

Comment on **essd-2022-224**

Anonymous Referee #1

Referee comment on "CALC-2020: a new baseline land cover map at 10□m resolution for the circumpolar Arctic" by Chong Liu et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-224-RC1>, 2022

The authors created a new 10-m land cover map for the circumpolar Arctic based on the combined use of Sentinel-2, Sentinel-2, and DEM and multiple-source training samples. It is challenging to map land cover over the Arctic due to the limited number of snow/ice-free observations. Yet, the current manuscript does not well demonstrate the land cover product is promising, particularly for the validation and the comparison. I have also lots of questions regarding the method. Please see my detailed comments below.

(1) All the samples filtered by the SOM may be helpful to improve the classification, but it will select similar samples for each category. The validation data, with the same resources as training data, will overestimate the accuracy of the land cover map. The samples from flux tower sites may be OK, because of the real independence to the training samples, but the number is too limited, and several important land cover types are lacking, such as cropland, forest, snow/ice (Table S2). Here I highly recommend the authors create new and independent validation samples stratified on the Arctic land cover map (see Olofsson et al. 2014, RSE, good practices).

(2) The description of the sample is not clear, i.e., how many samples were in the study area (L 133), how many changed pixels were removed (L 134), how did the authors determine the land over type for the pixels within the 90-m buffer (L 135), what was the strategy (L 139), etc. The strategy has a 40% maximum proportion and a 3% minimum proportion during the training data selection. How did the authors apply the strategy to "each" used land cover product, such as NLCD, Canada land cover, and GlobeLand30? This strategy was originally designed for selecting the training samples, that will be directly used to train the final random forest model; however, when the lichen/moss class was interpreted, how did the authors know its proportion? By the way, how did the authors immigrate the category of the existing products to Table 1, which are so different?

(3) The approach of comparing the new map to existing land products is not objective. For example, why did the authors focus on the agreement between the new map and the

existing products, rather than focusing on the accuracy? How did the authors align the different definitions of land cover between products? What are the means for DM, DS, and DD? What this comparison analysis delivered to audiences is a huge disagreement among the maps, and so what? How about the accuracy of each product? At least, the authors can provide the accuracy against the samples from flux tower sites.

Other comments:

L 110: What were discarded from the Sentinel-2 QA band? Snow/ice observations were excluded?

L 157: The number of each category should be given as well.

Section 2.4.2 Man-made impervious surface mapping: As far as I understand, the impervious surface from CAMI-2020 has the highest priority, and the authors classified the other 9 land cover types only. If that, I suggest clarifying that the classification method did not include the impervious surface, like in Section 2.4.3.