Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2021-435-RC2, 2022

Comment on essd-2021-435

J. Uusikivi (Referee)

Referee comment on "A 41-year (1979–2019) passive-microwave-derived lake ice phenology data record of the Northern Hemisphere" by Yu Cai et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-435-RC2, 2022

Overall review

The authors have used an automated method to extract ice phenology data from passive microwave data. The data set presented here and explained in the article is generally very interesting and will be useful for research community. The data set is likely the longest and most comprehensive ice phenology data set from satellite-based observations for that large number of lakes. This data covers multiple climatological areas and lake sizes and is therefore well worth publication. Data set is usable in the present format.

Comments:

I would like to present 2 recommendations to improve the usability of the data and the manuscript.

- Data set does not include any sort of error estimates for dates, duration, or maximum ice cover area. In the manuscript is long discussion on the errors and their possible sources, but these should be quantified in the data, or at least in the manuscript. It is very difficult to compare this data set to other similar data sets without this information. In the manuscript one major target for this data is climate research, it is difficult to draw conclusion if error marginals are unknown. To use this data to complement data gaps of in situ archives of ice phenology, more precise definition of the errors and their sources compared to the GRLIPD and GLERL ice cover data sets should be included.
- Using 37 GHz H-polarized data has some limitations in distinguishing ice and open water. Signal can be strongly affected by open water surface roughness from wind (for example, K.-K. Kang et al.: Estimating ice phenology on large northern lakes from AMSR-E; doi:10.5194/tc-6-235-2012). This problem and its implications to the data is

not discussed in the manuscript at all, and it is not covered in any of the references provided. By discussing this matter or providing references that discuss this, will make this data much more reliable and usable.

I also have some minor comments to consider:

- on line 176: "When the lake is water covered, the TB for land-contaminated pixels will be higher than that of a pure pixel, while when the lake is ice covered, the TB will be lower than that of pure pixel." Last 2 words: Is it pure pixel of ice/water/land?
- on line 271: "When the lake area was large enough, the gradual freeze-up or break-up within the pixel can be ignored, but for small lakes, it may lead to certain deviations in the lake ice phenology results." What are the certain deviations?
- on line 312: "Overall, if the overlapping time between the two dataset was longer, the lake ice dates could show a higher consistency." How or Why that could be the case?
- on line 353:" This is because a buffer of 6.25 km was used to exclude pixels near the lake shore, which happens to be the place where lake ice forms first." If this is the only explanation in the difference between GLERL data and this data set, one would expect the difference to gradually wannish as one nears 100% ice coverage. This is not the case in all the lakes in all the years. Why is that?