



Comment on **essd-2021-219**

Anonymous Referee #2

Referee comment on "A 1 km global cropland dataset from 10000 BCE to 2100 CE" by Bowen Cao et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-219-RC2>, 2021

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:

- Understandably, the date of 10000 BCE was the starting 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?
- 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?
- 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?