Review of 'PolarLakes: A Bi-Weekly Dataset of Supraglacial Lakes on Antarctic Ice Shelves from Multi-Sensor Satellite Observations (2015-2024)'

Overall, this is an excellent manuscript, which introduces an incredibly valuable dataset to the Antarctic community. I have a few major comments a list of minor comments. Once these have been addressed, I would recommend this dataset for publication.

Major Comments

1. At present, this dataset refers to itself as 'pan-Antarctic'. Given that it only targets 22 ice shelves, and only specific regions of interest within these 22 ices shelves are covered, I find this slightly miss-leading. Please can the authors make it explicitly clear several times throughout the paper that only specific ice-shelf areas are mapped. Of course, the limitation of this approach is that as meltwater areas expand further towards the ice-shelf-ocean edge this product is unlikely to capture the expansion.

   Following from this, please can the authors make it clear if they intend for these products to be produced and made publicly available on a rolling basis. If so, have the authors considered mapping meltwater across the full extent of each ice shelf, rather than just a region of interest?

   Please can the authors also publish the regions of interest for each ice shelf. This will be necessary for future users of the dataset.

2. Throughout the manuscript it is difficult to understand what work has been independently undertaken for this manuscript, what comes from Dirscherl et al., (2020, 2021), and what comes from Baumhoer and Kohler et al. (2025). For example, it seems like you followed the approach of Dirscherl et al. (2021) to create two separate U-Net's for S1 and S2. However, in Section 2.3.1 there is a brief mention of using a Random Forest Classifier (Dirscherl et al., 2020) to create S1 ground truth labels. As a reader it is almost impossible to work out where various aspects of this dataset are sourced from, and it needs to be made much clearer. Please also apply this comment to Section 3. Furthermore, Dirscherl et al. (2021) propose several ideas for improvement beyond the scope of the original paper (in section 4.3 Future Requirements). Were any of these followed up on before using the model in the context of this study?

3. It would be good to see a figure of some of the tiles used to train the U-Nets, as outlined in 2.3.1.

4. Section 3.2 ‘Lake detection confidence’ should perhaps be renamed to something that better reflects ‘scene visibility’. A clearer description of the methodological workflow is also needed here.

Minor Comments

L22-24: This sentence is trying to outline the differences in meltwater extent between the AP and the EAIS, but it lacks some clarity. I find your descriptions of these changes later in the manuscript much clearer. Please adjust.

L41-43: Discussion of hydrofracture processes and impact on ice-shelf stability is missing key reference to Banwell et al., 2013

L52-53: The seasonal evolution and distribution of surface lakes is also influenced by local ice-shelf surface and sub-surface conditions e.g. firn air content, blue ice extent, ice-shelf topography.

L64-64: ‘[…] followed by publications observing supraglacial lakes on ice shelves before their disintegration on Prince Gustav, Larsen A and Wilkins ice shelves’ > ‘[…] followed by observations of supraglacial lakes on Prince Gustav, Larsen A, and Wilkins ice shelves prior to their (partial) disintegration’ + references needed

L71: ‘ration’ > ‘ratio’

L93: ~~‘In the end’~~ > ‘Our research will provide…’

L138: It would be useful to the reader if the bands used for these spectral indices could be given here

L140: ‘Futhermore, ~~and~~’

L151: Elsewhere in the manuscript the austral summer is referred to as November to March, why does Sentinel 1 only use December to March? This needs clarifying within the text.

L160-162: It’s not immediately clear whether the number of 21,200 Sentinel-1 image tiles is a result of the augmentations (i.e., whether there were ~21,200/12 true image tiles before augmentation)

L183-184: ‘Train validation split for training was set to 80 to 20 %.’ This sentence needs to be clarified.

L184-186: How were the model weights initialised?

L190: The manuscript needs to be clearer about moving to inference here. So far 480 x 480 patches have only been mentioned in relation to training data so first it needs to be established that the patches described here are presumably the scenes described in 2.1.1 and 2.2 (this section should probably be numbered 2.1.2) tiled into the same

patch size, run through the model to get prediction probabilities and that it is those predictions that are reassembled.

L206: ‘size of > 0%’. Is this a typo? Otherwise it’s a completely ineffective threshold surely?

L210: Is this relative to the Corr et al. (2022) and Stokes et al. (2019) datasets? If so, please add references again here.

L247-248: ‘and for Sentinel-2 on 16 sites…’ >> This sentence doesn’t make sense to me, please re-write.

L338: ‘1924 square kilometers’ > Please write numerically.

L371: Please explain why recurrence maps were calculated for different time periods.

L393: ‘utilized’ > ‘utilised’

L444-446: ‘proprietary data’ could be clarified here, presumably you are referring to the labelled data used to train the models? And ‘pre-trained weights’ is ambiguous here, usually this implies transfer learning from an external backbone in the model training phase, but you contextualise with ‘that have been produced for this publication’. See major comment 2, it is unclear here what model training has been done for which publication.

Figure 4: The resolution of this figure is pretty low, making it blurry when printed.

Figure 8: Consider making the colour for the WAIS in the bars darker, as its incredibly hard to see.