

Please note that the referees' comments are in black and our responses are in blue.

Point by point reply below

General comments:

This is a short and interesting manuscript describing a reference dataset at 10 meter resolution for land cover and land use mapping. Authors trained experts to judge the land cover type of 10 m resolution sub-pixel through high resolution images, geo-tagged photos and other scientific datasets from multiple sources and compiled results with a global coverage. This product is an important and valuable training/validating data source for fine resolution land cover type or biodiversity mapping. But before I can recommend the paper for publishing on ESSD, I have the following concerns related to how authors present their work.

We would like to thank the Referee for their comments and the feedback on the manuscript. We appreciate the time and effort spent on providing a comprehensive review, which we believe has significantly improved the quality of the manuscript.

Specific comments:

1. It is interesting that you provide burnt areas as another land use cover type, but I'm curious about how you can use this information for other studies. Can you provide some examples?

This could be used as a source of verification of fire events that happened around the year 2015.

2. I have a concern about potential time frame mismatch across different products you used. When judging one location, it may be possible that the multiple image or scientific data products you used is obtained from different years, and the land cover type at the exact location might change (e.g., due to urban expansion, deforestation, etc.). I assume this to be one of the uncertainty sources which need experts to put more careful consideration before making decision. How did you handle this?

We checked the locations using the Google Earth Pro application where we could see the exact image dates. If the images were more recent or outdated, we undertook an additional check for changes by visual interpretation of NDVI time series and Sentinel-2

images available from 2015. If there was not enough data available, such locations were labeled as "not sure".

3. I would strongly recommend authors to provide a certain quantitative evaluation on the accuracy of your data product, as the accuracy evaluation to be one of the requirements for publishing your work on ESSD.

We have significantly improved the description of the quality assurance processes (section 2.4). We have added the accuracy number, which falls within the range of 90-95%, and the description of the process of how this number was obtained.

Technical corrections:

Line 64: Where is the start location (lon - lat) of your global systematic sample?

We have added the start location.

Line 67: "in areas with low classification accuracy" Which reference data you used to determine the classification accuracy? Through the assessment information from intermediate versions of the CGLSLC100 land cover map? Please clarify.

We have added further clarification in the text. Areas with low classification accuracy were determined exclusively by visual inspection of the intermediate version of the map. No specific sources of reference data were used to determine the classification accuracies.

Line 68: What is "the initial training data set"? The data produced in step (1)? Need to clarify.

We have now added a clarification to the text to indicate that the initial training data set was produced in step 1.

Line 80: "(NDVI) time series derived from Google Earth Engine (GEE)". Which satellite products did you use to calculate NDVI?

We have now listed the satellite products and added references. This includes Landsat 32-Day Composite (Collection2), MOD13Q1.005 Vegetation Indices 16-Day Global 250m, and PROBA-V C1 L3 Daily at 100m.

Line 81: "a time series of Sentinel 2 images that can be retrieved from Sentinel hub" The same question here. Which product of S2 you used? Or have you created a true color image?

We did not produce any S2 composites by ourselves. We used the ones provided by the Sentinel Hub service, including natural color (bands 4, 3 and 2) and false color images (bands 8, 4 and 3).

Line 95: "18 land cover experts" might be good to acknowledge them if they agree, since they're authors of the dataset.

We have listed all the contributors on Zenodo, and we have now added one more section to the manuscript describing the author's contribution and acknowledging all the experts involved.

Line 109: "Thus, out of 100 interpretations that were checked, an interpreter could have made up to 5 to 10 misclassifications, which were mainly random mistakes." Did you strategically design the distribution of tasks to have, for example, 10% of the dataset assigned to 2 or more interpreters at the same time for accuracy assessment purposes?

The performance of each expert was checked independently and on a regular weekly basis. There was no special design for the distribution of tasks. This was a near real-time continuous process, where at the end of each week, we randomly chose a subset of 100 locations out of all the annotations that were submitted during the past week by each expert, and we reviewed them. We have now improved this paragraph.

Figure 2. I saw desert regions are marked as snow and ice. It's a typo or "missing data" over these regions? If so please use another color to represent regions with "no data". Please clarify and revise the figure.

This is missing data. We have now adjusted the figure and have highlighted snow observations in a different color.

Line 140: "for ecosystem mapping and complex modelling of biodiversity". Can you provide more details about how to map biodiversity?

We have added an example of a possible use such as "an indirect uncertainty assessment of land cover maps used to produce terrestrial habitat types (<https://www.nature.com/articles/s41597-020-00599-8>)."

