Thank you for the comments and suggestions. These comments were very helpful for revising and improving our paper. We have responded to the comments point by point.

Reviewer #1:

General comments:

The paper takes on the very substantial challenge of recreating historical land use for the United States. The authors are correct that outside of coarse-level reconstructions such as HYDE, that there is nothing of higher resolution that goes back to pre-settlement by European colonists. The authors pitch this methodology and dataset as the solution to that data gap. I don't think the authors are completely successful in making that argument. After stating the need for higher-resolution data, and presenting a methodology that should allow for some higher-level of spatial detail than something like HYDE, the authors inexplicably leave out any kind of spatial analysis of how well their model result performs. It's difficult when spatially explicit data on land use is hard to come by (and hence the need for this work!), but some analysis of the full-resolution data compared to a data set such as NLCD (available 2001 to 2019) could have helped establish confidence in the model to capture spatial patterns well. Even a comparison to historical county-level data would help. However, the only "spatial" analysis of the data are some very coarse regional assessments that don't provide a reader much of a feel for the model's capability for generating realistic, high-resolution spatial patterns. Validation overall is a weak point of the paper. On the one hand, I understand the difficulties in trying to "validate" results such as this, when consistent reference data is absent or scattered. However, it's not acceptable from a modeling perspective to use HYDE, NLCD, and other data to parameterize the model, and also use those data in what's labeled as "validation" of results. Given the lack of spatially explicit reference data, I have no issue with the authors doing "consistency checks" with other data sets, including missed opportunities such as county-level ag census data that could have been used to provide a better feel for the spatial patterns produced by the model. But don't try to sell it as a real "validation" of model results.

Overall, the authors continually note the "Uncertainties" associated with coarser data such as HYDE, but fail to conclusively demonstrate their results are superior in terms of those uncertainties. A major need for the paper is recognition of the differences between source datasets, and the uncertainties it introduces into the modeling. For example, many parts of the model are parameterized by using multiple datasets that have inherent differences. Pasture, for example, is quantified at a state level by NRI, and then HYDE for older dates. Given how variable definitions are for “pasture” in the first place, it’s asking for trouble to mix and match datasets such as that. The authors do seem to make some attempts to harmonize differences in datasets, but the methodology isn’t well-defined enough to let me know if that’s really being done. In short, historical land use reconstruction is difficult because 5 different datasets may give you 5 different answers for how much “forest”, “urban”, “cropland”, or “pasture” is actually there for a given date! The spatial allocation of “change” on the landscape is also quite simplistic, based solely on probability surfaces with no stochasticity. The authors do attempt to mitigate the ‘static probability surface’ problem present for past applications such as CLUE and FORE-SCE by some simple weighting with population. But the actual allocation is fully dependent on the probability surfaces at the end of the day, and as a result, as you go back in time, you tend to see classes such as cropland and pasture become concentrated in the high-probability locations, with less fragmentation that’s there in later dates.

Finally, note I did download and look at some of the output results. For brevity I’ll keep my comments here to the “Boolean” land cover. Note that while results look reasonable at broad scales, the approach of parameterizing state-by-state does seem to cause some issues as you go back in time, as does the spatial allocation methodology. Going back in time reveals a number of obvious state and even what appear to be county boundaries, hard obvious lines where land use clearly differs on either side of a political boundary. Given the complete reliance on probability surfaces alone for the spatial allocation of change, land use looks more concentrated on the landscape for some classes going back in time. For example, on the 1850 map, cropland is concentrated in very large contiguous chunks in many areas, and very sharp and obvious political boundaries are present.

Response: Thank you for the comments and suggestions. According to your suggestions, we conducted another two data comparisons to increase the newly developed LULC data’s confidence in section 3.1.

First, we did a comparison between our reconstruction with USDA county-level agricultural data for cropland in 1850, 1920, 1960, and 2002.

An accurate cropland distribution map is quite critical for historical LULC reconstruction. To prove the reliability of the newly developed cropland, we compared it with county-level census data. The spatial pattern of cropland proportion (cropland area/county area) from this study is close to the census data in the four years. Both two datasets show the cropland expansion in the North Central, the Great Plains, the Mississippi Alluvial Plain, and California. The cropland abandonment can also be found in the Appalachian Mountains. In the early period (1850), our results and census data show the cropland distribution in the central area of Alabama and Georgia, New England, and the North Central. But our reconstruction overestimated the cropland fraction in the North Central states, the east area of Virginia, North Carolina, South Carolina, and the south of Georgia. Because the census data only recorded the area of 18 major crops, our data is higher than the census data in the county with high cropland area. For example, cropland derived from this study was higher than the census data in the southeast coast, Atlantic coast, the Mississippi Alluvial Plain, northwest area of Texas, west of Oklahoma, and California in 1920, 1959, and 2002. While the cropland proportion in the Appalachian Mountains and the south area of the Great Plains was lower than the census data. This underestimation may result from the low cropland fraction in satellite data because it is hard for satellite data to extract the small area cropland patch in the mountain area and classify the pasture/grassland with cropland in the south of the Great Plains.

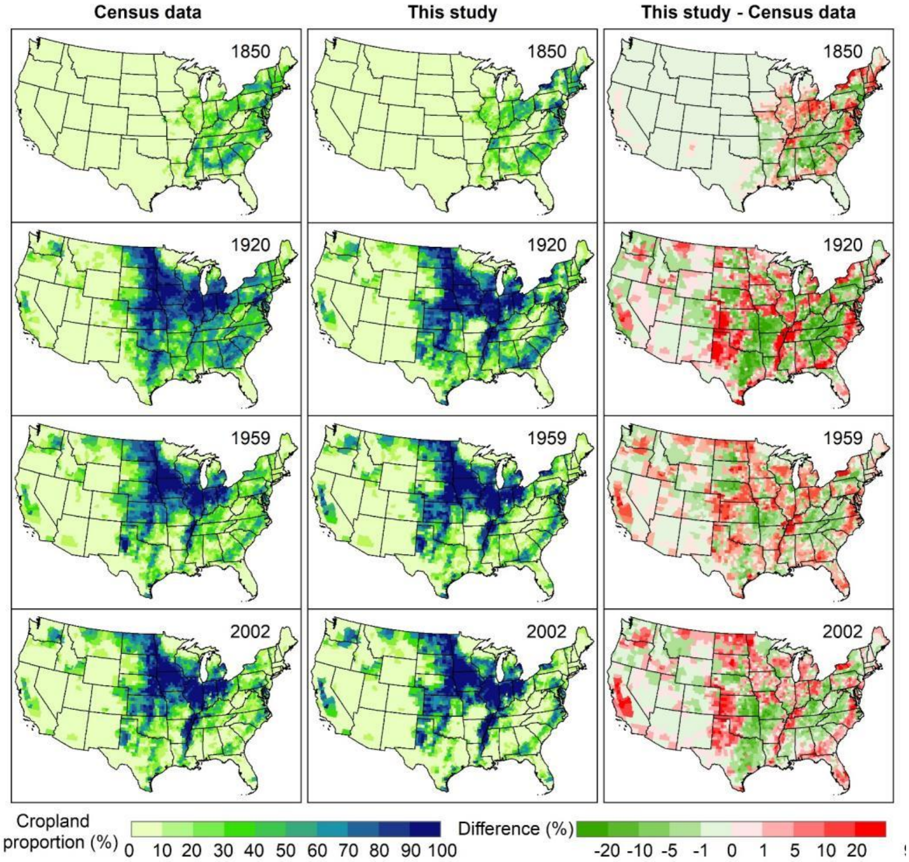

Figure R1. Spatial comparison of county-level cropland proportion between our reconstruction and census data in 1850, 1920, 1959, and 2002. First column: cropland proportion from census data; Second column: cropland proportion derived from this study; Third column: cropland proportion between this study and census data.

We also compared our data with NLCD data at the grid level and calculate the difference. The spatial patterns of urban, cropland, pasture, and forest in this study are close to the satellite-based LULC data from NLCD, and most grids have a relatively small difference. For urban land, our result overestimates the fraction in the low urban density area, but about 90% of urban land grids whose differences range from -10% to 10%. For cropland, the area with a positive difference is mainly distributed in the Northeast states, Alabama, and Missouri, in which 67% of grids have small differences values lower than 10%. 39% of grids have negative difference values and are mainly located in the states with high cropland area. Moreover, most states in our reconstruction have a lower pasture fraction than NLCD data except for Oklahoma, Arkansas, Texas, and Georgia. The grids with negative differences account for 49%. For forest, the reconstructed forest has a higher density than NLCD in the South, Pacific coast, and Great Lakes but underestimates the forest fraction in the central states, such as Missouri, Kentucky, and Ohio. There are 73% grids whose differences are relatively small and range from -10% to 20%.

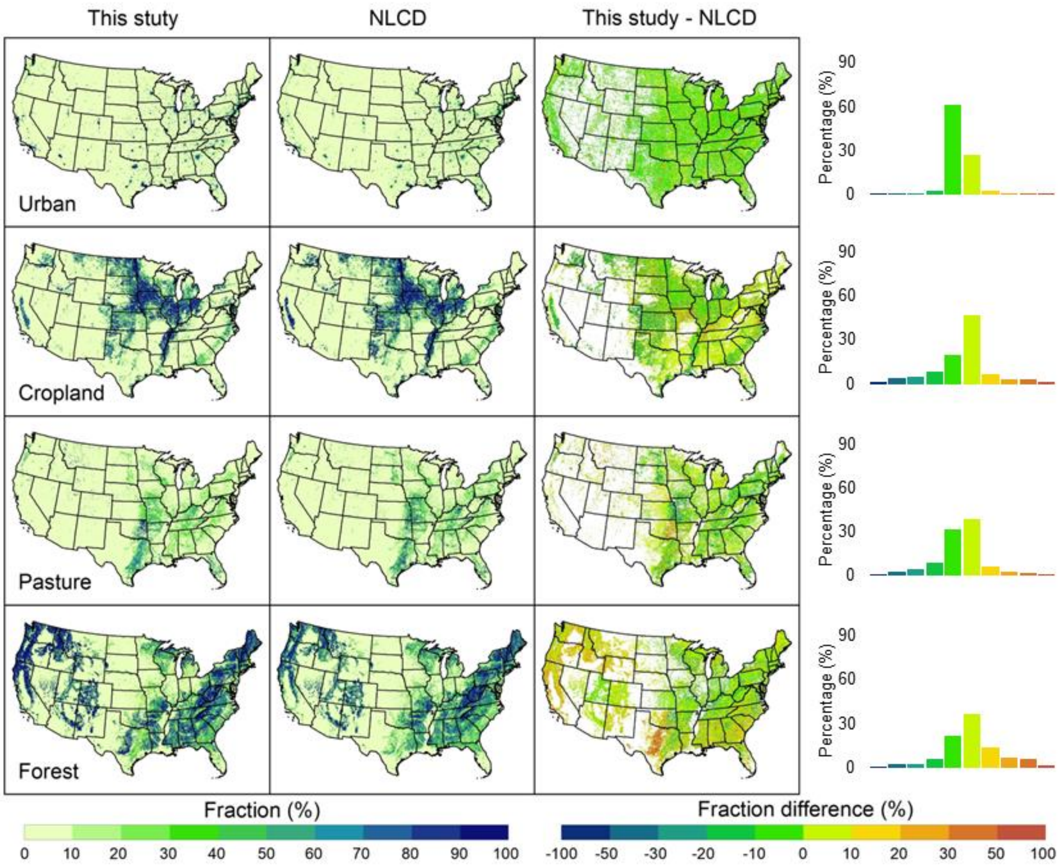

Figure R2: Spatial comparison between our reconstruction and satellite-based urban, cropland, pasture, and forest. First column: Reconstructed data in this study (average between 2001 and 2019); Second column: Satellite-based data (average between 2001 and 2019); Third column: Difference between first column and second column; Fourth column: Distributions of fraction difference between our reconstructed database and satellite-based data.

As you said, the definitional differences did increase the uncertainties of LULC modeling because we had to harmonize multiple datasets for each LULC type. Thus, we summarize and compare the definitions for cropland, pasture, and forest from the multisource dataset, please see Table R1, R2, R3. Moreover, to make the reader understand the urban, cropland, pasture, and forest reconstruction process easily, we revised the method described in section 2.2.

Table R1: Definition of cropland in different data sources.

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).
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 (USDA, 2020).
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).
HYDE	FAO categories of "arable land and permanent crops" (Klein Goldewijk et al., 2017).
This study	Same as USDA-ERS, but we only count the cropland harvested area.

Table R2: Definition of grazing land, pasture, and rangeland in different data sources.

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.
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 (USDA, 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 (USDA, 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.

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.
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. ***Pasture*****:** Pasture is high-intensity grazing land, or low intensity grazing lands where a conversion of the natural vegetation has occurred. ***Rangeland*****:** rangeland is low-intensity grazing land where the natural vegetation has not been converted.
This study	Same as NRI.

Table R3. Forest definitions from different data sources.

Data source	**Definitions**
Forest Inventory Analysis (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://www.fia.fs.fed.us/tools-data/maps/2007/descr/yfor_land.php).
National Land Cover Database (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).
Land Use Harmonization 2 (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).
This study	Same as Forest Inventory Analysis (FIA).

For the spatial allocation, we agree with you that the actual allocation is fully dependent on the probability surfaces and the landscape pattern at the end of the day. CLUE model and FORE-SCE model generate a LULC map at the predicted year by allocating the LULC demand (LULC area net change) to a LULC base map. This method works well for short-period studies because they can assume that the large-scale LULC pattern is not changed. To be honest, we also tried such a spatial allocation method for generating fractional and Boolean-type data. However, the contemporary LULC pattern is not representative for the historical LULC pattern even going back to the 1940s in CONUS (Sohl et al., 2016). Thus, we have to add some modifiers (such as population density) to improve the LULC probability. We know it is simple by adding a population weight, but it is effective. Because the distribution of human-related LULC types (e.g., urban, crop, and pasture) was always correlated with population density in the early period. In Figure R3, we can see that the county-level population and cropland proportion show the same spatial patterns in 1850.

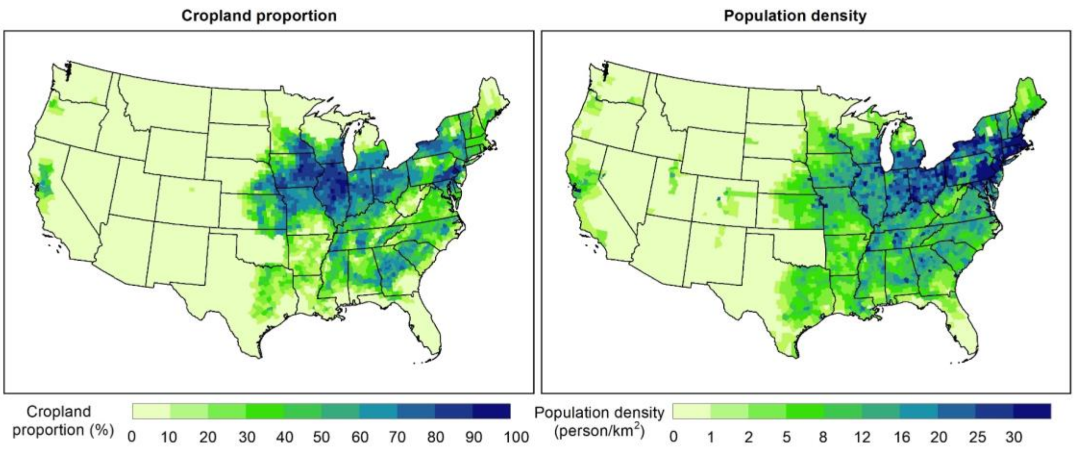

Figure R3: County-level population density and cropland proportion in 1850.

We add a detailed description for fractional type data and Boolean type data allocation methods and a workflow (Figure R4). To generate the fractional grid data, we assumed that the fraction of each LULC type at the grid level was determined by the total probability, which means that a grid cell (LULC class *k*) with a high probability will have a high fraction. Based on this principle and the state-level LULC area, we generated the fractional land use data at 1 km x 1 km resolution and annual time scale. The detailed information for generating fractional LULC data is shown in Figure R4 and the following steps: (1) prepare the input data: reconstructed LULC area and probability; (2) calculate the state target LULC fraction for class *k* and initialize an empty LULC fraction surface; (3) calculate a temporal fraction surface; (4) modify the temporal fraction surface, we assume that the fraction of water and barren is stable, and the sum of urban, crop, pasture, and forest fraction is lower than the maximum fraction in each grid cell; (4) add the temporal fraction data to the empty LULC fraction; (5) judge whether the unallocated LULC area is smaller than 0.01 km^2, if yes, the irritation will be stopped and begin to allocate another LULC class, else the unallocated area will assigned to target fraction and return back to step 3, the allocation was processed iteratively until unallocated area was less than the threshold (0.01 km^2) in all states.

Based on the fractional type LULC data, we further generated the Boolean type LULC data at 1 km x 1 km resolution. The detailed information for generating fractional LULC data is shown in Figure R4 and as following steps: (1) prepare the input data: reconstructed LULC area and fractional data; (2) generate a temporal LULC map (HistB) through identifying the dominate LULC type in each grid cell and initialize an empty LULC map ($HisB_E$); (3) calculate the area difference for class *k* between the HistB map and target area; (4) if the area difference is negative, we first sort the LULC fraction map where HisB map equal to *k*, the top m (equals to target area) grid cells where $HisB_E$ not be assigned a value will be assigned as *k*; then if the available number of grid cell (class *k*) is smaller than the target area, we will sort the LULC fraction map where HisB map not equal to *k*, and the top n (equals to unallocated area) grid cells where $HisB_E$ not be assigned a value will be assigned as *k*; (5) if the area difference is positive, the grid cells where HisB map equal to *k* and will be assigned *k* to $HisB_E$ not be assigned a value; then we will sort the LULC fraction map where HisB map not equal to *k*, and the top n (equals to unallocated area) grid cells where $HisB_E$ not be assigned a value will be assigned as *k*. If step (4) or step 5 is finished, the next LULC type will begin to allocate. After the four types allocations are finished, the grid cell not being assigned a value will be updated using the HistB map and LANDFIRE Biophysical Settings data.

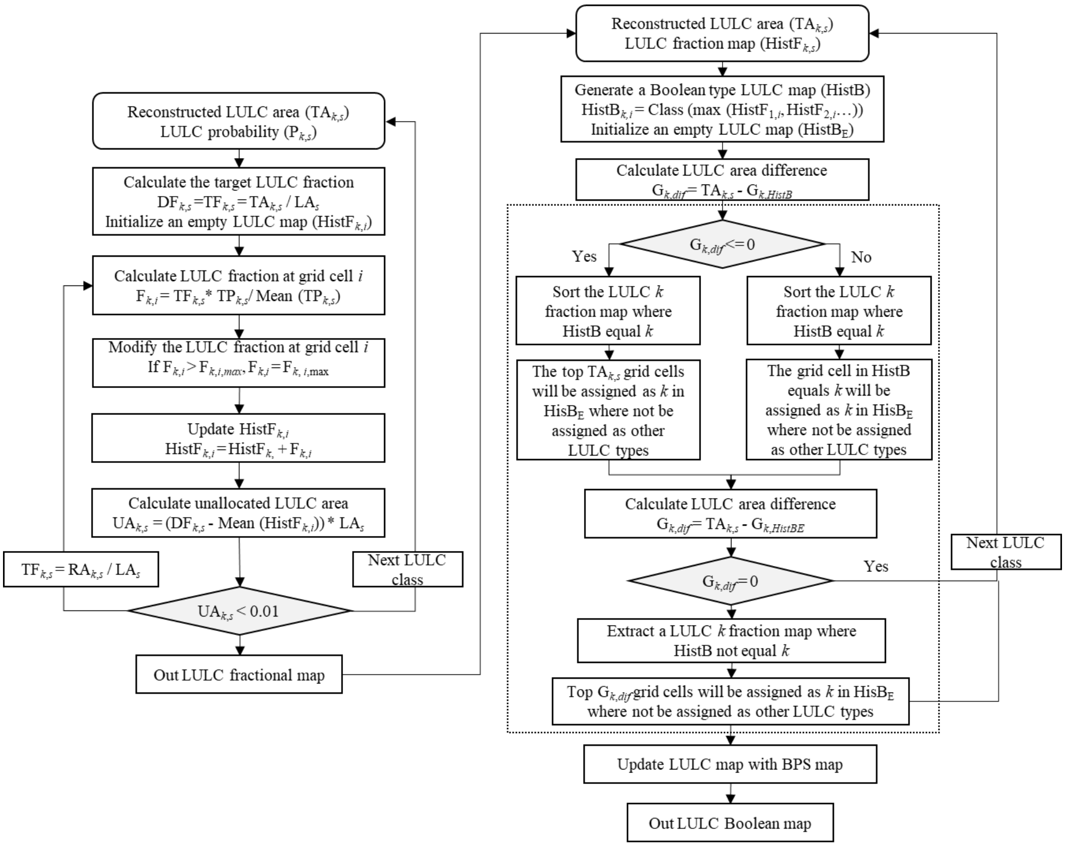

Figure R4: Workflow for generating fractional and Boolean type LULC data

(4) For the Boolean type data, you proposed several suggestions as follows:

- *Cropland and pasture become concentrated in the high-probability locations, with less fragmentation that's there in later dates*.
- *Going back in time reveals a number of obvious state and even what appear to be county boundaries, hard obvious lines where land use clearly differs on either side of a political boundary.*
- *Given the complete reliance on probability surfaces alone for the spatial allocation of change, land use looks more concentrated on the landscape for some classes going back in time.*

We re-check the Boolean type LULC data and analyze the reasons for these problems. The 'political boundary' issue resulted from the following two aspects. The first is that we conduct the spatial allocation at the state level. If the area of one LULC type has a large difference between two neighboring states, there would be a 'political boundary'. The second reason is that we use county-level population density (one value in a county) to modify the LULC probability in the early period, which will result in the hardlines between neighborhood counties. The LULC '*concentration*' problems also resulted from that we apply the population density data to modify the LULC probability surface. In our revised version, we optimized the population density weight and Boolean type spatial allocation method. A detailed description of the method can be found in section 2.4.1 and 2.4.2. Figure R5 shows the comparison of LULC at the local scale in 1850.

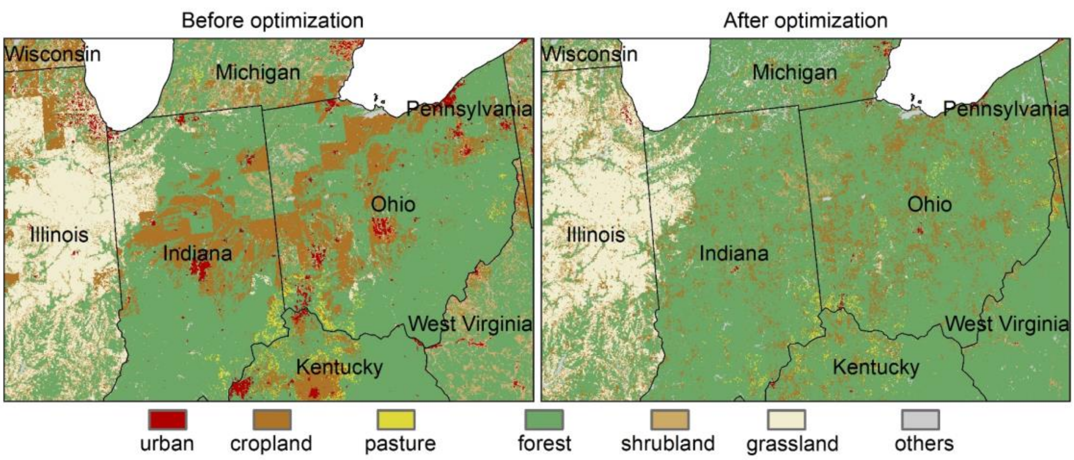

Figure R5: Comparison of Boolean type LULC map in 1850 between before and after optimization.

Specific comments:

Comment 1: Lines 28 – Would add one word…" In particular, managing agriculture and forest-related activities…"

Response: Thank you for this suggestion. We revise those sentences.

Comment 2: Line 34 – Would add words…"…arrival of Europeans, indigenous communities practiced agriculture and crop planting in the…"

Response: Thank you for this suggestion. We revise those sentences.

Comment 3: Line 36 – Would change "mainly occurred" to "initially occurred" to indicate these activities first started here, but expanded elsewhere later (as noted by the next sentences)

Response: Thank you for this suggestion. We revise those sentences.

Comment 4: Line 37 – "Driven" not "driving".

Response: Thank you for this suggestion. We revise those sentences.

Comment 5: Page 2 – Overall the paragraph at the top of page 2 could use some work. It's a rather disjointed history of US land change. For one it doesn't really talk about land change west of the Mississippi River, it's focused solely on Eastern US change. The organization is also a bit odd and disjointed.

The sentence on line 44, for example, seems like a very abrupt and odd ending to the final statement as to why a long-term land use dataset is needed. Perhaps a better organization by period (colonial, 19th century, 20th century), with a description of what occurred in each century? And perhaps a modification of the last sentence, adding "While general trends in historical US landscape change are known, we still lack a long-term dataset…"

Response: Thank you for this suggestion. We rewrite this paragraph. Please see Lines 32-49.

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, indigenous agriculture and crop planting existed in the eastern woodlands, the Great

Plains, and the southwestern US (Hurt, 2002). Since the establishment of the first colony in Virginia in 1607, farms began to expand by land clearing, which initially occurred in the eastern United States (Steyaert and Knox, 2008). During the colonial era, most people lived in the east of the Appalachian Mountains and agriculture was the primary livelihood. In the late 18th and 19th centuries, territorial expansion (e.g., Louisiana Purchase) open up new areas for agriculture. Driven by westward movement, land clearing, agricultural land reclamation, and deforestation expanded across the Appalachian Mountains into Ohio, the Mississippi River basin, the Great Lakes region, and the western US (Cole et al., 1998; Billington et al., 2001; Steyaert and Knox, 2008; Yu and Lu, 2018). Hardwood forests in the Mississippi River Valley were cleared for cotton and grain production (Hanberry et al., 2012). The lumber production center also shifted from the Northwest to Great Lakes states in the 1850s (Fickle et al., 2001). In California, agriculture and ranching expanded throughout the state and soon became an exporter of wheat as gold mining waned (Olmstead et al., 2017). Entered the 20th century, the agricultural intensification resulted in cropland abandonment in New England, the Atlantic coast, and the southeast US (Foster, 1992; Hall et al., 2002; Jeon et al., 2014; Zumkehr and Campbell, 2013). Meanwhile, the environmental protection movement accelerated forest restoration, and the national plantation forest area increased to 27 Mha in 2017 (Oswalt et al., 2014; Stanturf et al., 2014; Chen et al., 2017). While general trends in historical US landscape change are known, we still lack a long-term spatial-explicit LULC dataset to characterize LULC trajectories for the CONUS.

Comment 6: Line 55-56 – Be careful about highlighting “uncertainties” in datasets such as HYDE, as your historical landscape construction will also have substantial uncertainties. Your workflow itself uses HYDE data. There’s limited spatially explicit data available from which to base a model-based landscape reconstruction, and many of the datasets you’re using were also used by HYDE.

Response: Thank you for this suggestion. We agree with you that the uncertainties of HYDE should not be highlighted because HYDE data is also one of the input data in our reconstruction. The historical LULC pattern in the CONUS has changed a lot compared with the contemporary pattern. This is also the reason that we used the population density and human settlement extent to improve the LULC probability, even though it is a simple method.

Comment 7: Lines 87-88 – I wouldn't call the use of these other datasets "validation". It's a consistency check, not a validation, as these data sets too have uncertainties, and some are modeled just as you're modeling.

Response: Thank you for this suggestion. These data should be the input for consistency check rather than validation. We revise the description about the data comparison or consistency check. Please see section 2.4.

Comment 8: Line 90-91 – How was resampling done to get to 1-km grid cells? Is it fractional LULC within a given 1-km cell for datasets with native resolution <1 km?

Response: Thank you for this suggestion. The input data were resampled (nearest method) or aggregated to 1 km resolution. If the dataset with native resolution < 1 km, it will be fractional LULC within a given 1-km cell.

For the NLCD dataset, we generated the fractional type data to 1 km for each LULC type (urban, cropland, pasture, and forest) using the "aggregate" method.

Comment 9: Section 2.2.1 – This is an extremely simplistic methodology for calculating urban land area. To start, it's all based on one current dataset, NLCD. How was NLCD used? First of all, NLCD tends to underestimate low-density residential lands, which can bias your results. Secondly, NLCD "urban" classes also include extensive representation of road networks, which if counted as "urban", greatly overestimates urban land. For a rural state, for example, NLCD classes not only major roads, but every small section road has a 1-pixel-wide "urban" class representing it. Unless measures were taken to account for NLCD's underrepresentation of low-density residential lands, and to account for all the "urban" pixels that are really roads, it biases the results. The other problem is the very simplistic method for calculating land area. You're assuming the relationship between urban land per capita and total population is constant through time. Clearly it's not. Without accounting for that changing relationship, urban estimates can easily be biased.

Response: Thank you for this suggestion. In this study, we count the four components (Developed, Open Space; Developed, Low Density; Developed, Medium Density;

Developed, High Density) of NLCD developed land as urban land. And the developed land area between 2001 and 2019 is regarded as baseline data for historical urban land area reconstruction. Though NLCD has some shortcomings, we need to choose a dataset with a clear definition and spatial explicit map, which will be helpful for historical urban land reconstruction.

In the current method, we used the population and a stable urban land per capita to estimate the long-term urban area at the state level. We agree that it is a very simplistic method and also pointed out that it overestimates the total urban land area in the early period. To solve this problem, we apply the HISDAC data to reduce the bias between 1810 and 2000, and the method has been updated in *section 2.2.1*. Before 1810, there is no available data that can be used, and we assumed that urban land has the same change rate as the total population at the state level.

Section 2.1.1

In this study, we use the same definition as developed land in NLCD for urban land. Developed land in NLCD includes four components: open space, low intensity developed land; medium intensity developed land, and high intensity developed land. For 2001–2019, the NLCD developed land area was regarded as the urban land baseline. Before 2001, we applied the historical built-up area (HBUA) data from Historical Settlement Data Compilation for the United States (HISDAC-US) (Leyk et al., 2020; Uhl et al., 2021) as input to reconstruct the historical urban land area. The HISDAC-US describes the built environment of most of the CONUS from 1810 to 2010 at 5-year temporal and 250 m spatial resolution by using built-up property records, built-up property locations, and the built-up intensity (Leyk and Uhl, 2018; Uhl et al., 2021). Here, we assumed that the HISDAC data can capture the changing trend of urban land development. Then, the historical urban land can be estimated as follows:

$$HistUrban_{s,t} = HistUrban_{s,t+1} \times \frac{HISDAC_{s,t}}{HISDAC_{s,t+1}} \quad (1)$$

where $HistUrban_{s,t}$ and $HistUrban_{s,t+1}$ are the reconstructed urban land area of state *s* in year t and t+1; $HISDAC_{s,t}$ and $HISDAC_{s,t+1}$ are the HISDAC built-up area of state *s* at year *t* and t+1. The step was first conducted in 2000 and went back to 1810.

There is no reliable data on urban development before 1810. Following Liu et al. (2010), we used population to estimate the urban land area by assuming that urban land expanded at the same rate as total population during 1630–1810. The urban land area of each state can be calculated as follows:

$$HistUrban_{s,t} = HistUrban_{s,t+1} \times \frac{Pop_{s,t}}{Pop_{s,t+1}} \quad (2)$$

where $HistUrban_{s,t}$ and $HistUrban_{s,t+1}$ are the reconstructed urban land area of state s in year t and t+1; $Pop_{s,t}$ and $Pop_{s,t+1}$ are the total population of state s at year t and t+1.

Comment 10: Section 2.2.2 – You're using (at least) three different data sources to help establish cropland area. For historical land use, estimates vary widely, dependent upon methodology, data source, thematic definitions of a land use, etc. As a result, when switching from USDA-based data, for example, after 1889, and using HYDE before 1889, you'd expect an obvious break in estimated "cropland" amounts. How were those inconsistencies among historical land use datasets harmonized?

Response: Thank you for this suggestion. In the current method, we first convert cropland harvest area to planted area by assuming a stable linear relationship between these two datasets, then we calculated the cropland area by subtracting the double-cropped area from the planted area. However, these two steps will result in some uncertainties as you said in Comment 11 and 12.

To reduce these uncertainties, we decided to change the input data for historical cropland reconstruction, including the ERS national cropland harvested area (without double-cropped area), the state level CAHA-cropland harvested area (1879-2017) and HYDE cropland (1600-1879) to reconstruct the historical cropland area. Please see the new method described in section 2.2.2.

The definition of cropland varies in the existing literature and datasets (Bigelow and Borchers, 2017; Yu and Lu, 2018; Homer et al., 2020). Cropland defined by the U.S. Department of Agriculture (USDA) Economic Research Service (ERS) includes five components: cropland harvested, crop failure, cultivated summer fallow, cropland pasture, and idle cropland (Table 2). In this study, we only counted the cropland harvested area, which includes row crops and closely sown crops, hay and silage crops, tree fruits, small fruits, berries, and tree nuts, vegetables and melons, and miscellaneous other minor crops (https://www.ers.usda.gov/data-products/major-land-uses/glossary/#cropland). The USDA Census of Agriculture Historical Archive

(CAHA) provides state-level cropland harvested area with 4 or 5 years interval, which was applied for historical cropland reconstruction between 1879-2020. The dataset was first interpolated linearly to annual. To subtract the double-cropped area in the dataset, we applied the national cropland harvested area from ERS major land uses data to adjust the interpolated data during 1910-2020. The adjustment can be expressed as follows:

$$HistCrop_{s,t} = \frac{Cropland\ Harvested_{s,t}^{linear}}{Cropland\ Harvested_{conus,t}^{linear}} \times Cropland\ Harvested_{conus,t}^{ERS} \quad (3)$$

where $HistCrop_{s,t}$ is the reconstructed historical cropland area of state s in year t; $Cropland\ Harvested_{s,t}^{linear}$ is the linear interpolated cropland harvested area of state s in year t based on CAHA data; $Cropland\ Harvested_{conus,t}^{ERS}$ is the national total cropland harvested area without double-cropped area. For 2018–2020, the state level cropland area was calculated based on the state level area weight in 2017.

For 1879–1910, there is no national level statistics for cropland harvested area without double-cropped area. We applied the changing trend the CAHA cropland harvested area to reconstruct the historical cropland in this period, which can be calculated as follows:

$$HistCrop_{s,t} = HistCrop_{s,t+1} \times \frac{CAHA\_CHA_{s,t}}{CAHA\_CHA_{s,t+1}} \quad (4)$$

where $HistCrop_{s,t}$ and $HistCrop_{s,t+1}$ are the reconstructed historical cropland area of state s in year t and t+1; $CAHA\_CHA_{s,t}$ and $CAHA\_CHA_{s,t+1}$ are improved farmland area of state s in year t and t+1.

Before 1879, there is no available census data, we used the HYDE cropland dataset to reconstruct historical cropland area. The harmonization method can be expressed as follows:

$$HistCrop_{s,t} = HistCrop_{s,t+1} \times \frac{HYDE\_Crop_{s,t}}{HYDE\_Crop_{s,t+1}} \quad (5)$$

where $HistCrop_{s,t}$ and $HistCrop_{s,t+1}$ are the reconstructed historical cropland area of state s in year t and t+1; $HYDE\_Crop_{s,t}$ and $HYDE\_Crop_{s,t+1}$ are HYDE cropland area of state s in year t and t+1.

Comment 11: Paragraph starting on line 120 – You're assuming the relationship between harvested area and planted area from 1978 to 2017 is consistent decades and centuries before those data…a very dangerous assumption.

Response: Thank you for this suggestion. We agree that the relationship between cropland harvested area and the planted area is changed. Thus, we revise the reconstruction method. This step doesn't need to conduct in the new method. Please see the new method description in the response for Comment 10.

Comment 12: Line 123 – Yet another dataset, Borchers et al. 2014, was used to establish double-cropping at a regional level. Again, consistency among most of these datasets isn't great.

Response: Thank you for this suggestion. To reduce such kind of uncertainties, we revise the reconstruction method, and the subtracting work doesn't need to conduct. Please see the new method described in the response to Comment 10.

Comment 13: Line 124-125 – Another basic assumption that likely isn't true through time.

Response: Thank you for this suggestion. We revise the cropland reconstruction method, and we used the changing trend of HYDE cropland rather than national cropland per capita. Please see the new method described in the response to Comment 10.

Comment 14: Line 125-126 – Because the 1879 number was different you assumed it was incorrect? But data >1889 were "correct"? Was the reconstructed cropland area in 1879 substantially lower or higher than 1889?

Response: Thank you for this suggestion. On the one hand, there was no record for South Dakota in 1879. On the other hand, we assumed that there should not be a large difference between the cropland harvest area between 1879 and 1889. However, the cropland harvested area in South Dakota, Nebraska, and Kansas did have large differences in 1879 and 1889 (Figure R6). Thus, we thought the cropland harvested area in 1879 is not correct before.

We rechecked the changes in cropland harvested area and total population (Figure R7) between 1879 and 1889, we think the cropland harvested area recorded by CAHA should be right and can be used for cropland reconstruction. So, we used the CAHA data in the revised version.

The reconstructed cropland area in 1879 was not substantially lower or higher than 1889 (Figure R8).

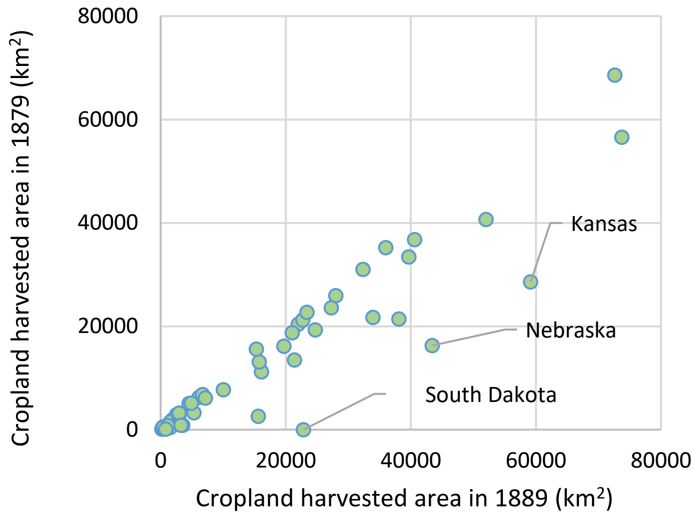

Figure R6: Comparison between cropland harvested area in 1889 and 1879.

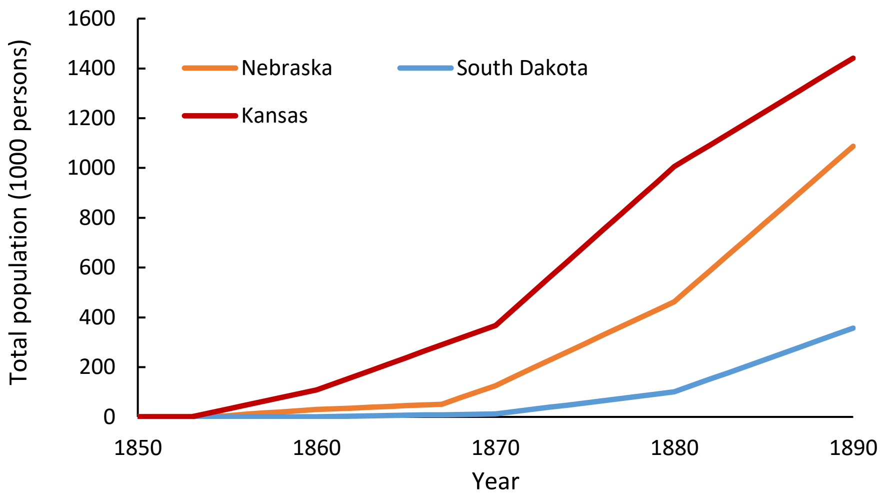

Figure R7: Changes of total population in Nebraska, Kansas, and South Dakota between 1850 and 1890.

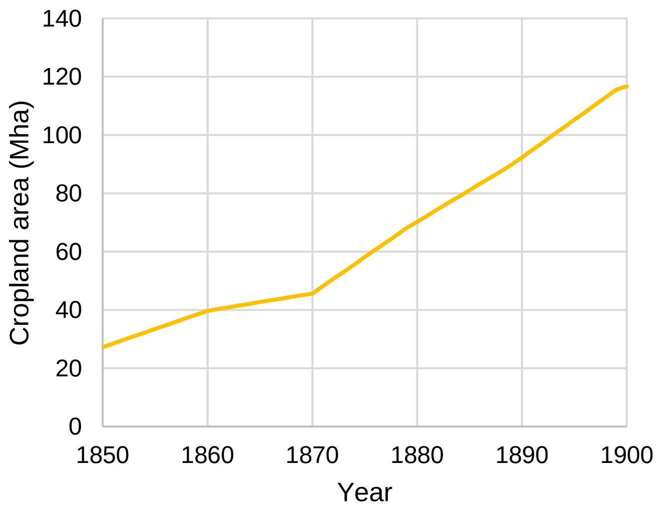

Figure R8: Changes of national total cropland area during 1850-1910.

Comment 15: Line 130-131 – Another basic assumption that likely doesn't hold region to region.

Response: Thank you for this suggestion. We change the harmonization method for integrating HYDE and reconstructed historical cropland. Please see the new method described in the response to Comment 10.

Comment 16: Line 132-133 – Again…how did you account for differences in the HYDE data, and the (mostly) USDA-based data after 1889? Is there an obvious break in cropland amount pre- and post-1889?

Response: Thank you for this suggestion. For the historical cropland reconstruction, we use the HYDE cropland area trends rather than absolute value. The harmonization method between HYDE and CAHA or reconstructed cropland area can be found in the response to Comment 10.

As our response in Comment 14, we found that cropland harvested area increased fast during 1879-1889, and this increase resulted from rapid population increase. The reconstructed data also show that cropland area in South Dakota, Nebraska, Kansa increased rapidly during 1879-1889. But there is no obvious break in cropland amount pre and post-1889 (Figure R8).

Comment 17: Line 135-136 – One of the greatest difficulties in historical landscape reconstruction is the definition of "pasture", vs. "grassland", vs. "hay" vs. "rangeland", etc. There is no one definition that's universally accepted. Your definition here states Pasture includes areas "for the production of seed or hay crops". Many definitions of "cropland" include alfalfa, hay, and other crops in "planted area" or "cultivated crop" area. For "Pasture", you're introducing yet another completely new dataset to establish pasture area, NRI. Are the definitions of "pasture" for NRI the same for NLCD, HYDE, and the US Census of Ag?

Response: Thank you for this suggestion. We summarize the definitions of grazing land, pasture, and rangeland from different sources (Table R2).

As you said, the definitions 'pasture' from NRI is not same as that in HYDE and NLCD or US Census of Ag. Therefore, it is difficult to reconstruct historical pasture area by harmonizing these datasets. In this study, we used the pasture definition from

NRI, and the state level pasture area between 1982 and 2017 was set as baseline data. Before 1982, we applied the changing trend of HYDE pasture data to reconstruct the historical pasture area. The following harmonization method was conducted to calculate historical pasture land area:

$$HistPasture_{s,t} = HistPasture_{s,t+1} \times \frac{HYDE\_Pasture_{s,t}}{HYDE\_Pasture_{s,t+1}} \quad (6)$$

where $HistPasture_{s,t}$ and $HistPasture_{s,t+1}$ are the reconstructed historical pasture areas of state *s* in year t and t+1; $HYDE\_Pasture_{s,t}$ and $HYDE\_Pasture_{s,t+1}$ are the HYDE pasture area of state *s* in year t and t+1.

Comment 18: Line 139 – Note Wasianen and Bliss took great pains to harmonize those definitional differences across their harmonized dataset.

Response: Thank you for this suggestion. We agree that they did make great efforts to harmonize those definitional differences across multisource dataset.

Comment 19: Section 2.2.3 – Again…it's extremely simplistic to assume things such as "pasture per capita" and that that ratio is consistent over time, and space.

Response: Thank you for this suggestion. We revise the historical pasture reconstruction method.

In this study, we used the pasture definition from NRI, and the state level pasture area between 1982 and 2017 was set as baseline data. Before 1982, we applied the changing trend of HYDE pasture to reconstruct the historical pasture area. The following harmonization method was conducted to calculate historical pasture area:

$$HistPasture_{s,t} = HistPasture_{s,t+1} \times \frac{HYDE\_Pasture_{s,t}}{HYDE\_Pasture_{s,t+1}} \quad (6)$$

where $HistPasture_{s,t}$ and $HistPasture_{s,t+1}$ are the reconstructed historical pasture areas of state *s* in year t and t+1; $HYDE\_Pasture_{s,t}$ and $HYDE\_Pasture_{s,t+1}$ are the HYDE pasture area of state *s* in year t and t+1.

Comment 20: Section 2.2.4 – Definitions of what is "Forest" vary greatly among data sets. You're introducing yet another data set in FIA that may have a definition of "forest" that differs from HYDE or from NLCD. How closely does the FATD data match with HYDE estimates, for example?

Response: Thank you for this suggestion. The definitional differences between FIA, NLCD, and LUH2 can be found in Table R3.

The HYDE data doesn't provide forest area estimation. So, the following figure shows the comparison between USDA-FR and NLCD, LUH2 between 2000-2020, between LUH2 and FATD in 1630. Both NLCD and LUH2 forest area are lower than USDA-FR. The forest area in Rocky Mountain states such as Nevada, Utah, New Mexico from NLCD and LUH2 is lower than that from USDA-FR. The forest area in 1630 derived from LUH2 and FATD does not match well.

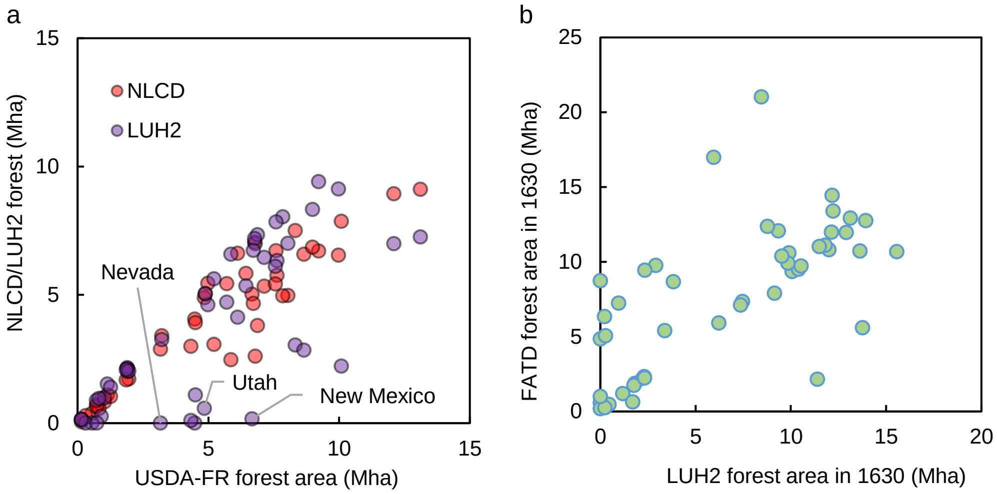

Figure R9: Comparison between the average forest area (2000-2020) derived from USDA-FR and NLCD, LUH2.

Comment 21: Lines 156-157 – Yeah you lost me here with what you're trying to do, needs a better explanation.

Response: Thank you for the suggestions. We rewrite the method to harmonize the FATD and USDA-FR forest data.

Comment 22: Section 2.2.5 – See main comments above related to how you balanced the four LULC classes.

Response: Thank you for this suggestion.

Comment 23: Line 195 – What was used to establish the "land use change boundary"? That is, what was the source of "settled area" data"?

Response: Thank you for this suggestion. We assumed that in the area where human was not settled, there was no urban land, cropland, and pasture. The human settlement boundary data were used to constrain the probability of urban, cropland, and pasture. The Exploration and Settlement maps were made by the U.S. Dept. of the Interior, Geological Survey and can be accessed at https://maps.lib.utexas.edu/maps/histus.html#exploration.html. We assumed that the LULC would not be changed as the pre-colonial era, though there were Native Indiana people.

Comment 24: I don't mind the use of something like this to constrain the allocation of change, but do wonder about full-resolution results. Are there any hard border issues obvious in the data when change occurs at the edge of those defined boundary layers? Overall with the boundary and effect of population density, I appreciate you trying something other than assuming a static probability surface through time.

Response: Thank you for this suggestion. We check the fractional and Boolean type gridded data, there are hard border issues in some years which resulted from the county-level population application to modify the LULC probability surfaces. We use a simple method to generate the historical gridded population by combing the gridded population data with 1-km resolution 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*.

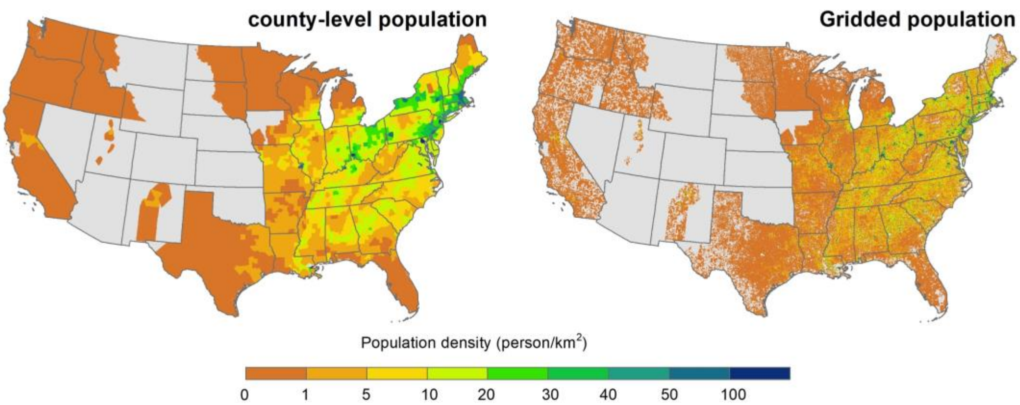

Figure R10: Comparison between county-level and gridded population density data in 1850.

Comment 25: Section 2.3.2 – There needs to be more explanation here. You've basically summarized the entire actual allocation to the pixel level in one sentence. I certainly get that higher probability areas will likely have a higher proportion of a given LULC class, but it's all deterministic and it's all based solely on the probability surface? There's no stochasticity? With such a sparse description of methodology, it's also hard to see how this simple description of the methodology ends up with the aggregate totals from the allocation stage matching the quantitative estimates you established for each of the LULC classes.

Response: Thank you for this suggestion. We rewrite the description of spatial allocation strategies for generating fractional and Boolean type LULC data and make it as detailed as possible (section 2.3.2). We also add a random item when calculating the total transition probabilities (section 2.3.1). Please see the spatial allocation method described in the response to the general comments.

Comment 26: As noted in the main comments, I have other concerns about the allocation strategy.

Response: Thank you for this suggestion. We rewrite the spatial allocation strategy and make it as detailed as possible. Please see the spatial allocation method described in the response to the general comments.

Comment 27: Section 2.4 – Comparison to other LULC datasets isn't a validation, it's a consistency check. That's particularly true when every dataset has it's own production methodologies, data sources, and thematic definitions, all of which makes even direct comparison problematic. Beyond that, you're comparing your results to some of the same datasets from which you parameterized your modeling, as noted in the overarching comments. Also note there aren't any details as to what methodologies you're actually using for "validation" in this very short, one-paragraph section.

Response: Thank you for this suggestion. We agree that the description or this step is a consistency check, rather than a data validation. Complete formal validation of model results was impractical because true, spatially explicit 'reference' data for historical LULC are difficult to obtain. In the historical LULC area reconstruction step, we assumed that the data used is reliable, which made it hard to conduct the validation. If we apply a rule to reconstruct the historical LULC area, we can use the census data to validate the accuracy of the prediction. For example, Sohl et al. (2016) used the LULC change data from the Trends project to reconstruct historical LULC proportions (demand) between 1938 and 1992, and compared the model results with census data, but such comparison still was a consistency check. Moreover, the definitional differences make it difficult to compare the newly developed dataset with other LULC products. We keep the LULC area comparison at the state level. Two new comparisons were conducted: Comparison between the newly developed cropland and USDA historical cropland area at the county level; Comparison between our reconstruction and NLCD developed land, cropland, pasture, and forest. Please see section 3.1.

Comment 28: Section 3.1 – As noted previously, this isn't very useful for inferring confidence in your results, when you're using the same datasets to parameterize the model as you are to "validate" model results.

Response: Thank you for this suggestion. As our response in Comment 27, it is hard to validate the newly developed LULC data. We keep the LULC area comparison at the state level. Two new comparisons were conducted: Comparison between the newly developed cropland and USDA historical cropland area at the county level; Comparison between our reconstruction and NLCD developed land, cropland, pasture, and forest. Please see section 3.1.

Comment 29: Line 273 (and throughout the results section) – If you're going to refer to a specific driving force of change, and, for example, point to a specific policy, you should name the policy and reference it (Immigration and Naturalization Act of 1965). While it certainly did change the nature of immigration to the country, you do give it too much focus as "the" causes of urban land increases after 1965. There's a lot more at play there than immigration policy.

Response: Thank you for this suggestion. We agree that urban land expansion is driven by multiple factors, while it is largely determined by population and economy growth. We rewrite the related sentences about urban land expansion.

Comment 30: Lines 276-277 – You state "cropland area did not change significantly" from 1930 to present day. First, it's always problematic to use the term "significantly" in a journal paper, given the scientific meaning of the word. Secondly, I would argue there were "substantial" trends in agriculture after 1930, including some of those you mention (e.g., biofuel impacts).

Response: Thank you for this specific suggestion. We rewrite the related sentences and this word should be used carefully. What I want to express is that the change magnitude of the national total cropland area is not like the period of 1850-1920 (Figure R11). In fact, cropland was abandoned in the southeast US and expanded in the Great Plains.

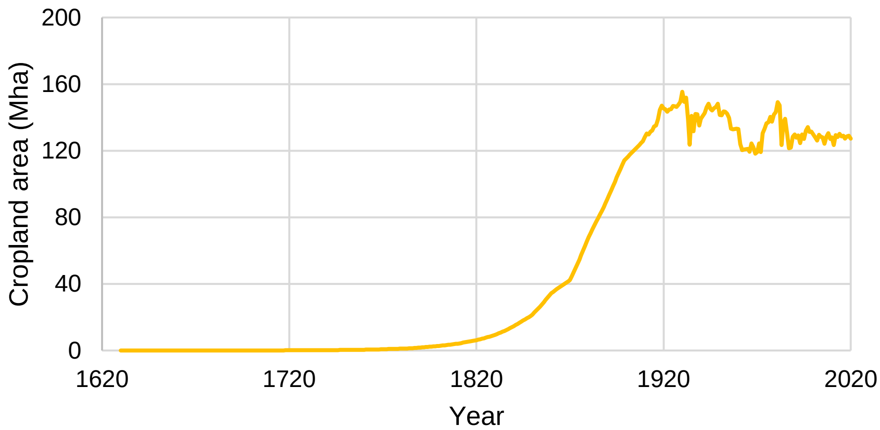

Figure R11: Changes of national total cropland area derived from the newly developed LULC dataset during 1630-2020.

Comment 31: Figure 7 – On a national-scale map figure, it's difficult to see patterns of the individual land use transitions. Perhaps it would be augmented by a complementary confusion matrix of changes or some other tabular data approach that allows you to see (and easily quantify) transition types.

Response: Thank you for this suggestion. We add a LULC conversion table including the information of major LULC conversions during 1630–1850, 1850–1920, 1920–2020, and 1630–2020. Please see Table 3.

Comment 32: Section 3.4 – This isn't the most effective section to me. As noted in the main comments, a major premise of the paper was to provide a "high resolution" historical landscape reconstruction for the US. Much of the "regional" information here is also discussed in the overall results above. I'd have much rather seen some real examples (and preferably validation) of landscape pattern at finer scales, given the focus on higher resolution with this paper.

Response: Thank you for this suggestion. Considering the differences in natural environmental conditions and social-economic development, land use and land cover change showed spatial heterogeneity in the CONUS during 1630–2020. The purpose of section 3.4 is to give a general description of how LULC changes among regions.

To show the improvement of newly developed LULC data, we add several figures to show LULC changes at a finer scale in the discussion section, please see Figure R12, 13, and 14.

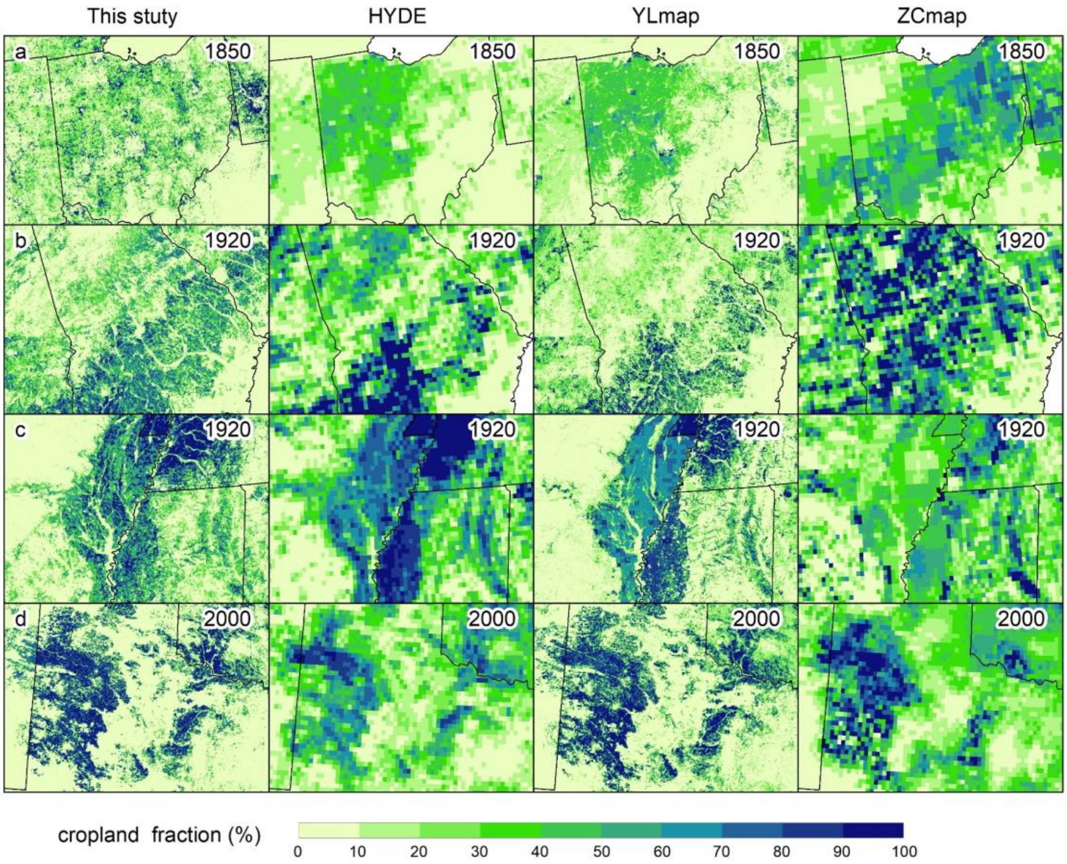

Figure R12: Visual comparison between our cropland and HYDE3.2, YLmap, and ZCmap in four different sites (a-d). The locations of image center points are as follows: a. Ohio in 1850 (83.05 °W, 40.17 °N), b. Georgia in 1920 (83.58 °W, 32.77 °N), c. Arkansas in 1920 (90.56 °W, 34.76 °N), d. Texas in 1920 (100.92 °W, 32.81°N).

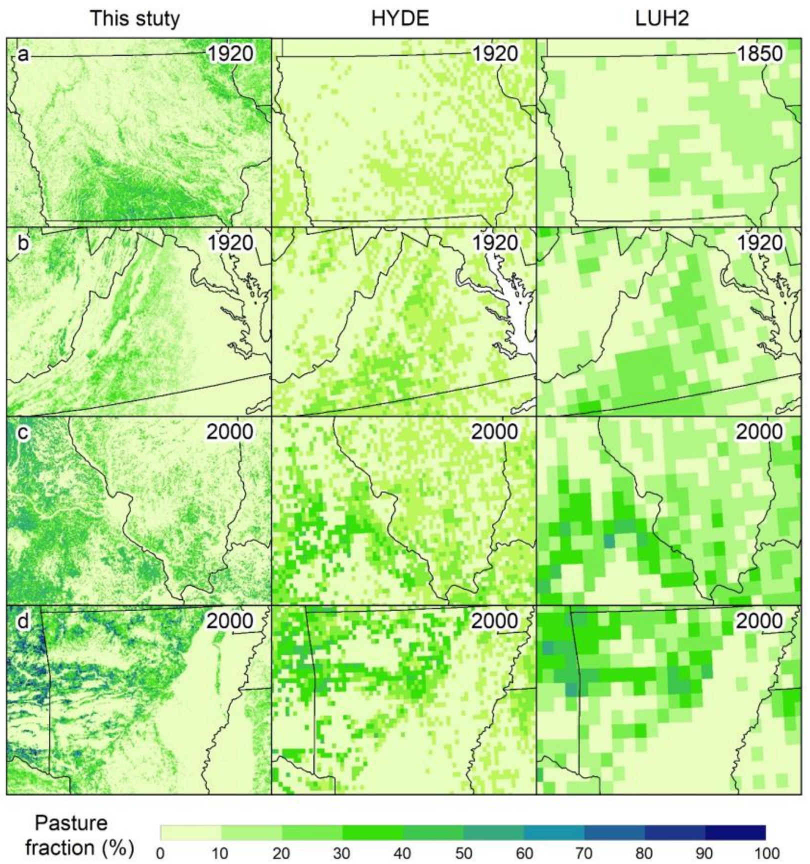

Figure R13: Visual comparison our pasture and HYDE3.2, LUH2 in four different sites (a-d). The locations of image center points are as follows: a. Iowa in 1920 (93.64 °W, 42.03 °N), Virginia in 1920 (78.72 °W, 37.96 °N), c. Illinois in 2000 (90.07 °W, 38.68 °N), d. Arkansas in 2000 (92.56 °W, 34.97 °N).

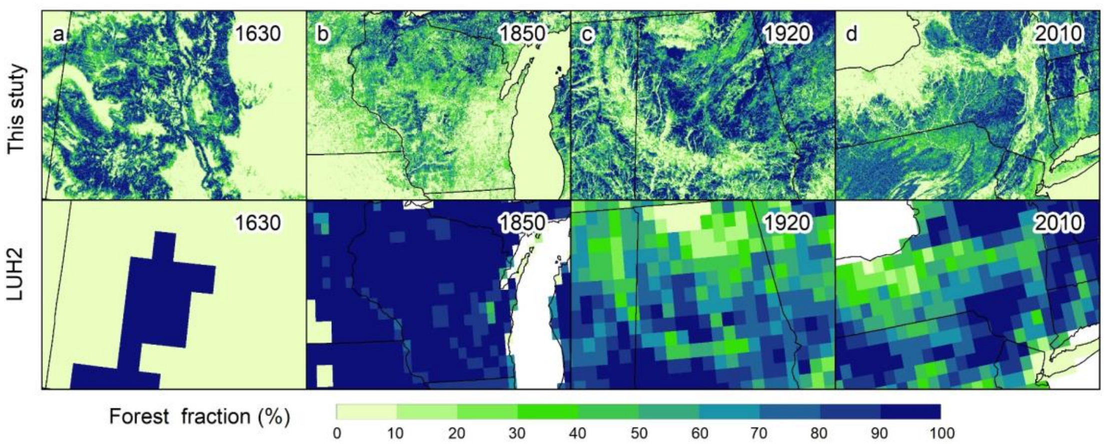

Figure R14: Visual comparison between our forest and LUH2 in four different sites (a-d). The locations of image center points are as follows: a. Colorado in 1630 (106.47 °W, 38.97 °N), Wisconsin in 1850 (89.85 °W, 44.54 °N), c. Alabama in 1920 (86.72 °W, 33.33 °N), d. New York in 2010 (75.14 °W, 42.21 °N).

Comment 33: Lines 338-339 – Agreed about the “per capita” approach.

Response: Thank you for this suggestion. We improve the urban land estimation method by using a changing urban land per capita. The HISDAC data is applied between 1810 and 2001 to reduce the bias in our estimation. The new method description can be found in the response to Comment 9.

Comment 34: Lines 339-341 – This doesn’t serve as any kind of adequate validation or even consistency check between datasets. Showing a national-scale map and stating the patterns are “consistent” isn’t valuable, and is very subjective at that scale.

Response: Thank you for this suggestion. The data validation/comparison or consistency check has been conducted in section 3.1 by comparing with NLCD data, agriculture census data, and state-level LULC area.

The national scale maps can give an overview of the spatial pattern of LULC in 1630, 1850, 1920, and 2010. We keep the national-scale map and add extra comparison figures at the regional scale to show our improvement. Please see Figure R12, R13, and R14.

Comment 35: Line 357-358 – Exactly why it’s not very valuable to compare your model results to HYDE…those data were used to help establish the model parameters themselves.

Response: Thank you for this suggestion. The HYDE data was used to reconstruct historical cropland and pasture area. But not all the periods applied the HYDE data. Moreover, we used the trend or interannual variations of HYDE data rather than the absolute value of LULC area. Thus, we compared the reconstructed historical LULC area and spatial pattern with HYDE.

Comment 36: Line 362 – Your product has higher spatial resolution than something like HYDE, but there’s no quantitative analysis of that spatial pattern that proves the superior value of that higher native resolution.

Response: Thank you for this suggestion. We add regional scale figures to show the data improvement than the dataset with coarse resolution (Figure R12, R13, and R14). Moreover, the HYDE or LUH2 have higher cropland acreage compared to US-specific datasets, like Yu and Lu, and USDA census data. We fixed this problem and went back to 400 years ago.

Comment 37: Figure 9 – It is difficult to compare all of these datasets given the definitional differences between them, particularly for pasture and cropland.

Response: Thank you for this suggestion. We agree that it is really hard to say which data is more accurate or reliable because of the definitional differences among them. But we can know whether the area of the reconstructed historical LULC dataset is in a reasonable range through data comparison. Meanwhile, the previous spatial LULC datasets are also a good reference to judge whether the reconstructed data has a reasonable spatial pattern.

Comment 38: Lines 382-383 – I'm not sure it's more "reliable", as sample-based, inventory approaches have flaws, just as satellite-based approaches have flaws. The bigger concern to me are the definitional differences, not the methodological differences.

Response: Thank you for the suggestions. We rewrite the sentence. The word 'reliable' may not suitable to describe the FIA data, I would say it has better consistency than other historical forest data for long-term study. But it is hard to say which forest data is more reliable because of the differences between forest definitions.

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 forest definition (forest cover greater than 10%). Thus, the forest area in this study was higher than the NLCD and Sohl et al. (2016) data.

In LUH2, the biomass density (BD) map is used to identify the potential forest (BD > 2 kg C m^{-2}) and non-forest at 0.5 × 0.5-degree resolution, which underestimates the forest in Rock Mountain and Northwest. NLCD is produced by using Landsat images and a comprehensive method and provides nationwide data on land cover and land cover change at a 30 m resolution (Homer et al., 2020). Spatially, it can capture the forest distribution better than LUH2. The FIA data provides critical status and trend information through a system of annual resource inventory that covers both public and private forest lands across the United States (https://www.fs.usda.gov/research/inventory/FIA), and it can provide forest trend data back to 1630.

Comment 39: Section 4.2 doesn't add a lot to the paper for me, particularly since you've already tried to explain some driving forces in the previous paragraphs of the paper. I'd much rather have the drivers woven into the story of what's happening in your results, than as a separate section.

Response: Thank you for this suggestion. The driving forces of land use and land cover change are quite complex in the United States. Though some driving forces of LULC change have been mentioned in the *Results* part, we think a comprehensive analysis of the driving forces of LULC change is still needed. We reorganized this paragraph and add more discussions. Please see Section 4.2 in the revised manuscript.

Comment 40: Lines 435-436 – I think reconstruction of historical land use is limited more by reliable, consistent historical data than methodology. Machine learning methods aren't going to be that valuable for historical reconstruction given the paucity and inconsistency of historical data for training.

Response: Thank you for this suggestion. We agree that the most important step in the historical LULC reconstruction study is to collect reliable and consistent data. The spatial allocation algorithm also impacts the reconstructed landscape pattern.

Comment 41: Section 4.3 – Somewhere in here you absolutely need to highlight the difficulties with trying to harmonize data sets with different definitions, data sources, and methodologies.

Response: Thank you for this suggestion. In this section, we add the discussion about the difficulties of harmonizing data sets with different definitions, data sources, and methodologies. Please see Section 4.3 in the revised manuscript.

Collecting reliable and consistent historical data is critical to reconstruct the historical LULC area. In this study, the LULC area is reconstructed at the state level, and LULC change at the sub-state scale is uncertain. Most of the input datasets were recorded at 5–10 years intervals, which made some insignificant fluctuations cannot be captured. Another difficulty to reconstruct the historical LULC dataset is the way to harmonizing the input datasets. Though we tried to gather the most reliable LULC datasets, the definitions for the same LULC type vary among them. For example, we applied two datasets (CAHA cropland harvested area and HYDE cropland) to generate the historical cropland area for the study period, but the definitions of these two datasets are different. Moreover, the definitions of four major LULC types do not belong to a universal classification system, making it difficult to process the total area and a post-processing step need to be conducted. The assumptions made in the reconstruction step take some uncertainties for the historical LULC area.