Response to Reviewer #4

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.

General comments:

As the three reviewers indicated, the manuscript reconstructed the history of land use and land cover over the conterminous United States (CONUS) at annual time scale and 1 km x 1 km spatial resolution in the past 390 years (1630-2020). Obviously, this is an important task. The high spatial and temporal LULC dataset is crucial for understanding and predicting the dynamics of coupled natura-human system across CONUS. I have reviewed the revised manuscript as well as the point by point responses to the comments by three reviewers. Overall, the authors have done a good job in addressing these comments. They have revised the manuscript by following the suggestions and comments closely. Having said that, I also want to point out the authors could address the comment about uncertainty associated with the new datasets by reviewer #1 better. I agree with the reviewer about a major need for the manuscript is recognition of the differences between source datasets, and the uncertainties it introduces into the modeling. I think the authors could address this comment better in revising the manuscript.

I have another comment regarding the presentation of the result. Overall, the flow and organization of this manuscript is pretty good. However, right now the first heading under Results is 3.1 Comparison with other datasets. It seems to me this is not the best. I'd like the authors to present the high-resolution and long-term dataset first. Tell us some unique features of this dataset. After that, then compare this dataset with other datasets.

Response: Thank you for the comments and suggestions. According to your suggestions, we add the uncertainty analysis by comparing the multiple datasets in their overlap period and reorganized the ***Results*** section.

Comment 1: Having said that, I also want to point out the authors could address the comment about uncertainty associated with the new datasets by reviewer #1 better. I agree with the reviewer about a major need for the manuscript is recognition of the

differences between source datasets, and the uncertainties it introduces into the modeling. I think the authors could address this comment better in revising the manuscript.

Response: Thanks for this suggestion. 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. Therefore, we added the related data comparisons in the supplementary materials to evaluate and quantify the uncertainties. Please see the following description or supplementary materials. We also add the main results of the data comparisons in the ***Discussion*** section. Please see Lines 541-555.

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:

$$RD_{urban,t} = \left| \frac{HistUrban\_CR_t - HISDAC\_CR_t}{HistUrban\_CR_t} \right| \times 100\% \quad (1)$$

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 R1a). Because there are only four overlap time points (2001, 2005, 2010, and 2015), we further calculated the state-level relative difference (Figure R1b). 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 R1b). Thus, the uncertainty induced by data difference should be little.

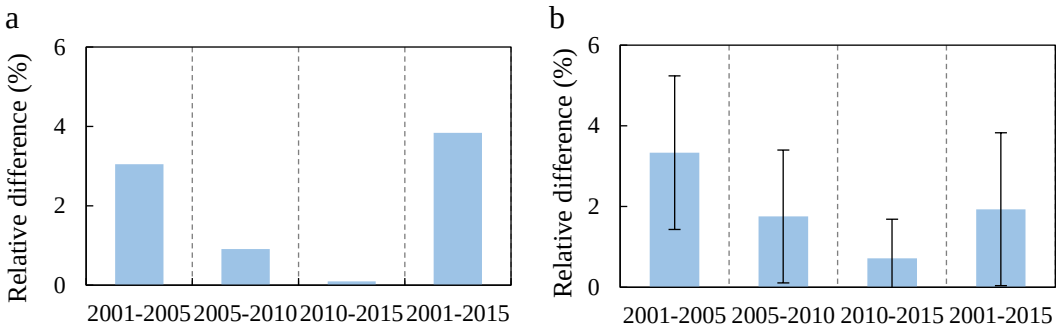

Figure R1. (a) National level relative difference in area change rate between the newly developed urban land data and HISDAC data during 2001-2005, 2005-2010, and 2010-2015. (b) State-level mean relative difference in area change rate between the newly developed urban land data and HISDAC data during 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:

$$RD_{crop,t} = \left|\frac{HistCrop_t - CHA_t}{HistCrop_t}\right| \times 100\% \quad (2)$$

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 R2a). 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 R2b), indicating that little uncertainty was introduced to the model.

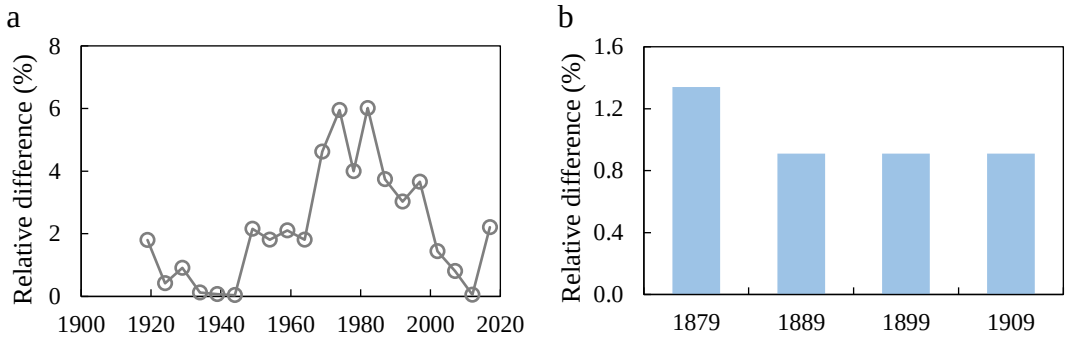

Figure R2. Relative differences in cropland area between the newly developed cropland data and the CAHA cropland harvested area data during 1919-2017 (a) and 1879-1909 (b).

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:

$$RD_{crop\_p,t} = \left|\frac{HistCrop\_CR_{p,t} - HYDE\_crop\_CR_{p,t}}{HistCrop\_CR_{p,t}}\right| \times 100\% \quad (3)$$

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 R3a). 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 R3b).

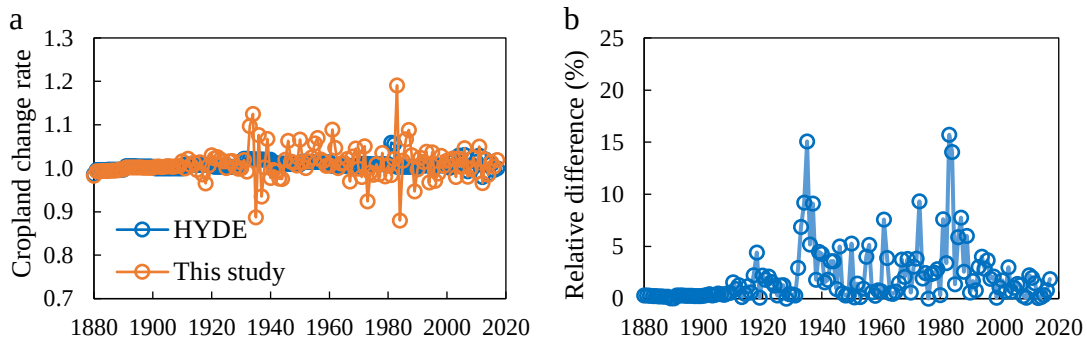

Figure R3. Cropland per capita relative difference between the newly developed data and HYDE3.2 cropland data during 1880-2017.

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 R4). 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 R4).

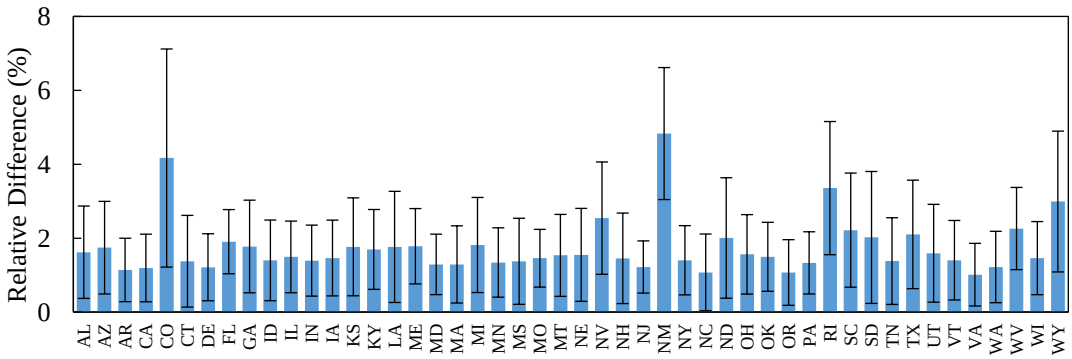

Figure R4. Mean relative difference of state-level cropland per capita change derived from HYDE3.2 cropland and the newly developed cropland land data during 2001-2017.

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:

$$RD_{pasture\_p,t} = \left|\frac{HistPasture\_CR_{p,t} - HYDE\_Pasture\_CR_{p,t}}{HistPasture\_CR_{p,t}}\right| \times 100\% \quad (4)$$

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 R5). 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%.

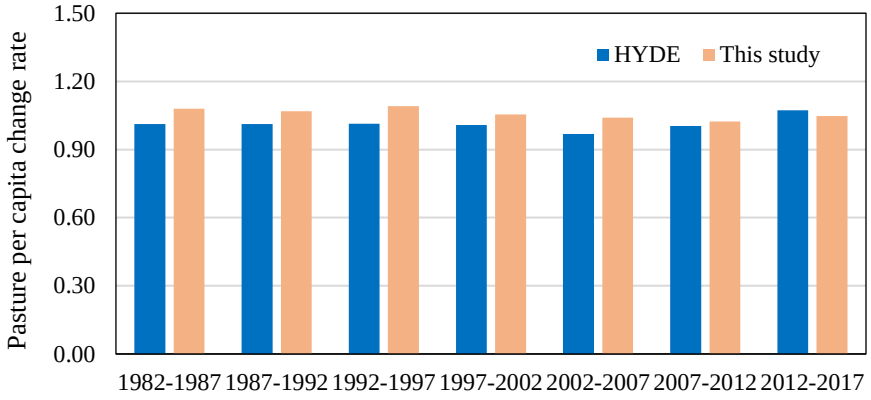

Figure R5: Pasture per capita change rate of HYDE3.2 pasture data and the newly developed pasture data during 1982-2017.

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 R6). 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.

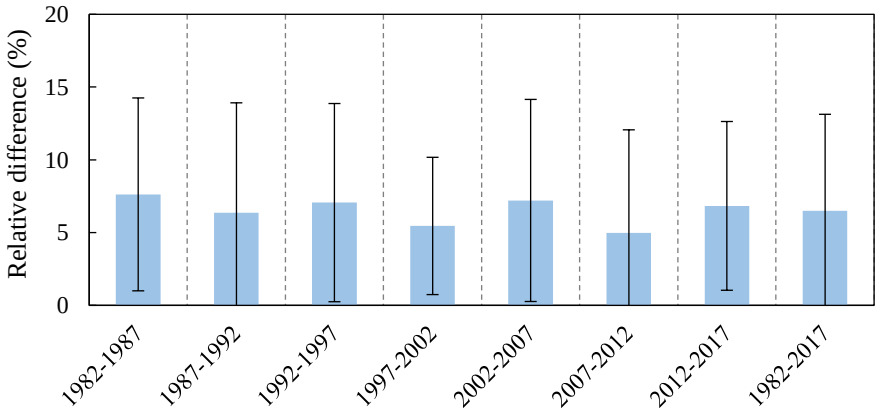

Figure R6: Mean relative difference of state-level pasture per capita change rate between HYDE3.2 pasture data and the newly developed pasture data during 1982-2017.

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.

Comment 2: I have another comment regarding the presentation of the result. Overall, the flow and organization of this manuscript is pretty good. However, right now the first heading under Results is 3.1 Comparison with other datasets. It seems to me this is not the best. I'd like the authors to present the high-resolution and long-term dataset first. Tell us some unique features of this dataset. After that, then compare this dataset with other datasets.

Response: Thank you for this suggestion. We reorganize of the ***Results*** section. We move the section 3.1 to the last part of ***Results*** (3.4). The previous section 3.2., 3.3, and 3.4 were revised as section 3.1, 3.2, and 3.3, respectively.

Specific comments:

Comment 1: The title of this manuscript can be further improved. I'd hope the title includes the word of "reconstruction" or "reconstruct" to reflect the task of this manuscript.

Response: Thank you for this suggestion. We revised the title for this manuscript. The new title for this manuscript is as follows:

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

Comment 2: Line 249-251. The last sentence of this paragraph is a little bit wordy and need to be cleared up.

Response: Thank you for this suggestion. We revised the sentence and make it clear. Please see Lines 250-251.

Comment 3: Line 269. The word "access" need to be replaced by the word "assess" instead. Similar errors need to be corrected elsewhere.

Response: Thank you for this suggestion. We check the similar errors and revised.