To the reviewers and editor,

We would like to thank the reviewers again for their valuable feedback, which has helped us make several meaningful improvements to both the dataset and the manuscript.

To make the data more transparent, accessible, and reusable, we added shapefiles delineating the areas of interest (AOIs) in which supraglacial lakes were detected, added a README.txt, and included additional TIFF files indicating the sensor contribution to each classification, so that users can clearly see how the results were derived. We also converted the TIFF files to cloud-optimized GeoTIFFs, ensuring easier access and full STAC compatibility for the wider community.

On the manuscript side, we incorporated additional literature references to better position our work within the current state of knowledge, re-phrased several sections for improved clarity, and added new figures in the supplementary material to make our results easier to follow and understand.

We believe these changes directly address the reviewers' comments and have substantially strengthened the quality, transparency, and usability of both the manuscript and the underlying dataset. Please find our answers to your comments in blue below. Line numbers refer to the final manuscript without track changes.

Kind regards,
Celia Baumhoer (in-behalf of all co-authors)

Reviewer 3

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

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.

Thank you for raising this point, which was also mentioned by the other reviewers. We have revised the abstract to clarify that we map the most frequently ponded ice shelves L17: "To

address this gap, we introduce the PolarLakes dataset, generated using a deep learning-based workflow that automates the bi-weekly detection and mapping of supraglacial lakes across 22 Antarctic ice shelves experiencing most frequent ponding". This selection covers approximately 98.6% of the total supraglacial lake area.

As the reviewer correctly notes, we restricted the classification area to the most frequently ponded regions to reduce computational cost. However, these areas still extend well beyond the current lake extent, meaning that growth of existing lakes will be captured, while entirely new lakes forming far from present locations would be missed. For full transparency, we now publish the shapefiles defining these mapping extents in the folder <shelfname>_stats_plots alongside the dataset, so that users can clearly see the spatial coverage.

We agree that for a fully rolling, continuously updated dataset, this fixed extent would eventually need to be revisited and expanded as lake areas grow or new lakes emerge elsewhere. Developing PolarLakes into such a living dataset was indeed our long-term goal. However, project funding concluded before we were able to implement continuous processing. We hope to pursue this extension in future work, should further funding become available.

2. Throughout the manuscript it is di icult 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?

We agree that it was very difficult for the reader to understand where the different parts of the processing pipeline came from. We clearly state this now in the beginning of the Methods Section L114: "The PolarLakes dataset is generated through a multi-step workflow that combines optical and SAR satellite data. The workflow builds on established methods for Sentinel-1 from Dirscherl et al. (2021a) and Sentinel-2 processing from Koehler et al. (2025). While both pipelines follow the approaches described in the respective publications, we computationally optimized them for large-scale processing and, for the first time, combined them into a unified classification of maximum lake extent that leverages both sensors. This dual-sensor integration allows the workflow to exploit the complementary strengths of SAR and optical imagery, providing more robust lake extent estimates than either sensor alone. "

We agree that the sentence with the Random Forest Classifier rather confused the reader than helped for understanding. We just wanted to say that lake prediction datasets form the Dischерl et al. 2020 study based on Random Forest Classifier were used to locate approximate lake occurrences. They were rather used exploratory than for delineating labels. This was solely done on Sentinel-1 imagery. We rephrased for clarity L181: "Ground truth labels for Sentinel-1 scenes are manually created for the classes "water" and "non-water". As this manual classification based on SAR data alone is very challenging we used Sentinel-2 imagery and supraglacial lake extent mapping products for support originating from Dirscherl et al. (2020). These products were just used to locate and explore approximate lake locations but not to create outlines. The outlines of the manual labels are solely based on the Sentinel-1 imagery as significant differences between Sentinel-2 and Sentinel-1 lake visibility exist. This is likely due to varying

acquisition times, data availability, and the impact of wind or freezing on lake visibility in SAR imagery compared to optical imagery."

Thank you for bringing up the "Future Requirements" mentioned in Dirscherl et al. (2021). We would have loved to work on these improvements, but unfortunately this project focused on upscaling the existing approach rather than improving the method itself. Therefore, we only optimized the coastline masking to make it computationally faster. We did not create additional training labels, nor did we experiment with the neural network architecture for potential improvements. Furthermore, while adding more bands with additional polarizations would likely be beneficial for classification accuracy, Sentinel-1 IW data only exists consistently across Antarctica with a single polarization. Hence, for our pan-Antarctic approach, it was not feasible to add more channels to the classification.

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.

We created an additional Figure S1 in the supplementary to show examples of Sentinel-1 and Sentinel-2 training labels as well as predictions of the test data. We hope this makes it easier for the reader to understand where differences between Sentinel-1 and Sentinel-2 are located. Further examples of training and testing data can be found in Dirscherl et al. 2021 and Koehler et al. 2026.

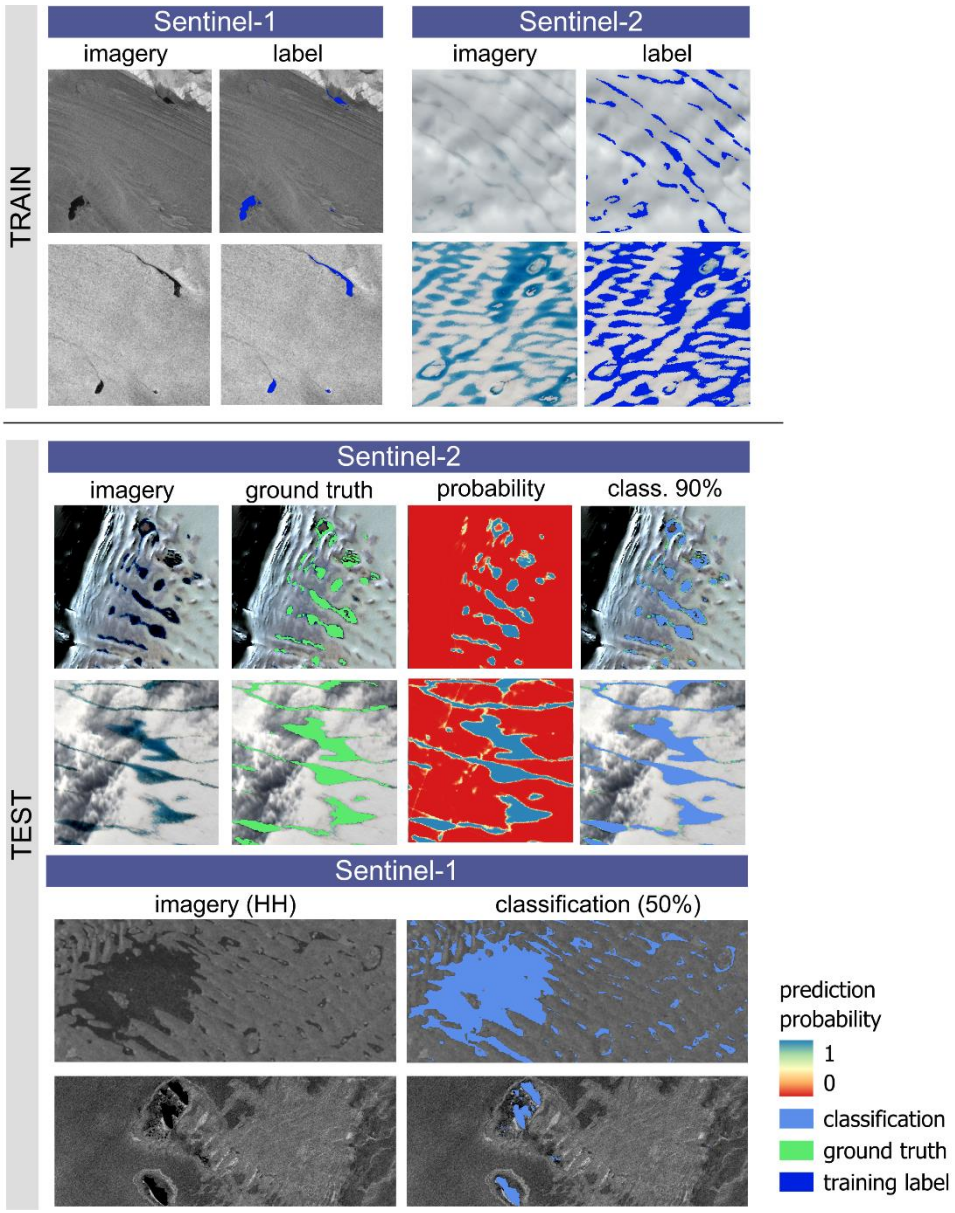

Figure S 1 The upper row shows the training labels marked in blue. Predictions were thresholded at 90% probability for Sentinel-2 and 50% for Sentinel-1, with thresholded predictions shown in pale blue. Ground truth labels for Sentinel-2 are shown in green. For Sentinel-1, testing was performed using a point-wise validation approach, so no ground truth label is shown. Contains Copernicus Sentinel-1 and Sentinel-2 Data.

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.

Good point. We renamed the section to: "Lake detection confidence based on optical scene availability"

Minor Comments

L22-24: This sentence is trying to outline the di erences 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.

We re-phrased for clarity L22: "East Antarctic lake extent exceeded that of the Antarctic Peninsula from the start of the record through the pan-Antarctic peak in 2020, after which the

pattern reversed. The maximum lake extent declined in East Antarctica and increased along the Antarctic Peninsula, experiencing a secondary regional peak in 2023."

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

Thank you. Added the reference.

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.

Indeed, we extended the sentence to L60: "Their distribution and seasonal evolution are shaped by a combination of air temperature, solar radiation, wind conditions, firn air content, blue ice areas, ice shelf topography and atmospheric modes which vary spatially and temporally across different regions of Antarctica (Arthur et al., 2020; Dirscherl et al., 2021b; Mahagaonkar et al., 2024)."

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

We added the correct citations and corrected the typo (Larsen B instead of Larsen A) L78: "First lakes were reported by Mellor and McKinnon (1960) on Amery Ice Shelf, followed by observations of supraglacial lakes on Prince Gustav (Arthur et al., 2020), Larsen B (Leeson et al., 2020), and Wilkins ice shelves (Braun et al., 2009) prior to their (partial) disintegration".

L71: 'ration' > 'ratio'

Corrected.

L93: 'In the end' > 'Our research will provide...'

Corrected.

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

We mention the bands for the indices now L162: "For the final datasets, image stacks are constructed that encompass the spectral bands in the visible green and red (Sentinel-2 bands 3 and 4), in addition to the Soil/Water Index (SWI) calculated from the blue and SWIR 1 bands(Dirscherl et al., 2020) and the Automated Water Extraction Index (AWEInsh) based on the SWIR1 an 2, green and NIR bands (Feyisa et al., 2014)."

L140: 'Futhermore, and'

Removed '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.

Here only the temporal extent of the training data is given and the training data is from December to March. We improved the phrasing for clarity to 176: " For Sentinel-1, the training and validation dataset includes image tiles covering periods of pronounced surface melt occurring mostly during January between 2016 and 2020, focusing on the months December to March."

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)

We re-phrased this sentence to make sure it is clear that this is after augmentation: "After tiling the labelled Sentinel-1 imagery by tiling it at 480 x 480 pixels, with a 200-pixel overlap, and augmenting it 12 times by rotating and flipping, in order to artificially increase the training data a total of 21,200 tiles were obtained."

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

This was probably a bit too much machine learning jargon. We re-phrased to: "During training the train-validation split was set to 80 to 20% in order to monitor train and validation loss."

L184-186: How were the model weights initialised?

We added "randomly initialized" to clarify we trained from scratch.

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.

Thank you, we fixed the section numbering. Additionally, we elaborate a bit more on the inference by writing L221: "For inference, the pre-processed satellite scenes are tiled into 480 x 480 pixel patches with 100 pixel overlap and fed to the model The pre-trained weights are used for each sensor individually to predict maximum lake extent. Afterwards during post-processing the individual 480 x 480 pixel patches are reassembled again by averaging predictions in overlapping regions."

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

Thank you for spotting the rounding error. The threshold was meant to be > 0.01%.

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

Added references.

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

We hope our re-phrasing makes this clear now L285: "The accuracy assessment is based on ten independent test sites from unique Sentinel-1 scenes and 8 unique Sentinel-2 scenes including 16 different sites."

L338: '1924 square kilometers' > Please write numerically.

Corrected.

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

We added the explanation as follows L423:" Note that the recurrences were calculated over different time periods depending on satellite data availability the time series lengths are different…."

L393: 'utilized' > 'utilised'

Changed to British spelling throughout the manuscript.

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.

"Proprietary data" was definitely wrong wording here. The Jupyter Notebook includes an example how to train the with published Sentinel-2 train-validation data. We also added citations to make clear where the weights are coming from L513: "The code incorporates Jupyter Notebook examples, which illustrate the methodology of training the U-Net on the published training data from Sentinel-2 or own data, and the subsequent execution of inference based on pre-trained weights from Koehler et al. (2025) and Dirscherl et al. (2021a) published at Zenodo (Baumhoer et al., 2025b)."

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

We will provide figures as vector graphics (PDF) for publication and not the low-resolution png version used for this word file.

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

You are right, it is very hard to see the comparably small extent of the WAIS compared to EAIS and AP. We selected a darker pink and longer y-axis to enhance the visibility for the WAIS contribution.