, and pasture showed a lower increase rate, and forest was the dominant land use type in the Eastern US. Numerous lands like Louisiana, Florida, Texas, Oregon, and New Mexico were acquired during 1800–1860 (Dahl and Allord, 1996; Fretwell, 1996). The westward expansion or movement opened up new agricultural areas and significantly affected the US land use pattern. For example, the rapid inland movement resulted in the conversion of wetland to cropland in the Ohio and Mississippi River Valleys (Dahl and Allord, 1996). In the Mississippi Valley and Midwest, forest land and grassland were cleared and put into growing crops (Steyaert and Knox, 2008; Hanberry et al., 2012). The national policy also had a substantial influence on LULC changes. For example, the Homestead Act issued in 1862 aimed to attract immigrants to develop the Western US (Shannon, 1977). As a result, cropland expanded with a rate of the population increase to produce more food during 1860–1920 (Fred, 1945; MacCleery, 2011). In addition, the development of fertile cropland in the Midwest resulted in the cropland abandonment in the Northeast and South, and the rapid development of railroads drove the growing cities and urban land increase. Though the population has more than tripled since the 1910s, cropland and forest land area maintained stable. The development of hybrid crops and the use of chemical fertilizers improve crop intensive level and productivity, reducing cropland reclamation (MacCleery, 2011). The conservation policy framework issued in the 1930s emphasized the importance of forest protection. Tree planting and stabilizing timber consumption also played an essential role in keeping forests stable (Chen et al., 2017). Though forest clearing for cropland reclamation continued in some states, offset

by cropland abandonment and forest regeneration in other areas, like New England (Foster et al., 1998; Hall et al., 2002) and Wisconsin (Rhemtulla et al., 2009). However, urban land increased at a higher rate, mainly driven by the rapid urbanization

420 and population increase (Leyk et al., 2020).

4.3 Uncertainties and future perspectives

This study aims to reconstruct the spatial and temporal dynamics of LULC changes in the CONUS and how agricultural activities and urbanization induced natural vegetation degradation or loss. Though we have gathered the most reliable land use census or inventory data, the historical census data only records net changes in the area. Moreover, the original census data

425 was recorded at 5–10 years intervals except in recent decades, which made some insignificant fluctuations cannot be captured. The assumptions made in the reconstruction section also increased the data uncertainties. In the generating spatial data section, we assumed that environmental factors-based land use probability was unchanged following previous land use simulation models (Verburg et al., 2006, 2009; Sohl et al., 2014; Liu et al., 2016). Though we improved the land use probability by integrating population density, human settlement extent, and current land use pattern, there were still uncertainties in describing

430 the historical land use trajectory.

More efforts are needed to generate accurate historical LULC maps to understand regional LULC history. On the one hand, more detailed data are required, such as the historical cropland map and survey data, the wood harvest, and tree planting information at the county or site level (Zumkehr and Campbell, 2013; Yu et al., 2019). The subclass of LULC (e.g., tree species, crop types) also needs to reconstruct more accurately (Thompson et al., 2013; Chen et al., 2017; Crossley et al., 2021).

435 On the other hand, land use simulation models should be improved to depict land use dynamics accurately by integrating machine learning methods. In the future, the impact of extreme climate events, war, and policies could be taken into account to better simulate LULC changes.

5 Data availability

The land use and land cover datasets for the conterminous United States are available at

440 https://doi.org/10.5281/zenodo.6469247 (Li et al., 2022). The annual gridded datasets (1km x 1km spatial resolution) with GeoTiff format include fractional and Boolean types. An Excel table is used to organize the annual urban, cropland, pasture, and forest area at the state level. A detailed data description is also provided.

6 Conclusions

This study developed spatially explicit LULC data at a spatial resolution of 1 km x 1 km and an annual time scale in the

445 CONUS during 1630–2020 by integrating multi-sourced data and the machine learning method. The results showed that extensive cropland and pasture expansion and natural vegetation loss occurred from 1630 to 2020 in the CONUS. New

reclaimed cropland was primarily converted from forest, shrubland, and grassland. Tree planting and forest regeneration increased the forest cover in the Northeast and Southern in the recent century. Compared to other LULC datasets, our data provided more accurate information with higher spatial and temporal resolution and better captured the characteristics of LULC

450 changes. The LULC data can be used for regional studies on various topics, including LULC impacts on the ecosystem, biodiversity, water resource, carbon and nitrogen cycles, and greenhouse gas emissions.

Appendices

Table A1: Spatially explicit variables adopted for artificial neural network (ANN) modelling

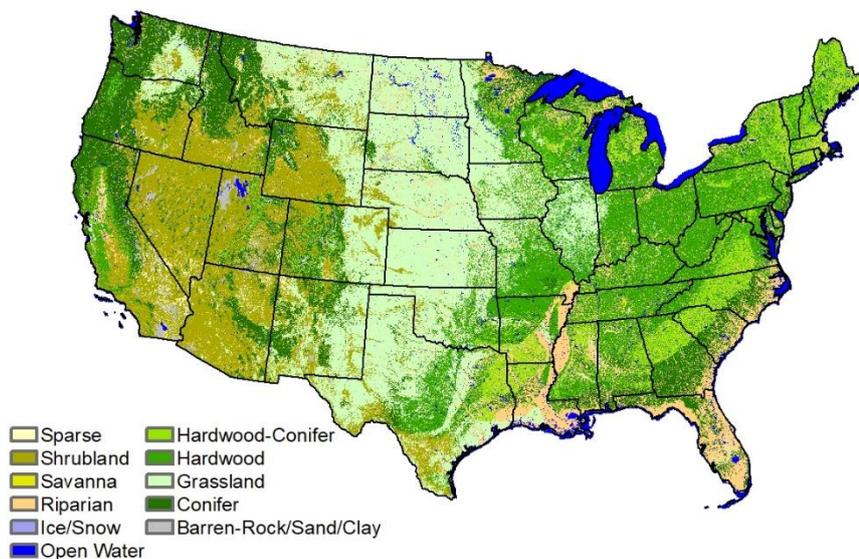| Variable	Description	Source	Resolution
Elevation	Digital elevation model (DEM)	Shuttle Radar Topography Mission (STRM)	90 m
Slope	Slope calculated from DEM	(https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/)	
Pop	Population density	Fang and Jawitz (2018) (http://doi.org/10.6084/m9.figshare.c.3890191)	1 km
$City_{dis}$	Distance to city	https://www.sciencebase.gov/catalog/item/537d23	vector
$Road_{dis}$	Distance to road	fee4b00e1e1a484c82?community=Data+Basin	vector
$Railway_{dis}$	Distance to railway		vector
$River_{dis}$	Distance to river	North America River and Lakes	vector
Soil clay	Soil texture clay fraction	Soil Survey Geographic (SSURGO) Data	250 m
Soil sand	Soil texture sand fraction		250 m
Soil SOC	Soil organic carbon		250 m
Crop PI	Crop productivity index		250 m
PPT	Precipitation	PRISM (https://prism.oregonstate.edu/recent/)	800 m
TMP	Mean temperature		800 m
Max TMP	July temperature		800 m
Min TMP	January temperature		800 m

455

Figure A1: Vegetation type pre-Euro-American settlement.
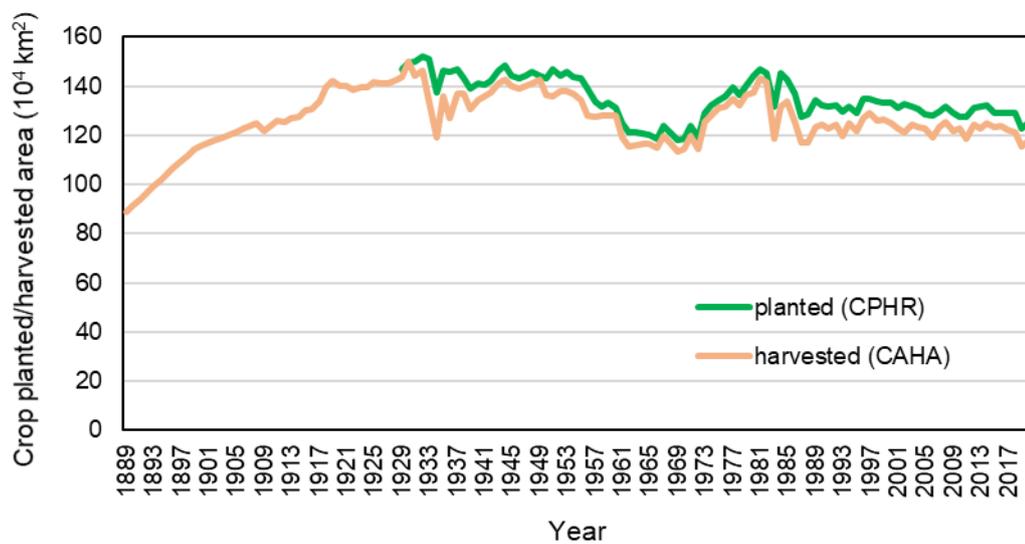
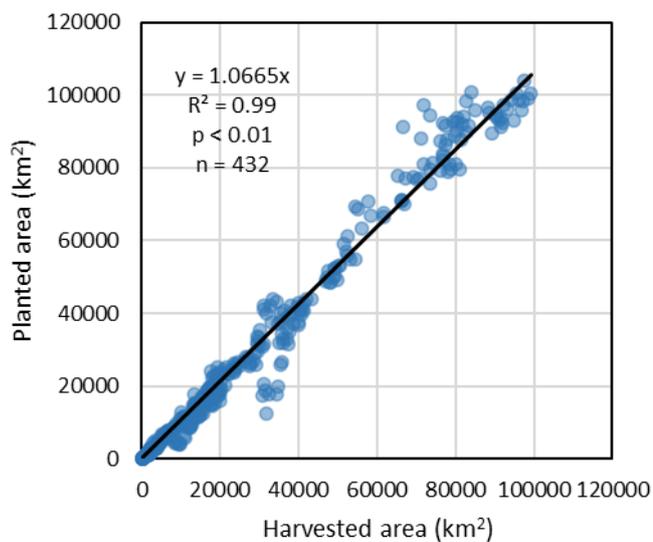460 **Figure A2: National crop harvested and planted area during 1889–2020.**

24

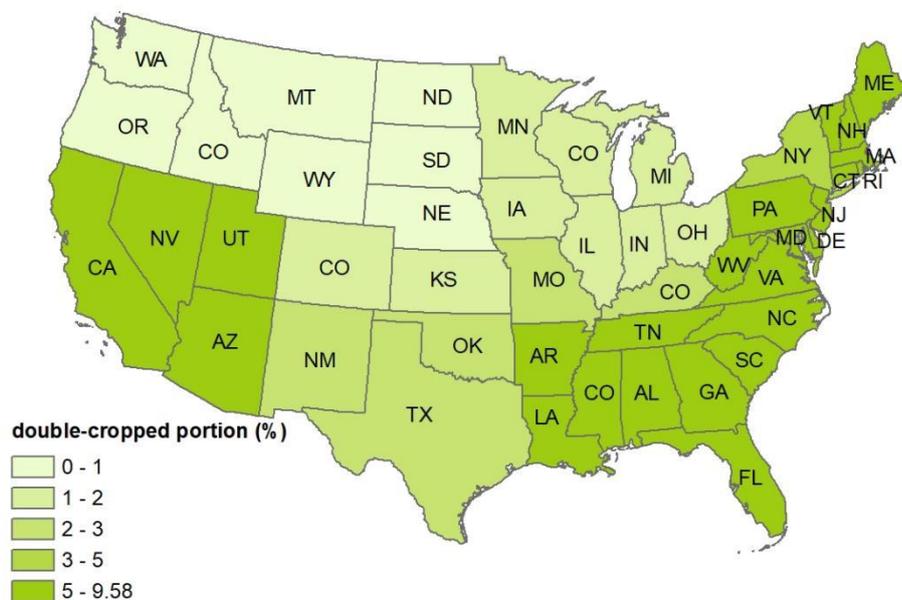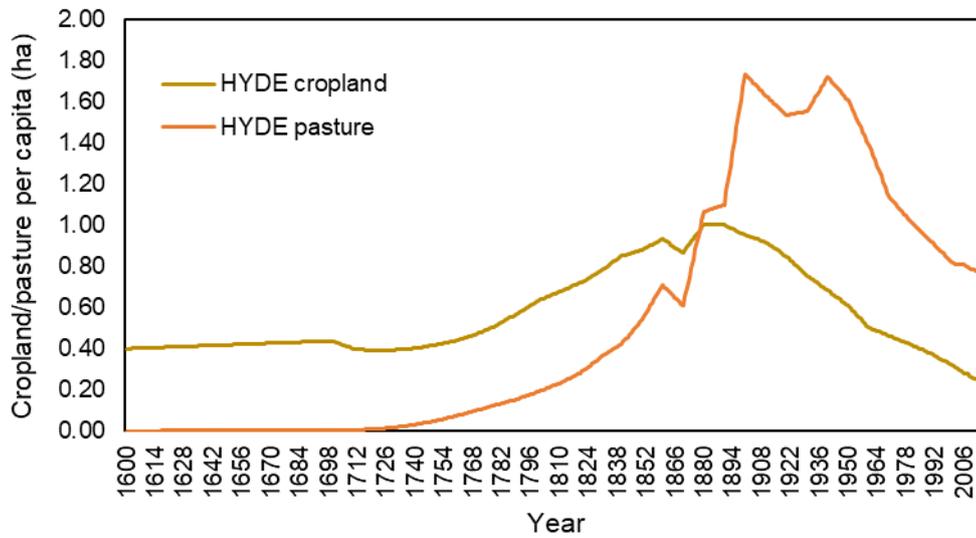**Figure A3: Scatter plot of crop harvested area and planted area during 1974–2017.**

465

Figure A4: Portion of double-cropped area at state-level.

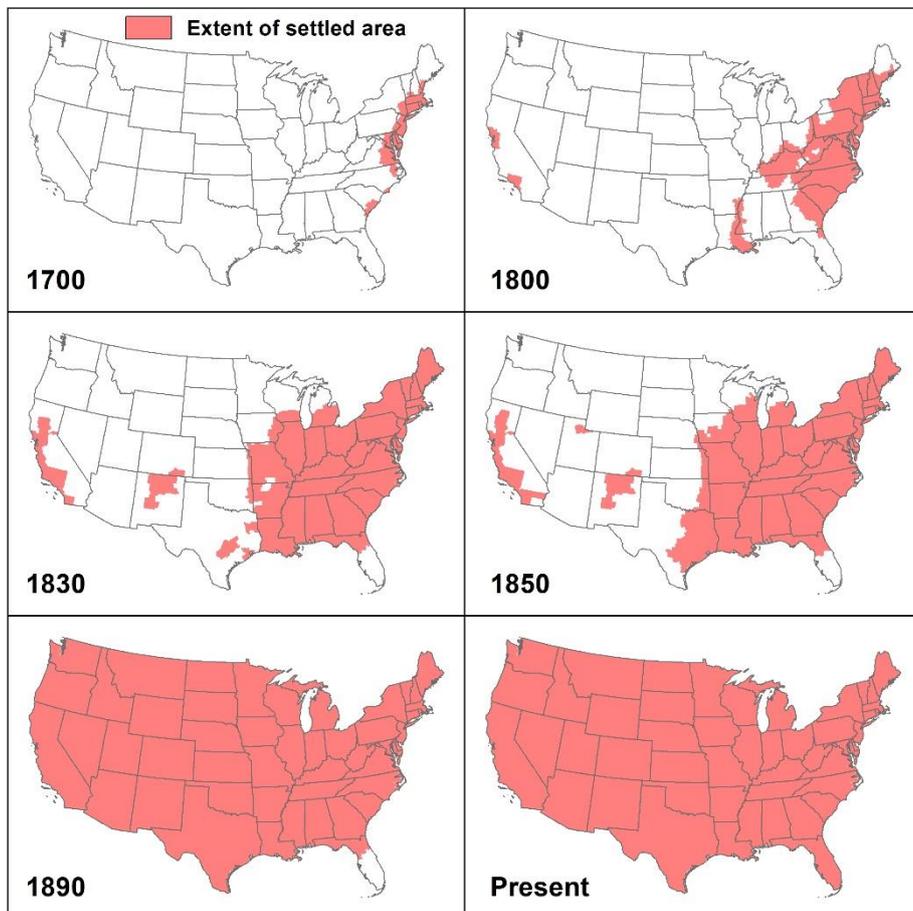Figure A5: Changes of the History Database of Global Environment cropland and pasture per capita during 1600–2017.

Figure A6: Extent of settled area in 1700, 1800, 1830, 1850, 1890, and present.
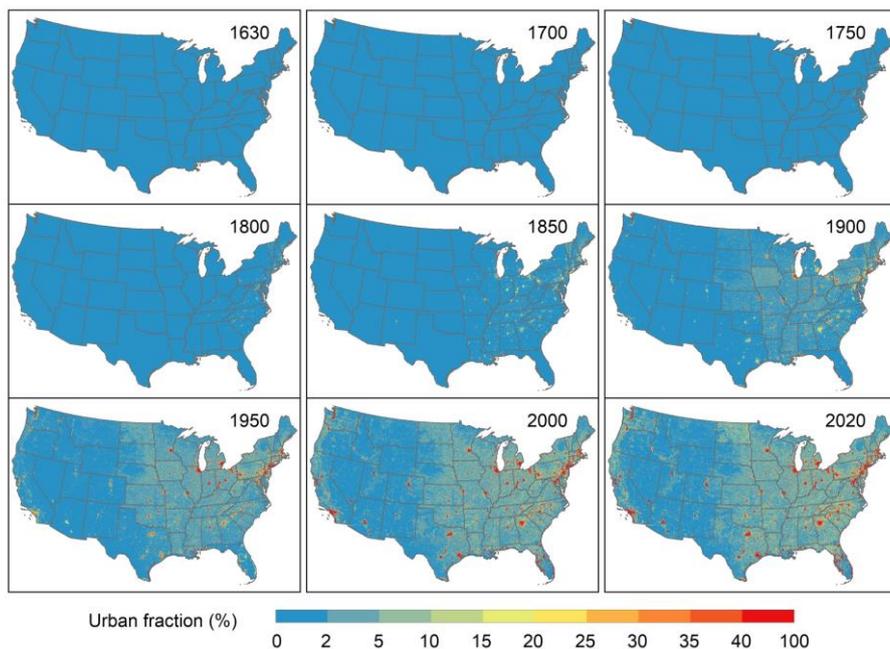
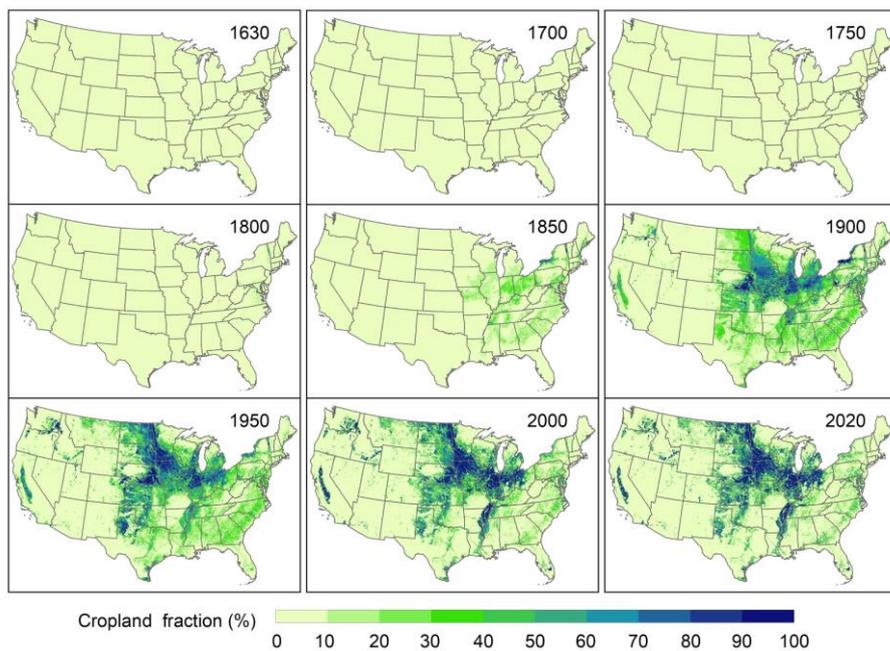Figure A7: Fractional urban land in the conterminous United States during 1630–2020.

475

Figure A8: Fractional cropland in the conterminous United States during 1630–2020.
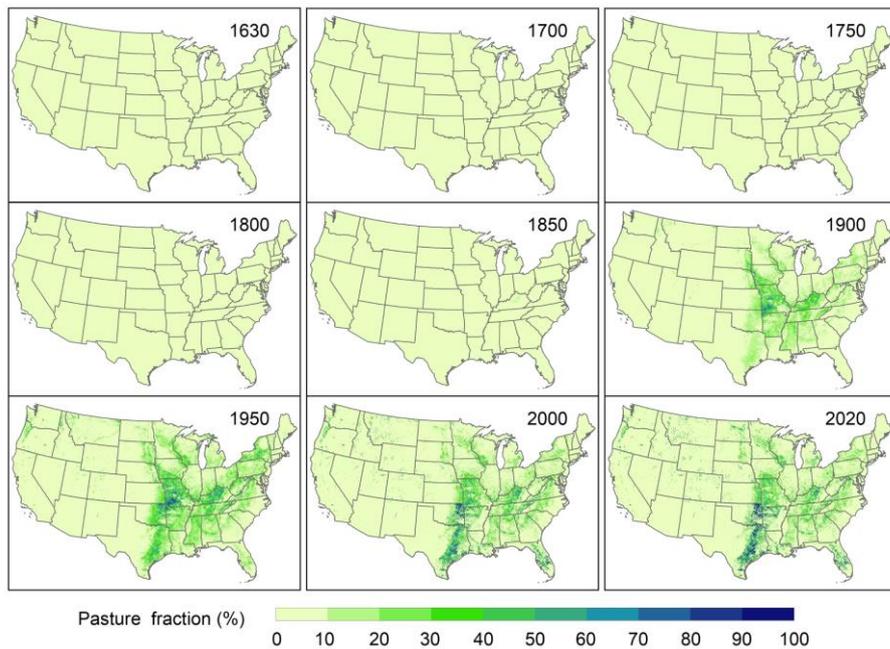
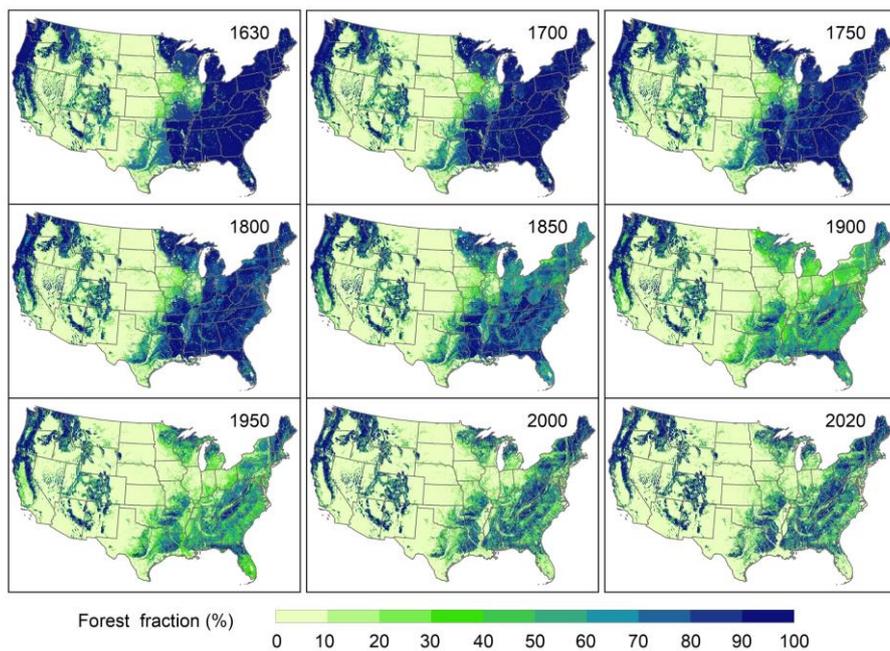**Figure A9: Fractional pasture in the conterminous United States during 1630–2020.**

480

Figure A10: Fractional forest in the conterminous United States during 1630–2020.

Author contributions

HT designed the research; XL implemented the research and analyzed the results; XL, HT, SP, and CL wrote and revised the
485 manuscript.

Competing interests

The authors declare that they have no conflict of interest.

References

495 Bartholome, E. and Belward, A. S.: GLC2000: a new approach to global land cover mapping from Earth observation data, Int.
J. Remote Sens., 26, 1959-1977, http://doi.org/10.1080/01431160412331291297, 2005.

Bigelow, D. P. and Borchers, A.: Major Uses of Land in the United States 2012, U.S. Department of Agriculture, Economic
Research Service, 2017.

Billington, R. A. and Ridge, M.: Westward expansion: a history of the American frontier, University of New Mexico Press,
500 2001.

Borchers, A., Truex-Powell, E., Wallander, S., Nickerson, C.: Multi-Cropping Practices: Recent Trends in Double Cropping,
U.S. Department of Agriculture, Economic Research Service, 2014.

Boryan, C., Yang, Z., Mueller, R., and Craig, M.: Monitoring US agriculture: the US Department of Agriculture, National
Agricultural Statistics Service, Cropland Data Layer Program, Geocarto. Int., 26, 341-358,
505 http://doi.org/10.1080/10106049.2011.562309, 2011.

Chen, G., Pan, S., Hayes, D. J., and Tian, H.: Spatial and temporal patterns of plantation forests in the United States since the
1930s: an annual and gridded data set for regional Earth system modelling, Earth Syst. Sci. Data, 9, 545-556,
http://doi.org/10.5194/essd-9-545-2017, 2017.

Chen, H., Tian, H., Liu, M., Melillo, J., Pan, S., Zhang, C.: Effect of Land-Cover Change on Terrestrial Carbon Dynamics in
510 the Southern United States, J. Environ. Qual., 35, 1533-1547, http://doi.org/10.2134/jeq2005.0198, 2006.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills,
J.: Global land cover mapping at 30 m resolution: A POK-based operational approach, ISPRS. J. Photogramm. Remote Sens.,
103, 7-27, http://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Cole, K. L., Davis, M. B., Stearns, F., Guntenspergen, G., and Walker, K.: Historical landcover changes in the Great Lakes
515 region, U.S. Fish and Wildlife Service, 1999.

Coulson, D. P., Joyce L.: United States state-level population estimates: Colonization to 1999, U.S. Department of Agriculture,
Forest Service, Rocky Mountain Research Station, 2003.

Crossley, M. S., Burke, K. D., Schoville, S. D., and Radeloff, V. C.: Recent collapse of crop belts and declining diversity of
US agriculture since 1840, Glob. Change Biol., 27, 151-164, http://doi.org/10.1111/gcb.15396, 2021.

520 Dahl, T. E.: Wetlands losses in the United States, 1780's to 1980's, U.S. Department of the Interior, Fish and Wildlife Service,
1990.

Dangal, S. R. S., Felzer, B. S., Hurteau, M. D.: Effects of agriculture and timber harvest on carbon sequestration in the eastern
US forests, J. Geophys. Res. Biogeosci., 119, 35-54, http://doi.org/2013JG002409, 2014.

Domke, G. M., Oswalt, S. N., Walters, B. F., and Morin, R. S.: Tree planting has the potential to increase carbon sequestration
525 capacity of forests in the United States, Proc. Natl. Acad. Sci. U. S. A., 117, 24649-24651,
http://doi.org/10.1073/pnas.2010840117, 2020.

Drummond, M. A. and Loveland, T. R.: Land-use pressure and a transition to forest-cover loss in the eastern United States,
BioScience, 60, 286-298, http://doi.org/10.1525/bio.2010.60.4.7, 2010.

Fang, Y. and Jawitz, J. W.: High-resolution reconstruction of the United States human population distribution, 1790 to 2010,
530 Sci. Data, 5, http://doi.org/10.1038/sdata.2018.67, 2018.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., and Snyder, P. K.: Global consequences of land use,
Science, 309, 570-574, http://doi.org/10.1126/science.1111772, 2005.

Foster, D. R., Motzkin, G., and Slater, B.: Land-use history as long-term broad-scale disturbance: regional forest dynamics in
535 central New England, Ecosystems, 1, 96-119, http://doi.org/10.1007/s100219900008, 1998.

Foster, D. R.: Land-Use History (1730-1990) and Vegetation Dynamics in Central New-England, USA, J. Ecol., 80, 753-772,
http://doi.org/10.2307/2260864, 1992.

Fretwell, J. D., Williams, J. S., and Redman, P. J.: National water summary on wetland resources, U.S. Government Printing
Office, http://doi.org/10.3133/wsp2425, 1996.

540 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5
global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168-182,
http://doi.org/10.1016/j.rse.2009.08.016, 2010.

Fuchs, R., Herold, M., Verburg, P. H., and Clevers, J. G. P. W.: A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe, Biogeosciences, 10, 1543-1559, http://doi.org/10.5194/bg-10-1543-2013, 2013.

545

Garrison III, C. E.: Forestry and Tree Planting in Virginia, Tree Planters' Notes, Reforestation, Nurseries, and Genetic Resources, 2012.

Goldewijk, K. K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene - HYDE 3.2, Earth Syst. Sci. Data, 9, 927-953, http://doi.org/10.5194/essd-9-927-2017, 2017a.

550 Goldewijk, K. K., Dekker, S. C., and van Zanden, J. L.: Per-capita estimations of long-term historical land use and the consequences for global change research, J. Land Use Sci., 12, 313-337, http://doi.org/10.1080/1747423x.2017.1354938, 2017b.

Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., and Penman, J.: The key role of forests in meeting climate targets requires science for credible mitigation, Nature Clim. Change, 7, 220-226, https://doi.org/10.1038/nclimate3227, 2017.

555 Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamaki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., Herrero, M., Kiesecker, J., Landis, E., Laestadius, L., Leavitt, S. M., Minnemeyer, S., Polasky, S., Potapov, P., Putz, F. E., Sanderman, J., Silvius, M., Wollenberg, E., and Fargione, J.: Natural climate solutions, Proc. Natl. Acad. Sci. U. S. A, 114, 11645-11650, http://doi.org/10.1073/pnas.1710465114, 2017.

560 Haines, M., Fishback, P., and Rhode, P.: United States Agriculture Data, 1840 – 2012, Inter-university Consortium for Political and Social Research, https://doi.org/10.3886/ICPSR35206.v4, 2018.

Hall, B., Motzkin, G., Foster, D. R., Syfert, M., and Burk, J.: Three hundred years of forest and land-use change in Massachusetts, USA, J. Biogeogr., 29, 1319-1335, http://doi.org/10.1046/j.1365-2699.2002.00790.x, 2002.

Hanberry, B. B., Kabrick, J. M., He, H. S., and Palik, B. J.: Historical trajectories and restoration strategies for the Mississippi

565 River Alluvial Valley, For. Eco. Manag., 280, 103-111, http://doi.org/10.1016/j.foreco.2012.05.033, 2012.

He, F., Li, S., and Zhang, X.: A spatially explicit reconstruction of forest cover in China over 1700-2000, Glob. Planet. Change, 131, 73-81, http://doi.org/10.1016/j.gloplacha.2015.05.008, 2015.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., and Riitters, K.: Conterminous United States land cover change patterns 2001-2016 from the 2016 National Land Cover

570 Database, ISPRS. J. Photogramm. Remote Sens., 162, 184-199, http://doi.org/10.1016/j.isprsjprs.2020.02.019, 2020.

Houghton, R. A., Hackler, J. L., and Lawrence, K. T.: The US carbon budget: Contributions from land-use change, Science, 285, 574-578, http://doi.org/10.1126/science.285.5427.574, 1999.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenoder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T.,

575 Lawrence, D., Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F. N., van Vuuren, D. P., and Zhang, X.: Harmonization of global land use change and management

for the period 850-2100 (LUH2) for CMIP6, Geosci. Model Dev., 13, 5425-5464, http://doi.org/10.5194/gmd-13-5425-2020, 2020.

Hurtt, G. C., Frolking, S., Fearon, M. G., Moore, B., Shevliakova, E., Malyshev, S., Pacala, S. W., and Houghton, R. A.: The

580   underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands, Glob. Change Biol., 12, 1208-1229, http://doi.org/10.1111/j.1365-2486.2006.01150.x, 2006.

Jeon, S. B., Olofsson, P., and Woodcock, C. E.: Land use change in New England: a reversal of the forest transition, J. Land Use Sci., 9, 105-130, https://doi/10.1080/1747423X.2012.754962, 2014.

Land, O.: Water resources, current and prospective supplies and uses, Economic Research Service Miscellaneous Publication,

585   1974.

Lark, T. J., Mueller, R. M., Johnson, D. M., and Gibbs, H. K.: Measuring land-use and land-cover change using the US department of agriculture's cropland data layer: Cautions and recommendations, Int. J. Appl. Earth Obs. Geoinf., 62, 224-235, http://doi.org/10.1016/j.jag.2017.06.007, 2017.

Lark, T. J., Schelly, I. H., and Gibbs, H. K.: Accuracy, Bias, and Improvements in Mapping Crops and Cropland across the

590   United States Using the USDA Cropland Data Layer, Remote Sensing, 13, 968, http://doi.org/10.3390/rs13050968, 2021.

Lark, T. J., Spawn, S. A., Bougie, M., and Gibbs, H. K.: Cropland expansion in the United States produces marginal yields at high costs to wildlife, Nat. Commun., 11, 1-11, http://doi.org/10.1038/s41467-020-18045-z, 2020.

Leyk, S. and Uhl, J. H.: HISDAC-US, historical settlement data compilation for the conterminous United States over 200 years, Sci. Data, 5, 1-14, http://doi.org/10.1038/sdata.2018.175, 2018.

595   Leyk, S., Uhl, J. H., Connor, D. S., Braswell, A. E., Mietkiewicz, N., Balch, J. K., and Gutmann, M.: Two centuries of settlement and urban development in the United States, Sci. Adv., 6, http://doi.org/10.1126/sciadv.aba2937, 2020.

Li, S., He, F., and Zhang, X.: A spatially explicit reconstruction of cropland cover in China from 1661 to 1996, Regional Environmental Change, 16, 417–428, https://doi.org/10.1007/s10113-014-0751-4, 2016.

Li, X., Tian, H., Pan, S., and Lu, C.: Land use and land cover changes in the contiguous United States from 1630 to 2020

600   (v1.1), Zenodo, https://doi.org/10.5281/zenodo.6469247, 2022.

Li, X., Yu, L., Sohl, T., Clinton, N., Li, W., Zhu, Z., Liu, X., and Gong, P.: A cellular automata downscaling based 1 km global land use datasets (2010–2100), Science Bulletin, 61, 1651-1661, 10.1007/s11434-016-1148-1, https://doi.org/10.1007/s11434-016-1148-1, 2016.

Liu, M. and Tian, H.: China's land cover and land use change from 1700 to 2005: Estimations from high-resolution satellite

605   data and historical archives, Global Biogeochem. Cy., 24, GB3003, http://doi.org/10.1029/2009gb003687, 2010.

Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S., and Pei, F.: A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects, Landsc. Urban Plan., 168, 94-116, http://doi.org/10.1016/j.landurbplan.2017.09.019, 2017.

MacCleery, D. W.: American forests: a history of resiliency and recovery, U.S. Department of Agriculture, Forest Service,

610   2011.

Mergener, R., Botti, W., and Heyd, R.: Forestry and Tree Planting in Michigan, Tree Planters' Notes, Reforestation, Nurseries, and Genetic Resources, 2014.

Oswalt, S. N., Smith, W. B., Miles, P. D. and Pugh, Scott, A.: Forest Resources of the United States, 2012: a technical document supporting the Forest Service 2010 update of the RPA Assessment, U.S. Department of Agriculture, Forest Service, Washington Office, 2014.

Oswalt, S. N.; Miles, Patrick D.; Pugh, Scott A.; Smith, W. Brad. Forest Resources of the United States, 2017: a technical document supporting the Forest Service 2020 RPA Assessment, U.S. Department of Agriculture, Forest Service, Washington Office, 2019.

Pan, Y., Chen, J., Birdsey, R., McCullough, K., He, L., and Deng, F.: Age structure and disturbance legacy of North American forests, Biogeosciences, 8, 715-732, http://doi.org/10.5194/bg-8-715-2011, 2011.

Peng, S., Ciais, P., Maignan, F., Li, W., Chang, J., Wang, T., and Yue, C.: Sensitivity of land use change emission estimates to historical land use and land cover mapping, Global Biogeochem. Cy., 31, 626-643, http://doi.org/10.1002/2015gb005360, 2017.

Reuss, L. A.: Inventory of Major Land Uses in the United States, U.S. Department of Agriculture, 1948.

Rhemtulla, J. M., Mladenoff, D. J., and Clayton, M. K.: Legacies of historical land use on regional forest composition and structure in Wisconsin, USA (mid-1800s-1930s-2000s), Ecol. Appl., 19, 1061-1078, http://doi.org/10.1890/08-1453.1, 2009.

Rollins, M. G.: LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment, Int. J. Wildland Fire, 18, 235-249, http://doi.org/10.1071/WF08088, 2009.

Shannon, F. A.: The farmer's last frontier: agriculture, 1860-1897, M. E. Sharpe, 1977.

Sohl, T. L., Sayler, K. L., Bouchard, M. A., Reker, R. R., Friesz, A. M., Bennett, S. L., Sleeter, B. M., Sleeter, R. R., Wilson, T., and Soulard, C.: Spatially explicit modeling of 1992–2100 land cover and forest stand age for the conterminous United States, Ecol. Appl., 24, 1015-1036, https://doi/10.1890/13-1245.1, 2014.

Sohl, T., Reker, R., Bouchard, M., Sayler, K., Dornbierer, J., Wika, S., Quenzer, R., and Friesz, A.: Modeled historical land use and land cover for the conterminous United States, J. Land Use Sci., 11, 476-499, https://doi/10.1080/1747423x.2016.1147619, 2016.

Stanturf, J. A., Palik, B. J., and Dumroese, R. K.: Contemporary forest restoration: A review emphasizing function, For. Ecol. Manag., 331, 292-323, https://doi/10.1016/j.foreco.2014.07.029, 2014.

Steyaert, L. T. and Knox, R. G.: Reconstructed historical land cover and biophysical parameters for studies of land-atmosphere interactions within the eastern United States, J. Geophys. Res. Atom., 113, D02101, https://doi/10.1029/2006jd008277, 2008.

Thompson, J. R., Carpenter, D. N., Cogbill, C. V., and Foster, D. R.: Four Centuries of Change in Northeastern United States Forests, Plos One, 8, e72540, https://doi/10.1371/journal.pone.0072540, 2013.

Tian, H., Chen, G., Zhang, C., Liu, M., Sun, G., Chappelka, A., Ren, W., Xu, X., Lu, C., and Pan, S.: Century-scale responses of ecosystem carbon storage and flux to multiple environmental changes in the southern United States, Ecosystems, 15, 674-694, https://doi/10.1007/s10021-012-9539-x, 2012.

645   Tian, H., Banger, K., Tao, B, Dadhwal, V. K.: History of land use in India during 1880-2010: Large-scale land transformation reconstructed from satellite data and historical achieves, Glob. Planet. Change, 121, 76-88, https://doi:10.1016/j.gloplacha.2014.07.005, 2014.

U.S. Department of Agriculture.: Summary Report: 2017 National Resources Inventory, Natural Resources Conservation Service Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa, 2020.

650   Uhl, J. H., Leyk, S., McShane, C. M., Braswell, A. E., Connor, D. S., and Balk, D.: Fine-grained, spatiotemporal datasets measuring 200 years of land development in the United States, Earth Syst. Sci. Data, 13, 119-153, https://doi/10.5194/essd-2020-217, 2021.

Verburg, P. H. and Overmars, K. P.: Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model, Landscape Ecol., 24, 1167-1181, https://doi/10.1007/s10980-655   009-9355-7, 2009.

Verburg, P. H., Schulp, C. J. E., Witte, N., and Veldkamp, A.: Downscaling of land use change scenarios to assess the dynamics of European landscapes, Agr. Ecosyst. Environ., 114, 39-56, https://doi/10.1016/j.agee.2005.11.024, 2006.

Waisanen, P. J. and Bliss, N. B.: Changes in population and agricultural land in conterminous United States counties, 1790 to 1997, Global Biogeochem. Cy., 16, 84-81-84-19, https://doi/10.1029/2001gb001843, 2002.

660   West, T. O., Page, Y. L., Huang, M., Wolf, J., and Thomson, A. M.: Downscaling global land cover projections from an integrated assessment model for use in regional analyses: results and evaluation for the US from 2005 to 2095, Environ. Res. Lett., 9, 064004, https://doi/10.1088/1748-9326/9/6/064004, 2014.

Winkler, K., Fuchs, R., Rounsevell, M., and Herold, M.: Global land use changes are four times greater than previously estimated, Nat. Commun., 12, 1-10, https://doi.org/10.1038/s41467-021-22702-2, 2021.

665   Williams, C. A., Gu, H., MacLean, R., Masek, J. G., and Collatz, G. J.: Disturbance and the carbon balance of US forests: A quantitative review of impacts from harvests, fires, insects, and droughts, Glob. Planet. Change, 143, 66-80, https://doi/10.1016/j.gloplacha.2016.06.002, 2016.

Williams, C. A., Gu, H., MacLean, R., Masek, J. G., and Collatz, G. J.: Disturbance and the carbon balance of US forests: A quantitative review of impacts from harvests, fires, insects, and droughts, Glob. Planet. Change, 143, 66-80, 670   https://doi/10.1016/j.gloplacha.2016.06.002, 2016.

Yang, J., Tao, B., Shi, H., Ouyang, Y., Pan, S., Ren, W., and Lu, C.: Integration of remote sensing, county-level census, and machine learning for century-long regional cropland distribution data reconstruction, Int. J. Appl. Earth Obs. Geoinf., 91, 102151, https://doi/10.1016/j.jag.2020.102151, 2020.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., 675   Granneman, B., Liknes, G. C., Rigge, M., and Xian, G.: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies, ISPRS. J. Photogramm. Remote Sens., 146, 108-123, https://doi/10.1016/j.isprsjprs.2018.09.006, 2018.

Yu, Z. and Lu, C.: Historical cropland expansion and abandonment in the continental U.S. during 1850 to 2016, Glob. Ecol. Biogeogr., 27, 322-333, https://doi/10.1111/geb.12697, 2018.

680   Yu, Z., Lu, C., Tian, H., and Canadell, J. G.: Largely underestimated carbon emission from land use and land cover change in the conterminous United States, Glob. Change Biol., 25, 3741-3752, https://doi/10.1111/gcb.14768, 2019.

Zumkehr, A. and Campbell, J. E.: Historical U.S. Cropland areas and the potential for bioenergy production on abandoned croplands, Environ. Sci. Technol., 47, 3840-3847, https://doi/10.1021/es3033132, 2013.