Earth Syst. Sci. Data Discuss., author comment AC2
https://doi.org/10.5194/essd-2020-382-AC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Zhen Yu et al.

Author comment on "A historical reconstruction of cropland in China from 1900 to 2016" by Zhen Yu et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-382-AC2, 2021

Spatial mapping of croplands distribution is essentially needed for multiple agricultural and environmental studies. In particular, the long-term croplands spatial datasets, like the one introduced by this study, enable the long-term tracking of cropland changes and analyzing their driving factors. Hence, I believe that the current study is introducing a very useful cropland dataset and so, it has great potential to be published in ESSD after considering a few minor revisions. I have listed three comments below for your consideration.

Response: We thank the reviewer for valuing our work!

- What is the final spatial resolution of the produced cropland grids? It will be better to have this information in the abstract.

Response: We thank the reviewer for the suggestion. The final spatial resolution of the cropland product is 5 km × 5 km. We will add this information to the abstract.

- It is clear that the cropland areas at the provincial level estimated from the constructed cropland dataset were compared with other earlier studies, and all were compared to official statistical data. However, I believe that this comparison might be biased because the official statistical data (from NLRB) was part of the tabular data used to construct the final cropland layers, right? That is why the correlation between the produced cropland layer and NLRB data was very high (R-Square = 0.999, k=1.0, in Figure 3). Therefore, I believe the accuracy assessment, through traditional error matrix and estimated OA, PA, Kappa, etc., can be performed here for selected years when crowdsourcing validation samples (or field samples) are available to reflect the efficiency of the cropland spatial allocation method introduced by the current study. Some crowdsource validation sets are publicly available and can be used for validation such as the one from Geo-Wiki (https://www.nature.com/articles/sdata2017136) and Global validation sample set v1 (http://data.ess.tsinghua.edu.cn/).

Response: We thank the reviewer for the suggestion. Yes, we used the official statistical data from the National Land and Resources Bulletin (NLRB) for the comparisons of different cropland datasets. The NLRB is also used in the reconstruction of the cropland in this study since it is the most authoritative report officially released by the Chinese government. We provided the provincial

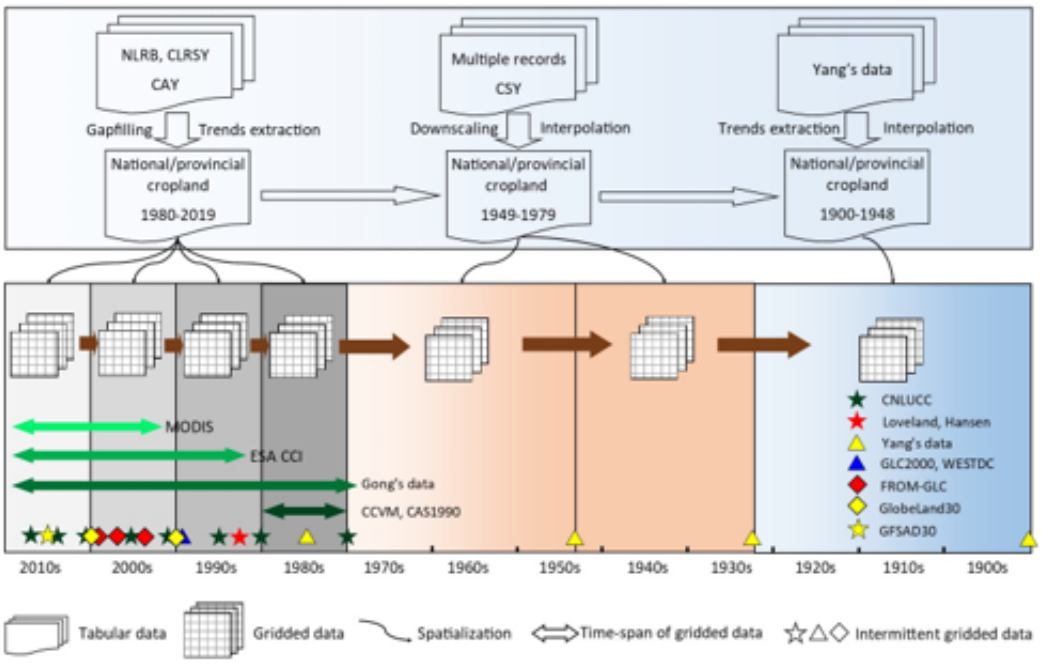
level comparison to illustrate the model performance in the reconstruction of the data.

The reviewer's suggestion is very helpful because the ground-true observations provide the validation at the finer resolution. However, direct validation is not suitable for the cropland product reconstructed in this study as traditional error matrix and accuracy assessment is designed for validation of classified land cover maps. The cropland product, as well as the HYDE data, describes cropland in percentage, and thus the site-based validation is not directly applicable.

Despite such limitation, we agree that the validation is important and necessary. Therefore, we chose the intermediate product (Boolean type map at 100-m resolution before aggregating to 5-km cropland percentage map) derived from multiple data sources for the validation using the Global validation sample set v1 (http://data.ess.tsinghua.edu.cn/). Note that the intermediate product describes the high possibility of absence (0) or presence (1) of cropland in each grid cell, but other land cover types were not available. Therefore, we directly compared the intermediate product with cropland sites provided in the Global validation sample set. We found that, in total of the 356 cropland sites from mainland China, 219 (62%) of the sites were correctly identified in our intermediate product (from 2001 to 2010), which is in the range of 198 (56%) to 248 (69%) identified from the MODIS product (MOD12Q1, from 2001 to 2014) and the GFSAD30 (2015), respectively. However, since the intermediate product is not the focus of this study, we will add this information to the supplementary file.

- In figure 1, GFSAD30 was produced for 2015 and so I think it belongs to the 2010s rather than the 2000s, right?.

Response: We thank the reviewer for pointing out this. Exactly, the GFSAD30 was produced for 2015 and it was mistakenly placed in Figure 1. The figure will be updated to:



☐High resolution figure can be found in the supplement pdf file☐

Please also note the supplement to this comment:
https://essd.copernicus.org/preprints/essd-2020-382/essd-2020-382-AC2-supplement.pdf