

Earth Syst. Sci. Data Discuss., referee comment RC1  
<https://doi.org/10.5194/essd-2021-435-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on essd-2021-435**

Tamlin Pavelsky (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-RC1>, 2022

---

### **Review**

A 41-year (1979-2019) passive microwave derived lake ice phenology data record of the Northern Hemisphere

By Cai, Duguay, and Ke

For *Earth System Science Data*

By Tamlin Pavelsky, University of North Carolina

### **Summary**

This manuscript presents a new dataset of lake ice phenology for several dozen large lakes in the Northern Hemisphere based on passive microwave data. This dataset, based on a new passive microwave dataset beginning in 1979, represents the most comprehensive passive microwave lake ice phenology dataset to date. The dataset is compared against in situ observations for several lakes, including the Laurentian great lakes, and another passive microwave dataset from the AMSR-E and AMSR-2 instruments and generally found to be consistent with these other datasets, though with greater consistency for larger lakes at higher latitudes.

## Overall Review

Fundamentally, the dataset presented in this paper is likely to be useful to the scientific community. I recommend some alterations to the paper and to the dataset itself, which I believe will improve it. However, I believe it should likely be published after consideration of reviewer comments. The strengths of the dataset are strongly related to the strengths of passive microwave remote sensing in general. There is a long record of global data, it is not impeded by clouds or other atmospheric effects, and there is a considerable literature suggesting that it can be used to detect ice status in large lakes. The dataset presented here is quite clearly the most comprehensive passive microwave lake ice phenology dataset, and it is likely to be of use to researchers interested in ice phenology patterns, including as a point of comparison for datasets collected using other methods (e.g. ground-based surveys, optical remote sensing, active-source radar).

There are 2 primary weaknesses that I believe should be addressed before the dataset and paper are finally published:

- The authors refer multiple times to uncertainty in ice phenology for various lakes, but there is no uncertainty field included in the dataset itself for any of the data fields. It is probably impossible to include all sources of uncertainty, but that should not stop the authors from include those sources they can quantify. I have some experience in this area with optical datasets, and what we have generally included are gaps between viable observations (see, for example, Pavelsky and Smith, 2004 and Zhang et al., 2021). These gaps are also clearly present in the passive microwave data (though for different reasons) and are discussed throughout the paper. The authors should at least be able to represent this source of uncertainty. However, they may also be able to represent other sources related to lake size. For example, there are several lakes that are represented by only one passive microwave pixel. In these cases, ice flagging is binary and likely to be less accurate. In any case, I would like to suggest that the authors quantify uncertainty on all dates as fully as possible and include those estimates in the dataset. Of course this also entails including a description in the paper of what sources of uncertainty are included in the flags.
- There are a number of places in the paper where statements are made that may be true but which are not supported by any data or analysis in the paper. These include:
  - Line 308: "The main reason for the difference between lake ice dates from in-situ observations and passive microwave is their different observation ranges. In-situ records rely on observations of lake ice status visible from lake shores by human observers, while passive microwave satellites record *T<sub>B</sub>* from the entire lake surface (here within the pre-defined buffer)." While this seems reasonable to me, there is no evidence in the paper that it is true, and no other work is cited.
  - Line 336: "Therefore, AMSR-E/2 data can capture more information near the lake shore than SMMR, SSM/I & SSMIS data, which led to the directional differences between the lake ice phenology dates extracted from the two datasets." Same as above.
  - Line 355: "The correlations between the ice cover for Huron and Ontario were lower than that of the other three lakes, which is also because the ice cover of these two lakes were usually small, and the ice first forms near shore which may not be covered in the set buffer (Figure 5)." This assessment potentially conflicts with the assessment shown in the next paragraph indicating that problems with the MTT

algorithm my result in lower accuracies for Lake Ontario.

- Line 435: "The differences between the in-situ observations and the lake ice phenology extracted in this study were mainly due to the different fields of view of human observers versus satellite instruments." Again, while this is plausible, no evidence of this explanation is explicitly shown in the paper.

I would strongly recommend that the authors either qualify these statements, provide evidence for them, or remove them. I would tend to hope for one of the two former options, as long as more concrete evidence suggests that they are correct.

Other than these two substantial areas for potential improvement, all other comments (listed below) are minor.

### **Specific Points**

Line 14: I would write ". . . in middle and high latitude regions."

Line 17: An alternate to what?

Line 28: I would write "consistency" rather than "consistencies" in both cases in this line.

Lines 39-40: I would recommend citing a few papers here, as there are a number of papers that have looked at this. For example, Smejkalova et al., 2016; Sharma et al., 2016.

Line 42: Might also consider citing Knoll et al., 2019 here.

Line 50: I might change the wording here, as a reader could be confused into thinking that all 865 site records begin in 1443, when most of them begin much latter.

Line 72: I might, again, cite the Smejkalova et al. 2016 paper here.

Line 101: I would probably delete "Satellite" before Nimbus-7.

Line 104: I would replace "data of the" with "data from the"

Line 108: "is against" should be "is compared against" or similar.

Figure 1: perhaps it's unnecessary to do, but I believe there are a number of Patagonian lakes that might be close to the necessary size threshold for inclusion in this study. If the authors have not looked at these lakes to see if they are viable, I would recommend doing so given the paucity of similar records in the southern hemisphere.

Line 141: Typo at the end of the line. Should be "For comparison with lake ice phenology. . ."

Line 144: I would write "on 7 lakes" instead of "of 7 lakes."

Line 149: I would write "from the CETB dataset."

Line 155: I would write "from the National. . ."

Line 184: I would write "For the remaining pixels. . ."

Line 185: Why were thresholds of 5% and 95% chosen? Is there any sensitivity compared to say, 80-90% and 10-20%?

Line 194: This choice to omit ice covered periods of <30 days is significant. Would you then say that your estimates of ice duration likely underestimate total ice duration, as they ignore any intermittent ice cover occurring during the breakup or freezeup seasons? If so, I would explicitly mention this.

Line 228: I would write "For years with multiple lake ice records. . ."

Lines 246-249: A minor point, but given the size of these lakes, I'm not sure there's any need to include the fourth significant digit in latitude. All are nearly 0.5 degree (or more) in N-S extent.

Line 270: I would recommend making sure you stay in consistently the present or past tense for this sentence and the following one.

Line 276: I would write "periodically missing data" instead of "periodical data missing."

Line 301: I would write "compared" rather than "used to compare"

Line 302: I would write "complete ice cover" rather than "completely ice covered."

Lines 313-316: while I agree with the statement in this sentence to a degree, I would argue that the agreement with GLRIPD records is not particularly strong in many cases. As such, I'm not sure I agree that the analysis presented here provides strong evidence for this statement. Rather, I would say that remotely sensed observations can complement in situ measurements.

Lines 335-336: Make sure to keep verb tenses the same in this sentence.

Line 429: The conclusion that lakes at low latitudes and/or small areas tend to have larger uncertainties would be much more robust if there were a more consistent uncertainty quantification, as mentioned above.

## References

Knoll, L. B., Sharma, S., Denfeld, B. A., Flaim, G., Hori, Y., Magnuson, J. J., ... & Weyhenmeyer, G. A. (2019). Consequences of lake and river ice loss on cultural ecosystem services. *Limnology and Oceanography Letters*, 4(5), 119-131.

Pavelsky, T. M., & Smith, L. C. (2004). Spatial and temporal patterns in Arctic river ice breakup observed with MODIS and AVHRR time series. *Remote sensing of environment*, 93(3), 328-338.

Sharma, S., Magnuson, J. J., Batt, R. D., Winslow, L. A., Korhonen, J., & Aono, Y. (2016). Direct observations of ice seasonality reveal changes in climate over the past 320–570 years. *Scientific Reports*, 6(1), 1-11.

Šmejkalová, T., Edwards, M. E., & Dash, J. (2016). Arctic lakes show strong decadal trend in earlier spring ice-out. *Scientific reports*, 6(1), 1-8.

Zhang, S., Pavelsky, T. M., Arp, C. D., & Yang, X. (2021). Remote sensing of lake ice phenology in Alaska. *Environmental Research Letters*, 16(6), 064007.