

## Comment on [essd-2020-382](https://doi.org/10.5194/essd-2020-382)

Yu Ye (Referee)

---

Referee 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-RC3>, 2021

---

A spatially-explicit cropland distribution time-series dataset is necessary for the accurate assessment of biogeochemical processes in terrestrial ecosystems and their feedback to the climate system. This study reconstructed a continuously covered cropland distribution dataset in China spanning from 1900 to 2016 by assimilating multiple data sources and identified the abandonment and expansion of cropland, which has important contribution to this research area. However, some questions are as follows:

- Many scholars (Li, Yang, Wei et al.) have done the research on the gridding allocation of cropland in China during the 300 year. Why do you choose 1900 as the starting time point to repeat the allocation? Whether you have more dependable historical data sources for 1900 to 1949 or have you revised some time section's value?
- Whether this study considered the difference between pure and mixed cropland grid cells, or different proportion range of cropland grid cells in different remote-sensing products when they are used for reconstruction?
- For satellite-data period and pre-satellite era (1900-1979), this paper has used different spatializing approach. For gridding images, cropland fraction, distance to urban, correlation with previous years and resolution were considered for the weighted value in the potential cropland map in this paper. So, the results, to a large extent, will be decided by the dependability of different satellite products. Whether you have think about the priority of high resolution and dependability of remote-sensing products? Or change another word, for China, maybe GlobalLand30 or Gong's data is more fit for?
- Reconstruction of low cultivation ratio regions maybe has large uncertainty, especially Xinjiang and XiZang, where the results of crop distribution area or intensity all exists unreasonable. You had better analyze it more from the view of method uncertainty or discrepancy of different sets of products.
- For data in 1900-1979, what's the difference between Yang's constrained CA models with other scholars' method? For example, Wei's "the partition and layering-based gridded method" or Li's method are all based on land suitability for cultivation affected by climate, soil and elevation etc. What's the significance of this study on methodology?