

Earth Syst. Sci. Data Discuss., referee comment RC1  
<https://doi.org/10.5194/essd-2021-7-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Comment on **essd-2021-7**

Anonymous Referee #1

---

Referee comment on "The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019" by Jie Yang and Xin Huang, Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-7-RC1>, 2021

---

This is an interesting and well-written paper. The authors seek to develop an annual land cover dataset of China, and such kind of continuous time-series products are lacking and urgently needed in the current research. In this context, the authors first derived training samples from existing data and visual-interpretation, and then fed it to random forest classifier to obtain land cover results. They compared the results against those from the MODIS, CCI and GlobaLand30, with a relatively better accuracy achieved. In particular, it is very interesting to see the detailed comparison between the CLCD datasets and other well-known thematic products (e.g., GISA, GAIA, GFC, GSW). In addition, the authors also examined spatiotemporal patterns of land cover changes.

The study is technically sound. The use of the Landsat and Google Earth Engine for 30 m land cover mapping is a good choice, especially given the post-processing to ensure the consistency. The validation of the result is generally comprehensive and reliable. The analysis in the discussion section flows reasonably from the results. As such, I recommend that it be published after minor revisions.

1) Line 22. In the title, the CLCD annual dataset spans 1990-2019. But, the land cover changes spans 1985-2019. Can you clarify this issue?

2) Line 52. It is vague to use "over 90,000 visually-interpreted training samples". Please explicitly identify how many samples were used.

3) Line 91. What do you mean by "inconsistency between different Landsat sensors"?

4) Line 111. Why only "Quite Sure" Geo-Wiki samples were chosen?

5) Line 173. "trainings" should be "training".

6) Line 216. I'm not sure what do you mean by saying "dominant LC" here. Please clarify.

7) Line 237. It is interesting to compare against these existing datasets. But please clarify the reason why MODIS and CCI were chosen, since they have relatively coarse resolution.

8) Line 243. The time span of GSW is different from that described in Line 324.

9) Line 244. "date" ->"data".

10) Line 256. Again, what does "dominant LC" signify here? Please specify it.

11) Line 297. Why were these three years chosen? How about other years?

12) Line 324. "surfaec"->"surface".

13) Line 395. Users may be interested in the zoom-in maps of Fig. 13, but the spatial extent of some zoom-in maps seems inconsistent to the loss/gain images. It would be good to highlight the extent of zoom-in maps in the corresponding images.

14) Line 411. I notice that you have mentioned the uneven coverage of Landsat 5 for several times. Therefore, it is suggested to explicitly demonstrate or explain this issue.