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 1

This is a thorough presentation of a great dataset that I'm sure will be very beneficial to the community. I have two 'major'/general comments and many minor comments.

Major comments:

Throughout the manuscript, I struggled to understand the contribution that Sentinel-1 provides. Since you choose to include it, I assume it is useful. Figure 4b indicates that is contributes significantly a lot of time and Table 2 and Figure 5 indicate good performance as measured by an F1 score. However, there are many instances in the text where you state or imply that the Sentinel-1 mapping is not useful or reliable. None of the map figures indicate much of Sentinel-1 contribution; where it does contribute, it doesn't seem to agree much with Sentinel-2 (Figure 4b; though acknowledging the images are often on different days within the two-week period). I think this could be resolved by simply providing a figure showing clearly what is mapped as a pond in Sentinel-1 and how it compares with Sentinel-2, ideally also with comparison to some kind of ground-truth/test data.

Thank you for raising this point. In Figure 4b Sentinel-1 contributes in many times to a high percentage in case no suitable S2 imagery is available. As the scale is in percentage this can mean, that only a very small lake area was detected overall. S1 is a reliable source to map open surface water. This is rarely the case which makes it a minor data source but still provides additional information during the lack of S2 data or heavy cloud cover as can be seen in our new Figure S2 in the supplementary additionally to the Figures 6 and 10. Additional Figure S1 in the supplementary also shows examples of labels for Sentinel-1 and Sentinel-2 to provide a better understanding how lakes are mapped in the different data sources.

Additionally, for more transparency we extended our dataset by the folder <shelfname>_beweekly_contrbution_mosaics providing detailed information in which areas which sensor mapped the lake. We added this information to the Data Availability Section L497: "The folder *<shelfname>_biweekly_contribution_mosaics* provides the maximum lake extent with information which sensor classified each pixel. The integer values in brackets indicate classification from Sentinel-1 (1), Sentinel-2 (2), and overlap (3)."

There are now several datasets of surface meltwater presence in various forms, resolutions and time periods. I appreciated the discussion of the differences and similarities between this dataset and some of those, but I think the manuscript would be significantly strengthened by some quantitative comparison when/where suitable overlaps exist. I realise the definition of melt/ponded water differs in some datasets.

This is indeed a very interesting point and we dived into it in the method paper of Koehler et al. which is accepted and soon to be published. Figure 8 and 9 visualize and quantify mapping differences between the datasets of Corr et al. 2022, Stokes et al. 2019, Dirscherl et al. 2020 and Dell et al. 2024. Overall, the U-Net maps slightly larger lake areas due to a more consolidated mapping due to U-Nets segmentation approach compared to the pixel-based approaches. Please see the full paper and in-depth discussion here: https://www.cambridge.org/engage/coe/article-details/68b960c5728bf9025e37f559 .

We decided not to pick up this topic again in this paper, as Koehler et al. already elaborate a lot on this topic and comparing this dataset also including SAR would not be directly comparable to other products solely relying on optical data.

Finally, can you add detail regarding your plans/funding to maintain and update this dataset, unless the intention is for it to remain 'static'.

The dataset was developed within the 1-year Lakes4Antarctica project. We would love to extend the dataset to a living dataset and implement real-time monitoring. Unfortunately, currently no funding is available for this. In case further funding gets available, we will continue our efforts to create a living dataset regularly updated.

Dataset:

1) There seems to be an errant total_<start_year>_<end_year>_<shelf>.tif in each of the annual mosaic directories, which seems identical to the total_yyyy-mm-dd_yyyy-mm-dd_<shelf>.tif file

Thank you for spotting this. After re-naming we forgot to remove the old files. We deleted them now and keep only the most recent one.

2) I think the data would be more readily adopted by other users if they were provided also/instead as NetCDF files, so that temporal analysis doesn't require loading in many different files and parsing many filenames to retrieve dates.

Thank you for suggesting to make this dataset more accessible. The file format used at the DLR GeoService is GeoTiff. Therefore, we decided not to go for NetCDF files but provide Cloud Optimized GeoTiffs which make loading and access much easier. Furthermore, the dataset will be available within the STAC Catalog of the DLR Geoservice end of this year after transitioning to a newer data portal. Via STAC raster calculations and loading several years is very straight forward.

Minor comments:

Line 39: Please cite Tuckett et al. (2025) here as well. https://www.nature.com/articles/s41558-02502363-5

Added.

Line 48: Although I agree that meltwater drainage to the bed of grounded ice on the AP probably does lead to transient acceleration of grounded ice, I think the evidence for it is still ambiguous, so I recommend rephrasing this sentence accordingly.

We re-phrased to L52: "Supraglacial lake drainage potentially can accelerate grounded ice flow on the AP …."

Line 49: There are now some field observations indicating acceleration of grounded ice (though only 1 km upstream of the grounding line) driven by surface derived meltwater: https://www.nature.com/articles/s41467-026-72724-x.

Thanks for the hint. We added L54 "But there is recent evidence from field observations that indicate acceleration of grounded ice driven by surface derived meltwater also in East Antarctica (Sugiyama et al., 2026)."

Line 55: "supraglacial lake formation is expected to expand further inland and across higher elevations, raising concerns over the long-term stability of Antarctic ice shelves" I'm not sure the second part follows logically from the first part? In any case, I think this too much of a simplification of the current state of understanding. Please consider adding more detail and nuance to this based on e.g. van Wessem et al (2023): https://www.nature.com/articles/s41558-022-01577-1

We improved the explanation as follows L63: "As surface melt intensifies under projected future warming scenarios (Trusel et al., 2015), supraglacial lake formation is expected to expand not only on ice shelves that already experience surface pondingnot only on ice shelves already experiencing surface ponding but also on currently cold, dry ice shelves, which van Wessem et al. (2023) show to be more vulnerable than previously assumed, with melt-ponding thresholds low enough to be reached by the end of this century under continued warming."

Line 57: perhaps also add interannual variations and response to changes in the firn/snowpack ability to store/refreeze meltwater? As per Tuckett et al. (2025) and perhaps Dunmire et al. (2024): https://www.nature.com/articles/s43247-024-01255-4

Good point, we added L68"Besides the expected future increase in meltwater production, meltwater ponding also exhibits strong interannual variability, and the ice-sheet surface is becoming increasingly prone to ponding as its capacity to retain meltwater declines. Repeated melting and refreezing within the snowpack forms impermeable ice lenses that lower surface permeability, while reduced accumulation depletes the firn air content available to buffer meltwater (Dunmire et al., 2024; Tuckett et al., 2025)"

Table 1: please add Tuckett et al. (2025) here and to the associated text.

Added to the table and text.

Dataset name – “PolarLakes”: are there plans to expand this dataset beyond Antarctic ice shelves? If not, then perhaps something like IceShelfLakes would be a more suitable name? That might come up more in Google searches too. Up to you of course.

We weren’t sure jet whether we would expand the dataset in the future to further areas. Therefore, we selected the name PolarLakes. The keywords will make sure that the dataset is findable also with search strings like “ice shelf”.

Line 91: “recent supraglacial lake dynamics giving insights on critical hydrological processes”. Can you be more specific here? “dynamics” and “hydrological processes” are quite vague and when it comes to supraglacial lakes readers will likely think of hydrofracture and rapid lake drainage, but since that process typically occurs over hours rather than weeks, I suspect you’re referring to lake growth and slow drainage/refreeze here?

We clarified this by re-phrasing to L107 “The PolarLakes dataset provides high resolution information on recent supraglacial lake dynamics giving insights into slow lake drainage, refreezing and lake expansion, offering opportunities for improved ice sheet modelling efforts.”

Line 92: I think the dataset will also be very useful to those interested in ice sheet surface mass balance and firn processes.

Good point, we added L108“The biweekly resolution of this dataset offers new opportunities for evaluating surface mass balance models against observed melt-pond timing and extent, for constraining firn air content, and for assessing ice-shelf vulnerability to hydrofracture.”

Line 108: What do you mean by “accessible”? Aren’t they all free online?

Should say “available”.

Line 121: Consider naming the software/tool used first, as it’s not clear until the end of the paragraph where one can find this function.

We now name the ESA SNAP Software at the beginning of the paragraph.

Line 121: I suspect it makes no difference, but the type/precision of orbit data used should probably be stated.

The most up-to-date and precises data is used by this function automatically. We re-phrased to L164 “The basic pre-processing steps include updating the orbit metadata with the most up-to-date orbit file with the “Apply Orbit File” function.”

Lines 121-125: please provide a few more details on the thermal noise removal, radiometric calibration and speckle filtering algorithms (or citations for each). If the radiometric calibration just used the image annotation data, then there’s no need to provide details on that.

We now define the speckle filtering algorithm (Refined Lee), the remaining functions are the basic ones of ESA SNAP and can be looked up in detail in the user manual if needed.

Line 122: Should “naught” be “nought”?

Corrected.

Line 131: AOI hasn’t been defined yet

Added Area Of Interest (AOI).

Line 135/136: I struggled to understand this. What does “cloud masking algorithm within the image acquisition strip of the scene” mean? Is a scene comprised of multiple strips?

For clarity, we re-phrased to L160:“Minor instances of two to four pixels with no data values introduced to the cloud masking algorithm within each Sentinel-2 L1C scene are addressed through spatial interpolation.”. The Level-1C and Level-2A products are already in 110 x 110 km tiles in the UTM grid being image acquisition strips of the Level-1B products. But as we start processing with Level-1C we removed “image acquisition strips” as this might confuse the reader.

Line 143: please specify the original resolution of the SW infrared bands.

Added L168 ”The shortwave infrared bands (SWIR1 and SWIR2) with an initial spatial resolution of 20 m are resampled for indices calculation to correspond with the 10 m resolution of bands 3 and 4.”

Line 150: “the dataset” – training or validation or both?

We clarify be now by “the training and validation dataset”

Line 155: perhaps refer the reader to Section 3.1 for more details

Added reference to section 3.1.

Line 179: As someone who has never used a UNet, I have no idea what “the bottleneck” means.

We added an explanation to make this clear L209: “At the bottleneck (the deepest layer of the U-Net), an ASPP module…”

Line 185: Should that be “factor of 10” rather than 0.1? i.e. the learning rate gets smaller, not bigger.

“Reduced” is ambiguous here. We changed it to “multiplied by a factor of 0.1”

Line 193: Can you show some of these prediction probability tiles next to the corresponding image, and perhaps a graph showing the deviation from the ‘ground truth’ for a sample of probability levels? This would give the reader another measure of uncertainty in the mapping.

We thank the reviewer for this suggestion. We have added a new Figure S1 in the Supplementary Information, showing examples of training and testing data along with prediction probabilities for both Sentinel-1 and Sentinel-2. We hope this provides readers with additional insight into our classification approach. Please note that prediction probabilities are only stored as an output in the Sentinel-2 pipeline. In the Sentinel-1 pipeline, they are an intermediate step and are discarded after thresholding. We can therefore only show the thresholded prediction (at the 50% probability level) for Sentinel-1.

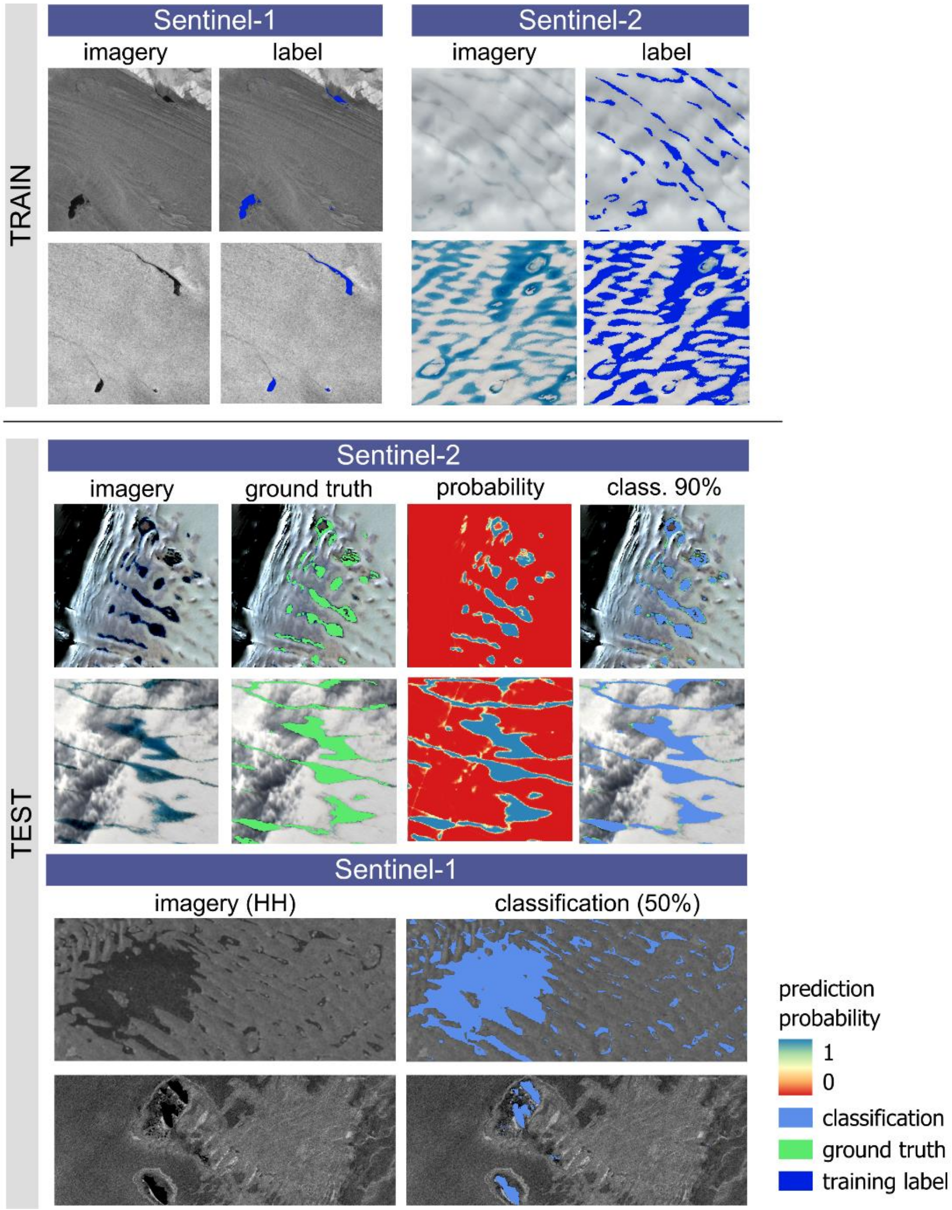


Figure S 1 The upper row shows the training labels marked in blue. Predictions were thresholded at 90% probability for Sentinel-2 and 50% for Sentinel-1, with thresholded predictions shown in pale blue. Ground truth labels for Sentinel-2 are shown in green. For Sentinel-1, testing was performed using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

Line 210: should the "observed in January 2017" be cited?

We clarify that this was done manually by adding L241 "we identified four additional ice shelves that experienced frequent ponding after 2017 (identified by manually screening optical imagery during January in the following years) and added them to the PolarLakes dataset…".

Line 215: can you specify what is "a sufficient amount"?

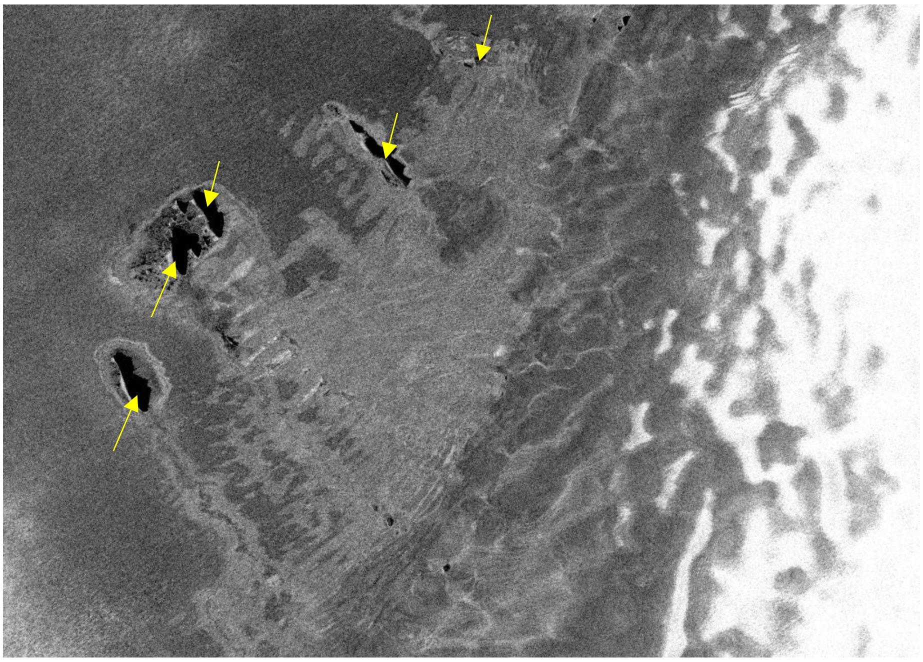
We now quantify it by saying L249 "The two-week interval turned out to include approx. two acquisitions per sensor of available Sentinel-1 and Sentinel-2 data to create a consistent data product with uniform time steps at 10 m spatial resolution."

Line 247: "covered extracted" – is that a typo?

Yes, removed “exctracted”.

Line 251: in the Sentinel-1 imagery, how do you know what is water and what is not water? From Figure 1 and other Sentinel-1 imagery I’ve seen, ponded water is clearly not always dark. I can see you used Sentinel-1 maximum extent maps for support, but presumably there is a time difference between those maps and the Sentinel-1 images, during which time cold snaps or snowfall could change the lake extent.

Indeed, identifying supraglacial lakes in Sentinel-1 imagery is challenging. To label the training/test data correctly, we cross-checked with optical imagery to ensure that similar-looking features (e.g., blue ice, shadow) were not misinterpreted as lakes. Open, wind-undisturbed water is distinctly darker than all other surface features, and we selected new example images in Fig. 1 specifically because this contrast is better visible. Only these darker areas are mapped (shown in dark blue in the biweekly merged product, and indicated by yellow arrows in the zoomed-in panel below). Areas that have already begun to refreeze are missed by S1 but still detected by S2, as shown in turquoise in Fig. 1. This is a clear caveat of the method, which is also why we state “Only a small amount can be contributed using Sentinel-1 (see **Fehler! Verweisquelle konnte nicht gefunden werden.** for the exact percentages of sensor contribution). An example of satellite sensor contribution to the classification result is given in **Fehler! Verweisquelle konnte nicht gefunden werden.** for George VI Ice Shelf highlighting the considerably larger fraction of detections based on Sentinel-2 data.” I hope this clarifies your questions.



Section 3.2: I struggled to understand the rationale behind this confidence flag. It seems to suggest that you don’t assign much confidence to the Sentinel-1 derived lake extents. That’s fair enough, but it doesn’t seem to tally with the results shown in Table 2 and Figure 5, which indicate that Sentinel-1 performs well.

We thank the reviewer for raising this point, as it highlights an important distinction that we should clarify further in the text. The strong performance of Sentinel-1 (S1) reported in Table 2 and Figure 5 and the confidence flag are not in conflict. They address two different sources of uncertainty.

S1's high performance holds specifically for open, wind-undisturbed water, which produces a distinctly low backscatter signal. Optically, this appears as a subset of the full lake extent. Slightly refrozen lake surfaces are still classified as supraglacial lakes in optical (S2) imagery, but in C-band SAR the onset of refreezing increases backscatter, causing the lake signature to disappear. As a result, S1 systematically underestimates lake extent relative to S2, even though its detection of open water is itself highly accurate.

Critically, this S1 behavior is consistent and predictable and it does not vary with external factors, so long as a scene is available. S2, by contrast, is strongly affected by cloud cover, which introduces variable, scene-dependent uncertainty in whether lakes are detected at all. Because this uncertainty is specific to S2 and not present in the same way for S1, we designed the confidence estimate to target S2 specifically, rather than applying it uniformly across both sensors.

We make this clearer now in the manuscript L341: "This metric is tailored specifically to Sentinel-2, as cloud cover introduces variable, scene-dependent uncertainty in whether lakes are detected at all. Sentinel-1, by contrast, is not subject to this constraint as its C-band SAR observations are available independently of cloud cover, and its detection performance is consistent whenever a scene is acquired. We therefore do not include Sentinel-1 in the confidence estimate, since its limitation lies not in cloud cover but in a distinct, systematic effect where the onset of surface refreezing increases backscatter and causes slightly refrozen lake surfaces to no longer be detected as open water, leading to a systematic underestimation of lake extent relative to Sentinel-2 rather than a source of scene-to-scene uncertainty. The confidence estimate therefore indicates the degree to which the mapped lake extent is supported by sufficient and unobstructed Sentinel-2 satellite observations".

Line 315: should that be "the annual **maximum** extent"?

Yes, corrected.

Line 335: should that be "intra-annual"?

Yes, corrected.

Line 337: typo? "much stronger"

Yes, corrected.

Line 350: I think the "research" you refer to here should be cited (presumably one of the references from the line above?)

We added references to make it clearer.

Line 375 and elsewhere: Re lakes forming "in depressions", consider adding "rather than crevasses", so pedants like me don't cleverly point out that all lakes must form in depressions whether they have smooth edges or not.

We added "rather than crevasses".

Line 384-385: I can't see a comparison between ponded area/locations and meltwater production here, so I'm not sure you can support this claim. An alternative explanation is that there was the same amount of melt but a lower capacity in the snowpack to accommodate the meltwater (perhaps because of reduced snowfall that year or because of year-on-year depletion of firn air content).

Thank you for adding another perspective. We re-phrased to "In certain years, ponding extends further across the flat ice-shelf surface. This potentially can be attributed either to higher meltwater production or to a reduced capacity of the snowpack to accommodate that meltwater driven by lower snowfall or by year-to-year depletion of firn air content."

Line 388: but isn't that the opposite to what you have found?

No, it isn't the opposite. But maybe we phrased this a bit too confusing. We improved the wording for clarity: "However, the timing of this recurrence differs between the AP and the EAIS. Ponding was more pronounced along the EAIS in the first half of the study period, up to 2020, and along the AP in the second half. This contrasts with the findings of Tuckett et al. (2025) who suggest that the EAIS may be becoming increasingly favourable to meltwater ponding. This discrepancy can be explained by the fact that their time series ends in 2021, shortly after the pan-Antarctic peak in 2020 and the subsequent decline in surface meltwater on the Riiser-Larsen Ice Shelf followed afterwards. The Larsen C Ice Shelf along the AP, in contrast, began to show extensive supraglacial lake formation following the 2020 pan-Antarctic maximum, reaching its peak in 2023."

Line 394-395: I couldn't follow this sentence. What is significantly smaller than the fused product? And how can the fused product increase the amount of detections from Sentinel-1, which is what this sentence seems to imply?

We re-phrased this part and hope it is now easier to understand L449: "Firstly, consistent time series analysis is only possible from the 2016–2017 melt season onwards, since Sentinel-2 data are not available before this period and the fused product relies on both Sentinel-1 and Sentinel-2. Secondly, Sentinel-1-only lake extents are substantially smaller than those from the fused product, since Sentinel-1 detects only open, undisturbed water rather than the full lake extent captured by combining both sensors. As a result, time series that span the pre- and post-2016–2017 period, or that mix single-sensor and fused-product extents, are not directly comparable. For consistency, we therefore recommend restricting time series analysis of the PolarLakes dataset to 2016–2017 onwards, or analysing the Sentinel-1 and Sentinel-2 records separately."

Figure 1: is the reader supposed to be able to read the legend in the U-Net Classification panels? I tried to zoom in but still couldn't at max zoom. Perhaps worth removing that or making a bigger popout of those panels? As it is, I don't think that panel provides any information.

Thank you for this comment. Figure 1 is intended purely as a high-level overview to illustrate the overall workflow, rather than a detailed architectural diagram meant to be read alongside its legend. Readers interested in the specific details of the U-Net architecture are referred to Koehler et al. (in review), where it is described in full.

Figure 1: the label "errors in regions that were only ponded any four weeks" is a little hard to parse – the main text description on lines 240-242 is much clearer (though note the typo – "be-weekly")

We removed "four weeks" and now also state in Fig. 1 "ponded by any two bi-weekly interval".

Figure 4b: on line 225-226, you state that only a small amount of lake extents are detected with Sentinel-1, but Figure 4 seems to suggest that Sentinel-1 often contributes significantly and sometimes wholly to the lake detection. Though all of the map images provided do indicate a low contribution from Sentinel-1. Please can you clarify this in the text?

We agree this might be misleading and explicitly explain this now L260: "In general, the largest lake extents are detected using Sentinel-2 imagery, provided cloud cover is low. Sentinel-1 contributes only a small proportion of the total mapped area (see **Fehler! Verweisquelle konnte nicht gefunden werden.** for exact sensor contribution percentages). Note that the sensor contribution shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**b is expressed as a percentage of the total detected area within each composite; consequently, a 100% Sentinel-1 contribution does not necessarily indicate a large detected lake area, but rather that no Sentinel-2 data were available for that composite, such that Sentinel-1 was the sole contributing sensor."

Figure 4b: there are quite a few occasions in this timeseries where there is a mix of contributions from Sentinel-1 and Sentinel-2, but with very little or no overlap. From the provided images (e.g. Figure 4a, Figure 6) I couldn't identify where these differences in mapping occur. Please can you provide an example? This could potentially go in the supplementary information.

We agree that from the provided figures it is hard to understand the sensor contributions. We added Figure S2 to highlight different fractions of both sensors contributing to the overall classification and added to the text: "The contribution of Sentinel-1 is high in areas of high cloud cover in Sentinel-2 imagery (see **Fehler! Verweisquelle konnte nicht gefunden werden.** a-c). In the first half in March 2021 the contribution of Sentinel-1 was very minor with a few lake pixels not visible in the imagery (see **Fehler! Verweisquelle konnte nicht gefunden werden.**d). For all dates with very high Sentinel-1 contribution a visual check revealed very small maximum lake extents where in Sentinel-1 where detected some small ponds not visible in cloudy Sentinel-2 imagery". We hope this makes it clearer for the reader.

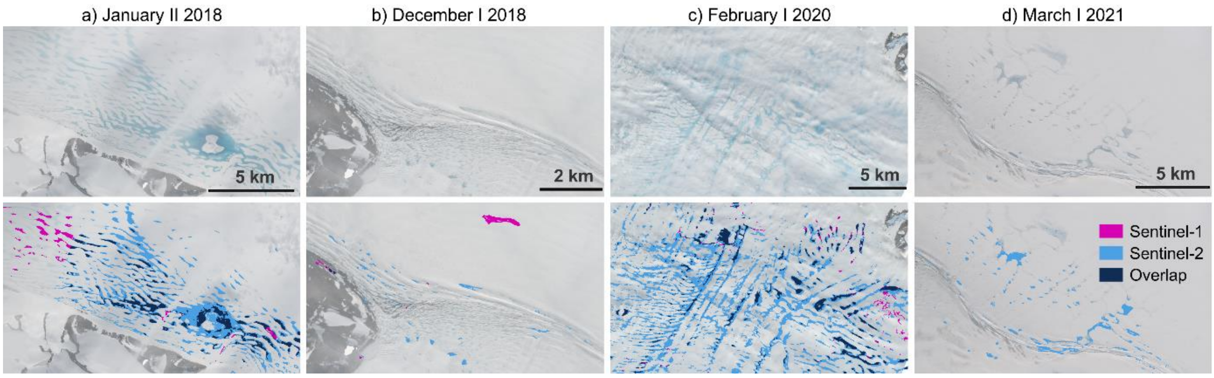


Figure S 2 Sentinel-1 and Sentinel-2 contributions to maximum lake extent classification on selected areas of George VI Ice Shelf for (a) late January 2018, (b) early December 2018, (c) early February 2020, and (d) early March 2021. Upper panels show the least cloudy Sentinel-2 image within each two-week period. Lower panels show the maximum lake classification with sensor contribution. The contribution of Sentinel-1 is high in areas of high cloud cover in Sentinel-2 imagery (panels a-c). In the first half in March 2021 the contribution of Sentinel-1 was very minor with a few lake pixels not visible in the imagery (see panel d). For all dates with very high Sentinel-1 contribution a visual check revealed very small maximum lake extents where in Sentinel-1 where detected some small ponds not visible in cloudy Sentinel-2 imagery. Contains Copernicus Sentinel-2 Data 2026

Figure 7: This is a nice figure but it isn't really used in the text. There is a passing reference in Section 5. Also, perhaps it's some kind of optical illusion but are there some dots (e.g. 2022-12) with a blue colour intermediate between "medium confidence" and "high confidence"?

Thank you for spotting the missing reference. We now included Figure 7 also in the text and state that the color bar is continuous. We tried different color combinations but it was challenging as red to green was not colorblind friendly. That's why we ended up with blue, turquoise to orange. The improved text reads now L356: "Users can access this information via the corresponding <shelf-name>_stats_with_confidence_flag.csv files or directly through the lake extent plots available for each shelf (see **Fehler! Verweisquelle konnte nicht gefunden werden.** as an example for George VI Ice Shelf with confidence estimates and maximum lake extent in km2 for every bi-weekly period), where confidence levels are visualized using a continuous color scheme with blue colors for high confidence and orange colors for low confidence."

Figure 10 and others like it: I find it very hard/impossible to distinguish between the flat colours used for "overlap" and "sentinel-2" and actual meltwater in the satellite image that has not been mapped. So when I look at Figure 10, I can't really see where the differences in mapping are. Are there other ways you could present this information? Perhaps just show the classifications in a separate panel to the satellite image? Or manually map the ponded water and graph/tabulate the differences to that?

We agree that the selected colorbar was too pale and made it difficult to distinguish between the imagery in the background and the mapped lakes. We changed the colorbar to pale blue and pink for a better separation in Figure 4, 6 and 10.

Table 2: Can you comment specifically on the reasons for the much lower F1 scores are Larsen C and Hull for Sentinel-2?

Yes, we were also interested in the drop of accuracy which can be explained by the lake morphology at Hull and the smaller dark clouds at Larsen-C. This is now also mentioned in the text at L310: "Furthermore, classification performance varies between test sites, reflecting differences in lake morphology and the presence of spectrally similar features (e.g., small, dark cloud cover misclassified in the Larsen C-1-2 region and very narrow and tiny lakes at Hull)."