Supplement of Earth Syst. Sci. Data, 15, 1005–1035, 2023
https://doi.org/10.5194/essd-15-1005-2023-supplement

Open Access Earth System Science Data

Supplement of

Four-century history of land transformation by humans in the United States (1630–2020): annual and 1 km grid data for the HIStory of LAND changes (HISLAND-US)

Xiaoyong Li et al.

Correspondence to: Hanqin Tian (hanqin.tian@bc.edu)

Table S1: Input datasets for the urban land reconstruction.

Data	Year	Variable	Data source
National Land Cover Database (NLCD)	2001, 2004, 2006, 2008, 2011, 2013, 2016, 2019	Developed land: Open Space; Low Intensity; Medium Intensity; and High intensity (30 m)	https://www.mrlc.gov/data (last access: Aug 11, 2022)
Historical Settlement Data Compilation (HISDAC) for the United States	1810-2015 (5-year interval)	Historical built-up areas (BUA) (250 m)	https://dataverse.harvard.edu/dataverse/hisdacus (last access: Aug 11, 2022)
Population	1630-1810 (annual)	Total population (State level)	Coulson and Joyce (2003). United States State-level Population Estimates: Colonization to 1999.

Table S2: Input datasets for the cropland reconstruction.

Data	Year	Variable	Data source
USDA Census of Agriculture Historical Archive (CAHA)	1879, 1889, 1899, 1909, 1919, 1924, 1929, 1934, 1939, 1944, 1949, 1954, 1959, 1964, 1969, 1974, 1978, 1982, 1987, 1992, 1997, 2002, 2007, 2012, 2017	Cropland harvested area (State level)	https://agcensus.mannlib.cornell.edu/AgCensus/homepage.do (last access: Jul 10, 2022)
USDA Economic Research Service (ERS)	1910-2020 (annual)	Cropland harvested area (National level)	https://www.ers.usda.gov/data-products/major-land-uses/ (last access: Jul 10, 2022)
History Database of the Global Environment (HYDE) v3.2 baseline	1600, 1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880	Cropland area (5 arcmin)	https://landuse.sites.uu.nl/datasets/ (last access: Feb 13, 2023)
HYDE population	1600, 1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880	Population (5 arcmin)	https://landuse.sites.uu.nl/datasets/ (last access: Feb 13, 2023)
Population	1630-2020	Total population (State-level)	Coulson and Joyce (2003). United States State-level Population Estimates: Colonization to 1999.

Table S3: Input datasets for the pasture reconstruction.

Data	Year	Variable	Data source
National Resource Inventory (NRI)	1982, 1987, 1992, 1997, 2002, 2007, 2012, 2017	Pasture (State level)	https://agcensus.mannlib.cornell.edu/AgCensus/homepage.do (last access: Jul 10, 2022)
History Database of the Global Environment (HYDE) v3.2 baseline	1600, 1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990	Pasture (5 arcmin)	https://landuse.sites.uu.nl/datasets/ (last access: Feb 13, 2023)
HYDE population	1600, 1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880	Population (5 arcmin)	https://landuse.sites.uu.nl/datasets/ (last access: Feb 13, 2023)
Population	1630-2020	Total population (State-level)	Coulson and Joyce (2003). United States State-level Population Estimates: Colonization to 1999.

Table S4: Input datasets for the forest reconstruction.

Data	Year	Variable	Data source
United States Department of Agriculture, Forest Resources of the United States, 2017 (USDA-FR)	1630, 1907, 1920, 1938, 1953, 1963, 1977, 1987, 1997, 2007, 2012, 2017	Forest area (State level)	https://www.fs.usda.gov/treesearch/pubs/57903 (last access: Mar 14, 2022)
Forest Inventory Analysis Trend data: 1630-2000 US forest area and population	1630, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000	Forest area (State-level)	https://www.fia.fs.usda.gov/ (last access: Feb 13, 2023)

Table S5: Definition of cropland in existing literatures.

Data source	Definition
USDA-ERS	***Cropland*:** Total cropland includes five components: cropland harvested, crop failure, cultivated summer fallow, cropland used only for pasture, and idle cropland (https://www.ers.usda.gov/data-products/major-land-uses/glossary/#croplandforcrops, last access: Aug 21, 2022).
USDA-NRI	***Cropland*:** A land cover/use category that includes areas used for the production of adapted crops for harvest. Two subcategories of cropland are recognized: cultivated and non-cultivated. Cultivated land comprises land in row crops or close-grown crops, as well as other cultivated cropland; for example, hayland or pastureland that is in a rotation with row or close-grown crops. Non-cultivated cropland includes permanent hayland and horticultural cropland.
NLCD	***Cultivated Crops*:** areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled (https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description, last access: Aug 21, 2022).
HYDE	FAO categories of "arable land and permanent crops" (Klein Goldewijk et al., 2017).

Table S6: Definition of grazing land, pasture, and rangeland in existing literatures.

Data source	Definition
USDA-ERS	***Cropland pasture*:** Cropland pasture includes acres of crops hogged or grazed but not harvested and some land used for pasture that could have been cropped without additional improvement. ***Grassland pasture and range*:** Grassland pasture and range encompass all open land used primarily for pasture and grazing, including shrub and brush-land types of pasture, grazing land with sagebrush and scattered mesquite, and all tame and native grasses, legumes, and other forage used for pasture or grazing—regardless of ownership. ***Forest land grazed***: Forested pasture and range consisting mainly of forest, brush-grown pasture, arid woodlands, and other areas within forested areas that have grass or other forage growth. https://www.ers.usda.gov/data-products/major-land-uses/glossary/ (last access: Aug 21, 2022)
USDA-NRI	***Pasture***: A land cover/use category of land managed primarily for the production of introduced forage plants for livestock grazing. Pastureland cover may consist of a single species in a pure stand, a grass mixture, or a grass-legume mixture. Management usually consists of cultural treatments: fertilization, weed control, reseeding, renovation, and control of grazing. For the NRI, includes land that has a vegetative cover of grasses, legumes, and/or forbs, regardless of whether or not it is being grazed by livestock (U.S. Department of Agriculture, 2020). ***Rangeland***: A broad land cover/use category on which the climax or potential plant cover is composed principally of native grasses, grass-like plants, forbs or shrubs suitable for grazing and browsing, and introduced forage species that are managed like rangeland. This would include areas where introduced hardy and persistent grasses, such as crested wheatgrass, are planted and such practices as deferred grazing, burning, chaining, and rotational grazing are used, with little or no chemicals or fertilizer being applied. Grasslands, savannas, many wetlands, some deserts, and tundra are considered to be rangeland. Certain communities of low forbs and shrubs, such as mesquite, chaparral, mountain shrub, and pinyon-juniper, are also included as rangeland (U.S. Department of Agriculture, 2020).
EPA	***Pastures*:** Pastures are those lands that are primarily used for the production of adapted, domesticated forage plants for livestock. ***Rangelands*:** Rangelands are those lands on which the native vegetation (climax or natural potential plant community) is predominantly grasses, grass-like plants, forbs, or shrubs suitable for grazing or browsing use. Rangelands include natural grassland, savannas, many wetlands, some deserts, tundra, and certain forb and shrub communities. https://www.epa.gov/agriculture/agricultural-pasture-rangeland-and-grazing (last access: Aug 21, 2022)
NLCD	***Pasture/Hay*:** Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation. https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description (last access: Aug 21, 2022)
HYDE	***Grazing land*:** Land used for mowing or grazing livestock, based on the FAO category "permanent meadows and pastures". Grazing land can be a variety of ecosystems, ranging from managed irrigated grasslands to unmanaged open savannah woodlands to semi-shrub/scrub, almost desert, lands (Klein Goldewijk et al., 2017). ***Pasture*:** Pasture is high-intensity grazing land, or low intensity grazing lands where a conversion of the natural vegetation has occurred (Klein Goldewijk et al., 2017). ***Rangeland*:** rangeland is low-intensity grazing land where the natural vegetation has not been converted (Klein Goldewijk et al., 2017).

Table S7. Forest definitions from different data sources.

Data source	Definition
FIA	Land at least 10 percent stocked by forest trees of any size, or formerly having such tree cover, and not currently developed for non-forest uses, with a minimum area classification of 1 acre (https://cfpub.epa.gov/roe/definitions.cfm?i=51, last access: Feb 13, 2023).
NLCD	Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover (https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description, last access: Aug 21, 2022).
LUH2	Forest was defined using a single tree canopy cover threshold to match the global forest extent provided by the FAO FRA report (Hurtt et al., 2020).

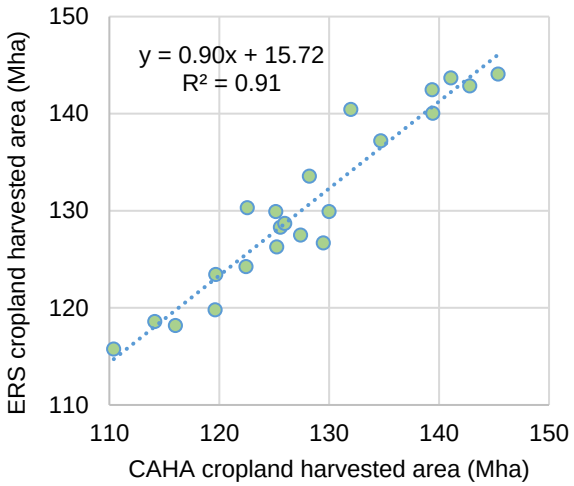

Figure S1: Comparison between the ERS cropland harvested area (without double-cropped area) and CAHA cropland harvested area between 1910 and 2017.

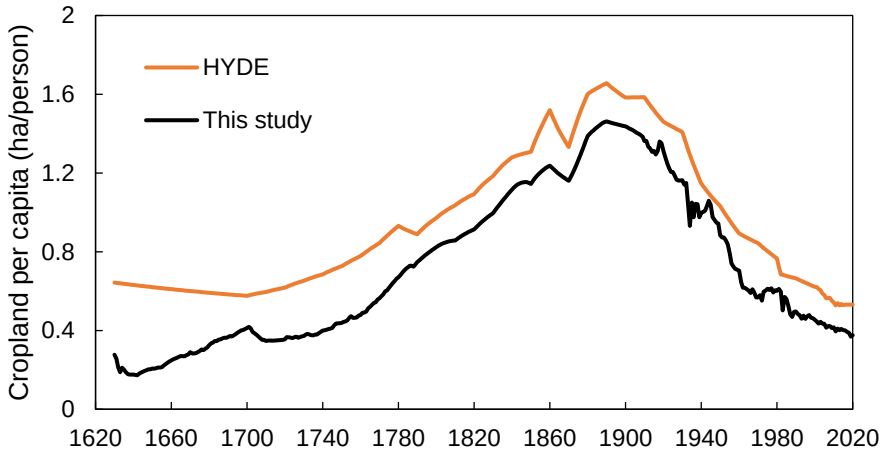

Figure S2: Nation-level cropland per capita comparison between HYDE and this study between 1630 and 2020.

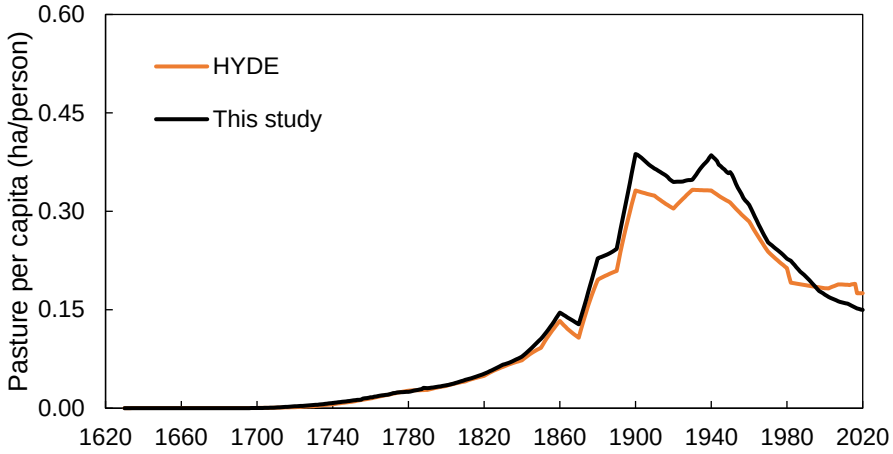

Figure S3: Nation-level pasture per capita comparison between HYDE and this study between 1630 and 2020.

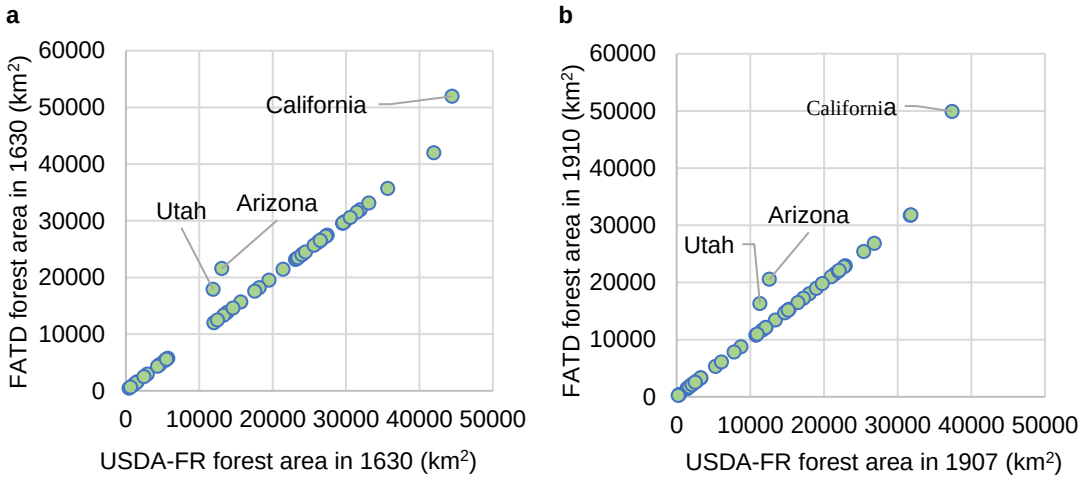

Figure S4: Forest area comparison between USDA-FR and FATD data in 1630 and 1907.

Generating historical population gird data

In the section 2.3.1, we use the population density to improve the LULC probability. However, the county-level population density data only have one value for each county, resulted in obvious political boundaries in the probability data. Therefore, a simple method was applied to generate the historical gridded population by combing the gridded population data with 1-km resolution in 2000 and county-level population, which can be expressed as:

$$Pop_{i,t} = \frac{Pop_{i,t}^{county}}{Pop_{i,2000}^{county}} \times Pop_{i,2000}^{grid}$$

where, $Pop_{i,t}$ is improved population density at grid cell i and year t; $Pop_{i,2000}^{county}$ and $Pop_{i,t}^{county}$ is the county-level population density at grid cell i in 2000 and year t; $Pop_{i,2000}^{grid}$ is the gridded population density at grid cell i in 2000 and year t. The gridded population density data in 2000 was download from https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11.

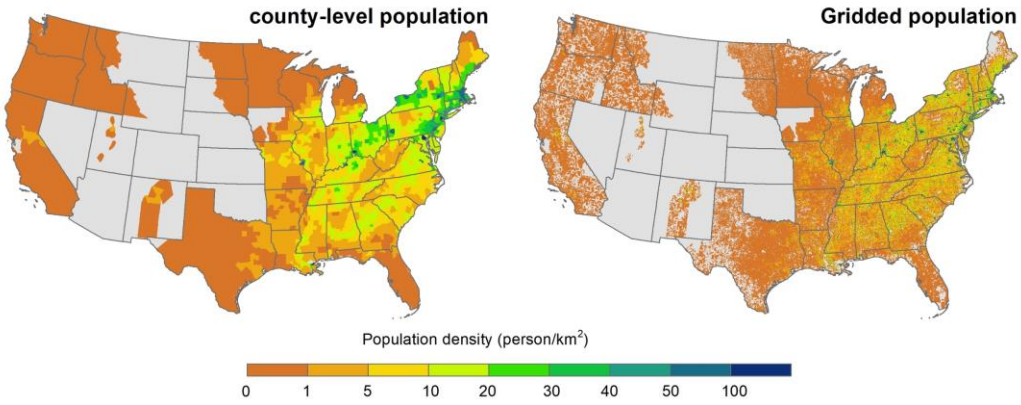

Figure S5: Comparison between county-level and gridded population density data in 1850.

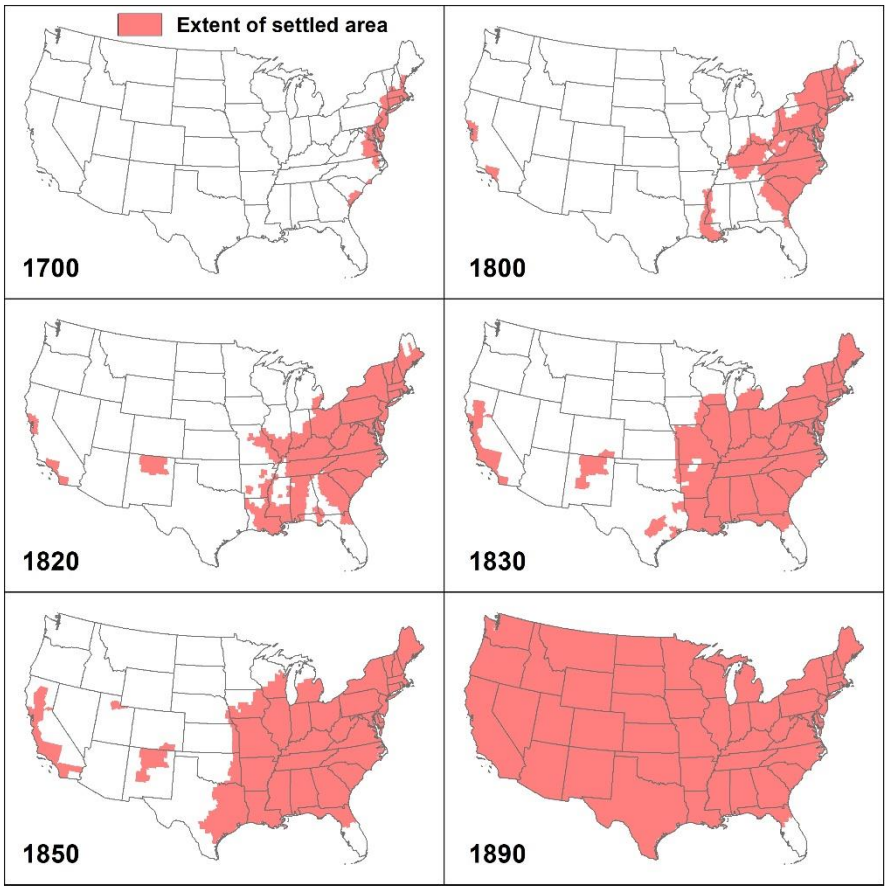

Figure S6: Extent of settled area in 1700, 1800, 1820, 1830, 1850, 1890.

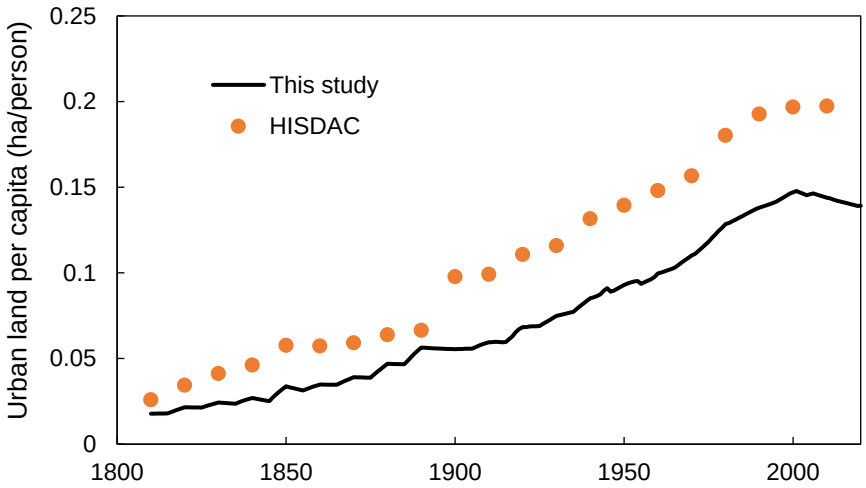

Figure S7: Urban land per capita change between 1810 and 2020. The value indicated by the orange dot is the urban land per capita derived from HISDAC built-up areas; the value indicated by the black line is the urban land per capita derived from the newly developed urban land.

Uncertainties of data integration

In this study, four major land use and cover types (urban, cropland, pasture, and forest) were reconstructed by integrating multisource datasets. Due to the differences among the datasets, some uncertainties would be introduced to the model in the data integration process.

Urban land

In the ***Result*** and ***Discussion*** section, we compared the Historical Settlement Data Compilation for the United States (HISDAC-US) built-up area and the newly developed urban land. The results showed that the urban land area derived from the HISDAC data was higher than that from our data. It is because the HISDAC data is rebuilt using the detailed property records and have a relatively coarse resolution. For example, the national total urban land area from this study is about 73% of that from HISDAC data between 2001 and 2015.

Considering the differences in the urban land area, we applied the annual change rate rather than the absolute value of HISDAC data as the input to reconstruct the historical urban land area for 1810-2001. We assumed that the HISDAC data could accurately capture the urban land expansion trends. To quantify the uncertainties, we calculated the relative difference in area change rate ($HISDAC_{t1}/HISDAC_{t2}$) between our reconstruction and HISDAC data in the overlap period (2001-2015), which can be expressed as follows:

$$RD_{urban,t} = \left|\frac{HistUrban\_CR_t - HISDAC\_CR_t}{HistUrban\_CR_t}\right| \times 100\% \quad \text{(S1)}$$

Where $RD_{urban,t}$ refers the relative difference in area change rate between the HISDAC data and the newly developed urban land data; $HistUrban\_CR_t$ and $HISDAC\_CR_t$ are the area change rate derived from the newly developed urban land and HISDAC data.

The mean relative difference in area change rate between the two datasets at the national level is 3.83% during 2001-2015 (Figure S8a). Because there are only four overlap time points (2001, 2005, 2010, and 2015), we further calculated the state-level relative difference (Figure S8b). The results show that the mean relative difference of the 48 states in 2001-2005, 2005-2010, 2010-2015, and 2001-2015 are 3.34±1.90%, 1.75±1.65%, 0.71±0.97%, and 1.93±1.89%, respectively (Figure S8b). Thus, the uncertainty induced by data difference should be little.

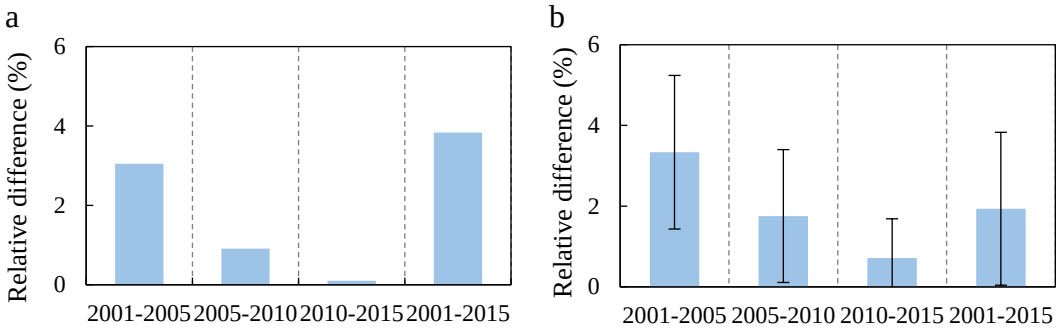

Figure S8: (a) National level relative difference in area change rate between the newly developed urban land data and HISDAC data during 2001-2005, 2005-2010, and 2010-2015. (b) State-level mean relative difference in area change rate between the newly developed urban land data and HISDAC data during 2001-2005, 2005-2010, 2010-2015, and 2001-2015.

Cropland

Four datasets, including the USDA Economic Research Service (ERS) cropland harvested area, USDA Census of Agriculture Historical Archive (CAHA), HYDE3.2 cropland, and total population, were used to reconstruct the historical cropland area. For 1910-2020, we used the ERS cropland harvested area (national level) to subtract the double-cropped area and optimize the interannual variations of CAHA cropland harvested area data. To quantify the uncertainties, we calculated the relative difference between the newly developed cropland area and the CAHA cropland harvested area, which can be expressed as:

$$RD_{crop,t} = \left|\frac{HistCrop_t - CHA_t}{HistCrop_t}\right| \times 100\% \quad \text{(S2)}$$

Where $RD_{crop,t}$ refers the relative difference in cropland area between the newly developed cropland data and CAHA cropland harvested area data in year *t*; $HistCrop_t$ is the cropland area derived from the newly developed cropland data in year *t*; CHA_t is the CAHA cropland harvested area in year *t*.

We found that the mean relative difference in cropland harvested area between the two datasets is 2.23±1.18%, but the data between the 1960s and the 1980s had relatively large differences ranged from 1.81% to 6.02% (Figure S9a). Therefore, the uncertainty induced by cropland area adjustment is little.

During 1879-1909, we adjusted the CAHA cropland harvested area based on the reconstructed cropland area between 1910-2020. The mean relative difference (1.02± 0.19%) between the CAHA cropland harvested area and the newly developed cropland (Figure S9b), indicating that little

uncertainty was introduced to the model.

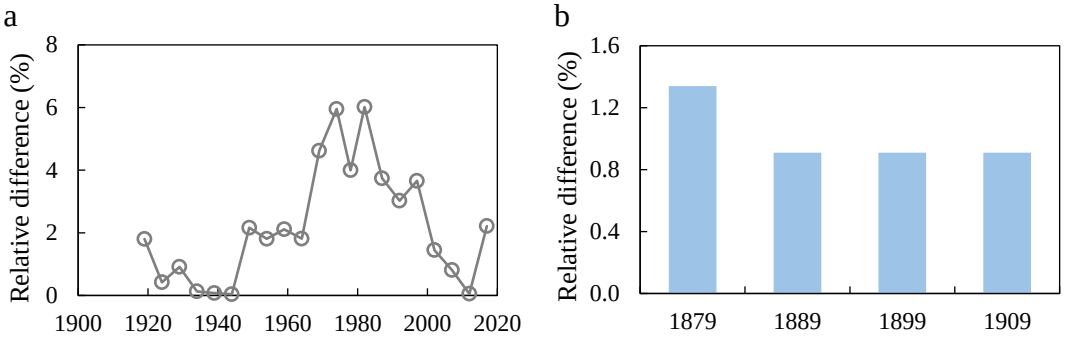

Figure S9: Relative differences in cropland area between the newly developed cropland data and the CAHA cropland harvested area data during 1919-2017 (a) and 1879-1909 (b).

For the period before 1879, we integrated the newly developed cropland data (1879-2020) and HYDE3.2 cropland data (1630-1879) to reconstruct the historical cropland area. However, the cropland definitions of this study and HYDE3.2 cropland data are different (Table S5), resulting in the uncertainties to the reconstruction.

Considering the cropland area and definition differences, we used the HYDE3.2 cropland per capita change rate rather than the absolute value of cropland area between 1630 and 1879. To quantify the uncertainties, we calculated the relative difference in cropland per capita change rate ($HistCrop_{p,t1}/HistCrop_{p,t2}$) between the newly developed cropland data and HYDE3.2 cropland data in the overlap period (1880-2017), which can be expressed as follows:

$$RD_{crop\_p,t} = \left|\frac{HistCrop\_CR_{p,t} - HYDE\_crop\_CR_{p,t}}{HistCrop\_CR_{p,t}}\right| \times 100\% \quad \text{(S3)}$$

Where $RD_{crop\_p,t}$ refers the relative difference of cropland per capita change rate between the newly developed cropland data and HYDE3.2 cropland data; $HistCrop\_CR_{p,t}$ and $HYDE\_crop\_CR_{p,t}$ are the cropland per capita change rate derived from the newly developed cropland data and HYDE3.2 cropland data in year *t*, respectively.

Compared with HYDE3.2 cropland data, the newly developed cropland data showed significant interannual variations during 1880-2017 (Figure S10a). The mean relative difference in cropland per capita change rate between HYDE3.2 cropland data and the newly developed cropland data during 1880-2017 is 2.10 ± 2.82%, and the relative difference values in most of the years were lower than 5% (Figure S10b).

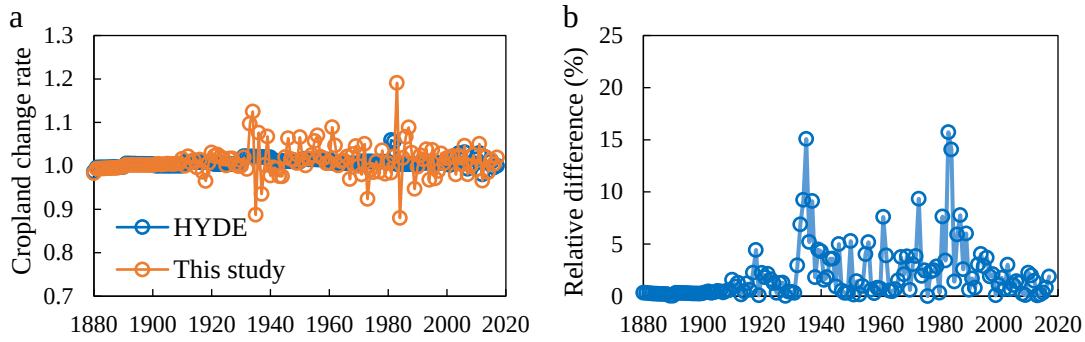

Figure S10. Cropland per capita relative difference between the newly developed data and HYDE3.2 cropland data during 1880-2017.

We further calculated the relative difference of state-level cropland per capita change between HYDE3.2 cropland data and the newly developed cropland data during 2001-2017 (Figure S11). The results showed that the two datasets matched well with the mean relative difference value of 1.21±1.45% (2001-2017). And the relative difference values in most states are lower than 3%, except Colorado (4.17%), New Mexico (4.83%), and Rhode Island (3.35%) (Figure S11).

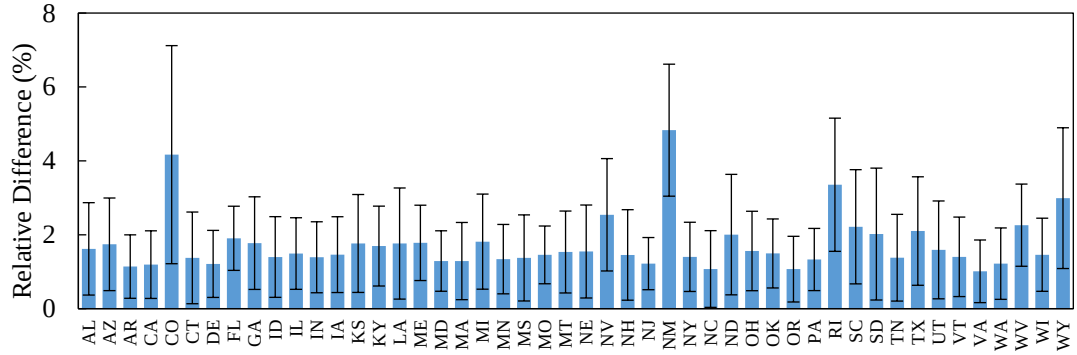

Figure S11. Mean relative difference of state-level cropland per capita change derived from HYDE3.2 cropland and the newly developed cropland land data during 2001-2017.

Pasture

In this study, three datasets (National Resources Inventory (NRI) pasture data, HYDE3.2 pasture, and total population data) were used for the historical pasture reconstruction. For the year before 1982, we integrated the newly developed pasture data (1982-2017) and HYDE3.2 pasture data (1630-1982) to reconstruct the historical pasture area. However, the pasture definitions of this study and HYDE3.2 pasture data are different (Table S6), resulting in uncertainties to the reconstruction.

Considering the pasture area and definition differences, we used the change rate of HYDE3.2 pasture per capita ($HYDE\_Pasture\_p_{t1}/HYDE\_Pasture\_p_{t2}$) rather than the absolute value of pasture area between 1630 and 1982. To quantify the uncertainties, we calculated the relative difference in pasture per capita change rate between our data and HYDE3.2 pasture data, which can be expressed as follows:

$$RD_{pasture\_p,t} = \left|\frac{HistPasture\_CR_{p,t} - HYDE\_Pasture\_CR_{p,t}}{HistPasture\_CR_{p,t}}\right| \times 100\% \quad \text{(S4)}$$

Where $RD_{pasture\_p,t}$ refers the relative difference in pasture per capita change rate between the newly developed pasture data and HYDE3.2 pasture data; $HistPasture\_CR_{p,t}$ and $HYDE\_Pasture\_CR_{p,t}$ are the pasture per capita change rate of the newly developed pasture and HYDE3.2 pasture in year *t*, respectively.

We calculated the pasture per capita change rate from HYDE3.2 pasture data and the newly developed pasture data in the overlap period (1982-2017) (Figure S12). The results showed that mean change rates in pasture per capita were 1.01±0.03 (HYDE3.2) and 1.02±0.02 (This study), respectively. The mean relative difference in pasture per capita change rate between HYDE3.2 pasture data and the newly developed pasture data is 4.89±1.94%.

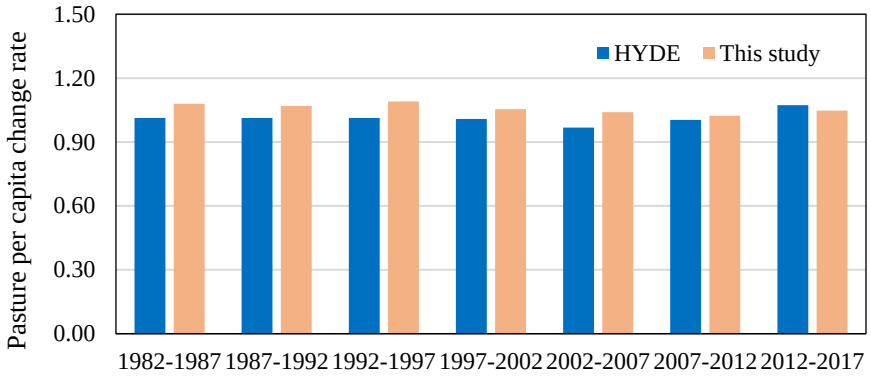

Figure S12: Pasture per capita change rate of HYDE3.2 pasture data and the newly developed pasture data during 1982-2017.

We also calculated the relative difference of state-level pasture per capita change between HYDE3.2 pasture data and the newly developed pasture data during 1982-2017 (Figure S13). The results showed that the mean relative difference of 48 states was 6.51±6.62%, and the relative difference values of all the states were lower than 8% in the seven sub-period.

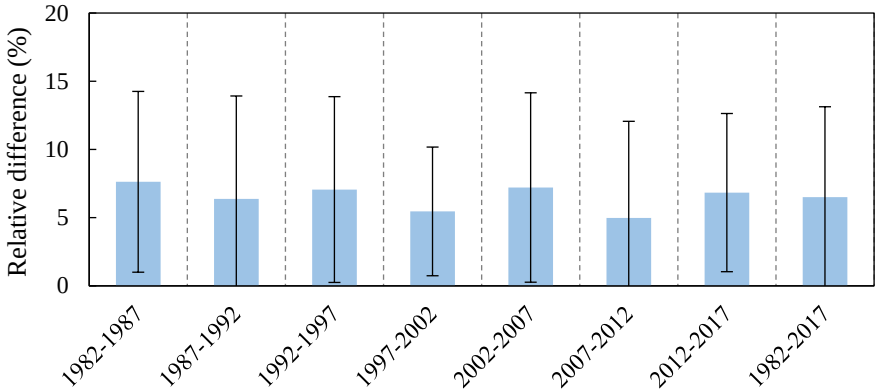

Figure S13: Mean relative difference of state-level pasture per capita change rate between HYDE3.2 pasture data and the newly developed pasture data during 1982-2017.

For the forest, we integrated two datasets (USDA and FATD) to generate the historical forest land during 1630-2020. In the overlap period, the forest area from the two datasets is the same. We didn't calculate uncertainties for the forest area.

Reference:

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, https://doi.org/10.2737/RMRS-GTR-111, 2003.

Klein Goldewijk, 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, https://doi.org/10.5194/essd-9-927-2017, 2017.

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., 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, https://doi.org/10.5194/gmd-13-5425-2020, 2020.

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, https://www.nrcs.usda.gov/nri (last access: Feb 13, 2023), 2020.