Response to comments

Title: A 1 km global cropland dataset from 10000 BCE to 2100 CE

MS No.: ESSD-2021-219

Referee #2

General comments

Comment 1:

This study targeted the construction of global cropland dataset starting from 10000 BCE and extending to the future in 2100 CE. Great effort was made to integrate, harmonize, and downscale multi-datasets to produce the final 1 km global dataset. The new dataset is expected to be very useful for a broad spectrum of studies and applications since it considers mapping the historical, as well as future, cropland distribution at a relatively optimal spatial resolution to such a large geographical scale and long period. However, going through the manuscript could raise several questions to the readers, which need to be considered by the authors. Some of those questions are listed as follow.

Response 1:

Thank you very much for the comments and suggestions. Please see the detailed point-by-point responses below.

Specific comments

Comment 1:

Understandably, the date of 10000 BCE was the staring of farming, but integrated datasets (Table 1) were back only to 1950 CE, regardless of the population dataset sourced from HYDE? Hence, defining the suitability for cropland was totally dependent on the population data during the period before 1950, which makes the quality of the cropland mapping during the period totally dependent on the quality of the single layer of the population. Hence, the reader could question the added value and uncertainty when starting the production of the maps from 10000 BCE, with the lack of data covering this long period?

Response 1:

Thank you for your comments. During 10000 BCE-1950 CE, the cropland suitability was not only dependent on the population data. As described in Section 2.2, except population, we also used the other variables in suitability evaluation for 10000 BCE-1950 CE. Nevertheless, since these variables are unavailable before 1950 CE, they remained unchanged for these years. It is true that using the unchanged variable data brings some uncertainties. However, it is the commonly used strategy under data

acquisition limitations in land use/cover simulations and there is no better one. Besides, despite the uncertainties, there is a great demand for cropland reconstruction since the start of farming. The complete cropland distribution information throughout the whole process of agriculture development is important and benefits a lot for the overall understanding of agricultural activities. In the study, we also evaluated the uncertainty for the years beyond the recent decades. As shown in Fig. 7, the results are acceptable. As for the added value, compared with HYDE 3.2/LUH2, we used higher resolution variable data and improved methods. According to Fig. 4-6, our maps can better capture the cropland distribution details and spatial heterogeneity, which are very valuable and can serve as a more comprehensive alternative for related applications.

We added the statement in **Section 4, Paragraph 4**: "However, using the unchanged variables is the commonly used strategy under data acquisition limitations in land use/cover simulations and there is no better one. Despite the uncertainties, there is a great demand for complete cropland distribution information throughout the whole process of agricultural development, which is important for the overall understanding of agricultural activities. Compared with HYDE 3.2/LUH2, we used higher resolution data and improved methods. Even if for the years beyond the recent decades, our maps can still better capture the cropland distribution details and spatial heterogeneity (Fig. 4-6). In the study, we also evaluated the uncertainty for these years and the results are acceptable (Fig. 7)."

Comment 2:

Suitability map played a crucial role in mapping production. The authors clearly stated that the influence of the variables defining the land suitability for agriculture was not equal throughout the whole period. For instance, population was the key variable defining the land suitability in earlier dates due to the traditional farming practices and weak global links (as mentioned in line 169). However, another assumption in line 175 was made by authors that I believe it needs further explanation and justification. The assumption is that "the impact of population on cropland distribution is negligible for most regions in the future years (2010 CE-2100 CE)", why?

Response 2:

Thank you for your comments. We agree that there is still a relationship between total future global cropland area and total future global population. However, the relationship between cropland area and population will become much weaker when it comes to finer grid scales (Meyer and Turner II, 1992). Future technology development will weaken the population demand in crop farming, making crop farming no longer heavily rely on population. Moreover, future trade activities will become more frequent. The demanded crops of many regions will be not produced locally but from other regions. Besides, the population continues to migrate and gather to urban, resulting in stronger population intensification. In some regions especially metropolises, although population increases a lot, the cropland area increases little or even decreases. Thus, at grid-scale, the

indication of future population change on future cropland change will become weak for lots of regions worldwide. Additionally, there is no suitable future population dataset. Therefore, we exclude the population variable in future cropland simulations.

We added the further explanation in **Section 2.2**: "In the future (2010 CE-2100 CE), technology development will make crop farming no longer heavily rely on population, and trade boom will enable many regions to meet the crop demands of the local population through purchasing from other regions. Besides, future population intensification tends to be stronger. Thus, at the fine grid-scale, the relationship between cropland area and population will become much weaker for lots of regions worldwide in the future. Additionally, there is no suitable future population dataset."

Reference

Meyer, W. B. and Turner II, B. L.: Human population growth and global land-use/cover change, Annu. Rev. Ecol. Syst., 23, 39–61, https://doi.org/10.1146/annurev.es.23.110192.000351, 1992.

Comment 3:

The comparison between the constructed cropland dataset and HYDE 3.2 dataset in Figure 7A showed a relatively close r value for recent years (1700-2000) but a high increase of RMSE value. A reader could expect that the error in quantifying cropland area would decrease when getting closer to the current time. But this is not the case in the constructed dataset, any justification from the authors about that?

Response 3:

Thank you for your comments. As interpreted in Section 3.1, "the RMSEs also decline with downscaling time step increases (i.e., from present to past), which is mainly attributed to the reduction of cropland proportion in pixels".

The RMSE is calculated based on the cropland proportion differences of the corresponding pixels between our dataset and HYDE 3.2. It is directly affected by the absolute values of cropland proportion. The cropland proportions are generally lower for the early stage, so, the RMSEs are also smaller. It should be noted that the RMSE is scale-dependent since it represents absolute differences rather than relative differences. Therefore, direct comparisons of RMSEs between different years/scenarios are actually invalid due to the different scales. Besides, the larger RMSE values do not represent larger errors, but only indicate greater differences between our data and HYDE 3.2.

We supplemented the statement in **Section 3.1**: "It should be noted that the RMSE is scale-dependent since it represents absolute differences rather than relative differences. Therefore, direct comparisons of RMSEs between different years/scenarios are actually invalid due to the different scales."
