Response to referee comments

Anonymous Referee #1

1, Generally, there are many problems in historical records, which cannot be directly used without evaluation and correction. But in this manuscript, the reliability evaluation of the data used to estimate cropland area was missing.

→We used two types of data sources in this study, statistics and previous studies. From the beginning of the 19th century, statistics were allowed to cover more and more social areas. Over time, the data became increasingly reliable, and the process was completed in the early 20th century (Linde and Palm, 2014). Thus, we need to validate the statistics before the 20th century. Cropland data in 1665, 1723, and 1809 for Norway, and cropland data in 1907, 1936, 1950, and 1980 for Denmark were from statistics. The rest of the cropland area data came from previous studies. All cropland data from previous studies we used was also based on statistics. The cropland area data before the 20th century from earlier studies was calibrated using regional historical maps (Linde and Palm, 2014; Dam and Jakobsen, 2008; Odgaard and Rømer, 2009). For cropland area data during 1875-1999 in Sweden and Norway, Li et al. (2013) used statistics directly, and the cropland data were smaller than the real values before the 20th century. Due to the historical census's unavailability, we compared the cropland area data in 1999 from Li et al. (2013) with that in 2000 from Eurostat agricultural census. Eurostat collects information on the structural characteristics of the agricultural holdings, including land use every 10 years as an agricultural census, with two or three additional, intermediate sample surveys carried out in-between in Norway. The difference was found between statistics and census in Norway, and we calibrated the cropland area data in Norway from 1690 to 1999 (See **Page 10, lines 2-19; Page 11, lines 1-9 in revised manuscript**).

2, The methods of historical cropland estimation based on proxy data were not introduced in detail, and the estimation and allocation results were not validated or calibrated (comparison with HYDE3.2 is not calibration). These problems will undermine the reliability of the dataset.

→We have added more information on the assessment and calibration of the data sources. Allocation method validation was also added. Because there was no real value of cropland area in each 1km × 1km grid from 1690 to 1999 in Scandinavia, the allocation results calibration was impossible to achieve. Therefore, we chose Non-irrigated arable land in 2000 from CORINE Land Cover to validate our allocation method (See **3 Methods**).

3, Additionally, the results of this manuscript and HYDE3.2 are close (Figure 5a), so what's the advantages of this dataset? Although some differences exist in spatial pattern (Figure 6), the conclusion "the results of this study are better that HYDE3.2, and are more close to real land use history" cannot be reached by readers.

→ HYDE 3.2 works with a global model with assumptions that are supposed to be valid worldwide. However, historical realities differ from these assumptions. Examples are shown in China (Fang et al., 2020), the U.S. (Yu and Lu, 2018), and the European part of Tsarist Russia (Zhao et al., 2020). The differences between HYDE 3.2 and regional results are also different. Because the total cropland area in Scandinavia is much smaller than China, the U.S., and Tsarist Russia, the absolute difference between HYDE 3.2 and this study is also small. But the relative differences were between -13% and 13%. The bigger differences are shown on the national scale. The relative differences between HYDE 3.2 and this study are -5% ~ 43%, 30% ~ 291% and -57% ~ 19% in Sweden, Norway and Denmark, respectively. Norway's cropland area accounts for only 10% of the total cropland area in Scandinavia. Although Norway has the largest relative difference, it has little effect on relative differences in Scandinavia's total cropland area. Besides, HYDE 3.2 is larger than this study in Sweden but smaller in Denmark before 1930, making a small difference between HYDE 3.2 and this study in Scandinavia (See **5.1 Comparison with global datasets**).

Spatial differences between HYDE 3.2 and this study are critical to the accuracy of the cropland area dataset. In HYDE 3.2, the cropland area at the country level was allocated to grids based on population density, soil suitability, rivers, slope, and temperature. These factors determined whether and how much cropland area was allocated in each grid. This study allocated the cropland area of each Parish/Municipality/County to grids. For grids in administrative units with small cropland areas, it will not be allocated too much even if they have high weight for cropland allocation. In Scandinavia, spatial differences between HYDE 3.2 and this study showed that HYDE 3.2 model allocated cropland area to some grids in administrative units without cropland area based on this study. Or no cropland area was allocated by HYDE 3.2 model in all grids in administrative units with cropland area (See **Figure 11**). These spatial differences also reflect the necessity for spatial explicit historical cropland reconstructions in Scandinavia, although the total cropland area is tiny.

4, Finally, the study area is small, and the proportion of cultivated land is also very small (The max value is about 7% in 1950). However, the dominated land cover type, forest, was excluded in this study, which greatly undermined the significance of this dataset.

→For spatial explicit land cover reconstructions, cropland area data is often the basis for reconstructing grassland and forest land. Especially for the spatial pattern change of forest land, knowing the spatial extent of cropland can determine where the original forest land has been reduced. Except for natural factors, cropland area changes are mainly driven by human factors. Human activities have great uncertainty, and it isn't easy to simulate with models at present. Scandinavia has more detailed historical cropland area statistics than forest records. At some time points, even cropland areas at the parish level are available, which is necessary for improving the accuracy of cropland area allocation. Thus, the forest was excluded in this study.

Although the study area is small and the total cropland area is also very small in Scandinavia, the amount of data in this study is large since we collected very detailed cropland data (Please see Table 1 below). After cropland area allocation, there are about 0.8 million 1km × 1km grids with cropland area data for each time point. The total number of grids is more than 7 million.

Table 1 The amount of cropland area data (before allocation) in this study

Time points	The number of administrative units with cropland area data
1690	3010
1750	3000
1810	4575
1875	2566
1910	982
1930	1100
1950	1104
1980	1013
1999	2570

5, Why is it necessary to reconstruct the area and distribution of historical cropland in this area? What is the importance or necessity of reconstruction in this area?

→There are several reasons that we choose Scandinavia as our study area. Although Scandinavia's total cropland area is less than the area of forest land, Scandinavia has a long history of land cultivation. Farmers were drivers of economic and social change for a long history. Over the past few centuries, agriculture in Scandinavia has undergone significant changes. Studies on agricultural history in Scandinavia mainly concentrated on agricultural policy, agricultural economy, settlement and population, and landscape history. But the spatial explicit historical land cover dataset of Scandinavia was scarce. The reconstruction of historical farmland depends on the richness and quality of data sources. Scandinavia has good historical cropland area data at the Parish/Municipality/County level. Time points with cropland area records are close in Sweden, Norway and Denmark, which helps us reconstruct the cropland area change from 1690 to 1999. We have explained the necessity to reconstruct the historical cropland area change in Scandinavia in the Introduction (See **Page 3, lines 26-29, Page 4, and Page 5, lines 1-16**).

6, The reliability evaluation of the data used to estimate cropland area was missing.

→We have added the assessment and calibration of our used cropland area data (See **Page 10, lines 2-19, Page 3, lines 1-9**). The newly developed dataset is also validated (See **5.1 Validation of the dataset developed in this study**).

7, The allocation results were not validated or calibrated.

→We have validated the allocation method (See **3.4 Allocation method validation**).

8, The results of this manuscript and HYDE3.2 are close (Figure 5a), so what's the advantages of this dataset? "Underestimation" and "overestimation" are inappropriate.

→ This question is the same as question 3. We have answered it.

9, In addition, it is suggested that the author should supplement the use of this dataset as widely as possible, particularly compared with previous datasets, including HYDE, so as to let readers understand the specific value of this dataset.

→ We have added comparisons with PJ and KK10 datasets. The differences were also analyzed (See **5.1 Comparison with global datasets**).

10, Technical corrections. Page 3, line 9, forestland is unavailable in SAGE dataset. Page 5, line 6, repeated "Li et al." Page 11, line 18-20, reference(s) are needed. Figure 4 is not indispensable. Page 21, line 15 "underestimated", and line 21 "overestimation" are inappropriate.

→We have corrected your mentioned errors. We used "underestimated" and "overestimation" to show the difference between HYDE 3.2 and this study. We have modified the sentences more clearly.

Anonymous Referee #2

1, Major limitation of this study is the data uncertainty and gaps in the methodology. For data uncertainty – it can be observed that, this study did not perform any validation of the dataset and entire dataset is clearly based on only statistical datasets in the region – therefore, use of satellite based dataset as explained in the introduction is irrelevant to the study and therefore its allocation to the girds is without base.

→Besides the statistics, data sources in this study include the results from previous studies. We have added the explanation of our used datasets. These data are also validated and calibrated (See **Page 10, lines 2-19, Page 3, lines 1-9**). We used the satellite-based dataset in cropland area allocation, and the maximum cropland extents for allocation is based on the satellite-based dataset. We also gave more information about the use of satellite-based dataset (See **Page 14, lines 2-25, Page 15, lines 1-8**).

2, CORINE dataset used in the background of grids need further explanation as the allocation in the 1700 cannot be similar as allocation in 2010.

→Our first version of the manuscript used The 300m CCI-LC maps developed by European Space Agency (ESA), but not CORINE dataset. Thanks for your advice and we found land cover classes from CORINE dataset are more than those from ESA, which are more suitable for cropland area allocation in Scandinavia. We allocated the cropland area to grid cells considering different time points and administrative units. The maximum cropland extent maps used in different time points were different in our revised manuscript. Based on the total cropland areas in different administrative units, we allocated the cropland area to the extents of "Arable land, Permanent crops, Discontinuous urban fabric", "Complex cultivation patterns", "Land principally occupied by agriculture, with significant areas of natural vegetation", "Pastures, Artificial, non-agricultural vegetated areas, Industrial, commercial and transport units" and "Forests within 1km of arable land and pastures" from CORINE datasets in turn, until all the cropland was allocated (See **Page 14, lines 2-25, Page 15, lines 1-8**).

3, Another limitation is 'cropland definition' – meaning of croplands is not clearly given in the dataset as data is the mixture of grasslands, fallow lands and sometime cropland area is converted using the volume of seeds to area.
→ The category "Cropland" defined by FAO (http://www.fao.org/) was used for the cropland in this study. The "Cropland" includes areas under temporary crops, temporary meadows and pastures, land with temporary fallow, and permanent crops. The collect cropland area data was calibrated to meet the definition of "Cropland" used in this study. However, Sweden and Norway before 1875 had only cropland data recorded as the volume of seed, so we converted the volume of seed records to cropland area, which belongs to "Cropland" can fill the gap in cropland area before 1875.

4, Result and discussion part mainly explain the changes in the croplands in the allocated croplands in the study area but author should provide more detailed results on the allocation on croplands itself, for example, how much is the error percent in the grids in each year or how the allocation is showing the granularity as compared to country level polygons, detailed statistical analysis on the allocation for uncertainty and area values.
→Due to the lack of "true value" of the spatial explicit cropland area in history, it is almost impossible to analyze the uncertainty of our cropland area allocation results. However, we selected arable land area in 2000 as an example to analyze the uncertainty (See **3.4 Allocation method validation**).

5, Lastly, the paper need serious grammatical English correction and some restructuring for example, methodology can be well explained with flow charts and results may have first section to explain the plain allocated maps itself rather than the change in croplands.
→We have added flow charts for explaining our methods. The first section was also added to explain the allocated maps. A professional person has polished the manuscript and serious grammatical English has been corrected.

Specific comments:
6, Line 10 in section 1 –"The decrease of natural vegetation is accompanied by

an increase in cropland area." need to support with references. Decrease of natural vegetation may have other drivers including increasing agricultural activities.

→ "The decrease of natural vegetation is accompanied by an increase in cropland area." was concluded based on the result of Pongratz et al. (2008) (See **Page 2, lines 8-10**). The expression was maybe misleading. We have deleted this sentence.

7, This study introduction should be focused on agriculture in Scandinavia and should provide more background on it in the introduction rather than detailed explanation about global croplands and its changes over time. Authors may provide more literature review on croplands in Scandinavia and avoid exaggerated details about global croplands and its changes.

→ One of this study's primary purposes is to verify and improve the global land use datasets using regional historical materials with higher resolution. Thus, we must introduce the advantage and the shortage of global datasets. We have revised the "Introduction", and the redundant parts were removed. According to your advice, we introduced the agricultural and land use history in Scandinavia in the "Introduction", please see **Page 3, lines 26-29, Page 4, lines 1-19**.

8, In methods, there are several gap in the information and analysis. For example, missing data of several counties for many time-stamps is calculated using interpolation but there is no any validation performed to support the output. Also, the allocation of croplands from county-level historical dataset to grids is not clearly explained and uncertainty in the conversion process remains firm.

→ Interpolation of missing cropland data based on cropland in neighboring counties or cropland change trends at adjacent time points is the two most commonly used data interpolation methods for historical cropland reconstruction. Since only Denmark had 3% missing data, interpolation results using different methods have little effect on Denmark's total cropland area. We have given more detailed explanations about cropland area allocation (See **3.3 Cropland area allocation**). The cropland allocation method was also validated (See **3.4 Allocation method validation**).

9, Methods did not explain the accuracy and validation of the resulting cropland dataset and therefore, the reliability and usage of this dataset is limited.

→ We have added the validation of our dataset (See **5.1 Validation of the dataset developed in this study**).

10, High resolution dataset term is leading in the entire paper – 30 arc second or 0.5 degree datasets may not be considered high resolution. Author may need to rethink on the use of the term high resolution.

→ Compared with the satellite-based data in modern times, 30 arc seconds (we have changed the resolution to 1km in the revised manuscript) is not high resolution.

However, the current global cropland area from HYDE 3.2 has the highest resolution of 5 minutes. At present, the highest resolution of the existing historical cropland cover dataset for Scandinavia is 5′×5′, also from HYDE 3.2. So for the historical cropland area dataset in Scandinavia, 1km is a high resolution. Moreover, since the 1km×1km resolution cropland dataset is obtained by allocating the cropland area of each administrative unit, further increasing the spatial resolution will lead to a decrease in the accuracy of the dataset.

11, Lastly, it is very important to know why Scandinavia region for this study and then why Agricultural lands for historical land use study if there can be another significant land covers which may affect. The context of the paper need to be updated with better explanation and focus on the main goal of the paper.
→For spatial explicit land cover reconstructions, cropland area data is often the basis for reconstructing grassland and forest land. Especially for the spatial pattern change of forest land, knowing the spatial extent of cropland can determine where the original forest land has been reduced. Except for natural factors, cropland area changes are mainly driven by human factors. Human activities have great uncertainty, and it isn't easy to simulate with models at present. Scandinavia has more detailed historical cropland area statistics than forest records. At some time points, even cropland areas at the parish level are available, which is necessary for improving the accuracy of cropland area allocation. Thus, the forest was excluded in this study. We have added the explanations about why we chose Scandinavia and "Cropland". Please see the **Introduction.**

12, Also, the goal of the study is misleading as it is different in the introduction line 26 in Section 1 vs in methodology in line 4 in section 3.
→ We have re-write the **Introduction** and **Methods**.

13, Overall the work is significant with further additions and analysis along with English corrections.
→ We have added more explanations and polished our manuscript.

References:

Linde, M., Palm, and L. A.: Sverige 1810: Befolkning, jordbruk, skog, jordägande, Rapport för Vetenskapsrådets project Databasen Sverige 1570-1805: befolkning, jordbruk, jordägande, Institutionen för historiska studier, Göteborgs universitet, 2014.

Dam, P., and Jakobsen, J. G. G.: Atlas over Danmark: Historisk-Geografisk Atlas, København: Det Kongelige Danske Geografiske Selskab, 2008.

Odgaard, B., and Rømer, J. R.: Danske Landbrugs-landskaber gennem 2000 år, Gylling: Narayana Press, 2009.

Fang, X., Zhao, W., Zhang, C., Zhang, D., Wei, X., Qiu, W., and Ye, Y.: Methodology for credibility

assessment of historical global LUCC datasets, Science China Earth Sciences, https://doi.org/10.1007/s11430-019-9555-3, 2020.
Yu, Z., and Lu, C.: Historical cropland expansion and abandonment in the continental U.S. during 1850 to 2016, Global Ecology and Biogeography, 27, 322-333, https://doi.org/10.1111/geb.12697, 2018.
Zhao, Z., Fang, X., Ye, Y., Zhang, C., and Zhang, D.: Reconstruction of cropland area in the European part of Tsarist Russia from 1696 to 1914 based on historical documents, Journal of Geographical Sciences, 30(8), 1307-1324, https://doi.org/10.1007/s11442-020-1783-y, 2020.