***Summary***

*The paper describes a semantic segmentation scheme to map landlocked lakes in Antarctica, using Landsat and Sentinel-1 satellite imagery as base data. Landsat images are segmented with a U-net, Sentinel-1 with a manually tuned threshold. The results are merged with a simple late fusion logic.*

We would like to thank you for reviewing and commenting the manuscript. Below, we present a detailed response to each of your comments, with the original comments in italics and the responses in blue. All the recommended modifications will be implemented in the revised paper that will be uploaded.

***Novelty/Relevance***

*There isn't any technical novelty. Methods are standard and used in somewhat ad-hoc manner without clear justification for the design.*

We agree that our methods are standard which had been widely applied, such as the application of the U-Net model to detect supraglacial lakes in SAR images (Dirscherl et al., 2021b). We also used U-Net model for landlocked lake detection with Landsat images. However, surrounding information such as rock and ice is necessary for the identification of landlocked lakes, while this information is unavailable in previous studies. Thus, we attempted to identify the surrounding situation of lakes through multi-classification and breadth-first search (BFS). We generated potential landlocked lakes' open water (LLOW) areas as a water mask, and successfully transferred surrounding situations of lakes from Landsat images to Sentinel-1 images through the mask process.

*The specific application appears to be new, I am not aware of any paper that described the specific case of land-locked Antarctic lakes. That being said, the distinction is perhaps a tad contrived, there has certainly been work on detecting supra-glacial lakes, so the only difference is really to check whether a lake is surrounded by rock or by ice.*

We agree that the primary difference between our work and previous studies is the detection of whether a lake is surrounded by rock or by ice. But the work is challenging and time-consuming. Limited by the complexity of rock and ice backscatters, SAR images can't distinguish rock and ice. Due to the interference of clouds, multispectral images are not suitable for generating long time-series datasets. Only the combination of SAR images and multispectral images can solve this problem.

***Strengths***

*Since the task has apparently not been studied before, there is potential to systematically map land-locked lakes with the method (it is not done at any scale,* *though). I am not an expert in Antarctic ecology or climate and cannot judge the relevance of this, but it is a mapping capability that hadn't been investigated.*

*The proposed method works moderately well, even if the segmentation performance is not surprising or spectacular for the fairly straightforward task.*

We would like to thank you for the positive comments on the significance of this work. In fact, the threshold method didn't perform well on the lake detection task. However, the detection of water in the Antarctica is more difficult than in other areas. Interfered by freeze-thaw process, strong wind and incidence angle (Dirscherl et al., 2021a; Li et al., 2021), the backscatters of water become unstable, which usually change significantly.

***Weaknesses***

*Technical decisions seem somewhat arbitrary and ad-hoc. Not seriously wrong, but the described scheme is just "a way to do it", not a carefully designed and justified "best way to do it".*

Thanks for your suggestion on our workflow. After our exploration and experiment, each step in the workflow is necessary for completing this task. For example, we employed the upsampling method and adjusted the patch size, to better identify tiny landlocked lakes with a few pixels in images. When the number of valid Landsat images is insufficient to form a LLOW time series, we changed the results derived from Landsat images into masks. Then we combined them with Sentinel-1 images to improve temporal resolutions of the LLOW time series. In addition, we also changed the parameters in the U-Net model for the better model performance.

*The evaluation is rather weak, using only a few small areas, and even excluding some lakes that are clearly visible within the image tiles. The study does not go beyond the four small proof-of-concept regions, there are no large-scale, wall-to-wall results.*

We agree that some obvious lakes were excluded in our results. When identifying potential LLOW areas, these lakes are not surrounded by rocks in classification results. We will attempt to improve our model performance to avoid the misidentification. In figures 4,5, and 6, we only present parts of the results. Considering that some LLOW areas are relatively small, especially in LHs, we increased the map scale and didn't show the entire area. We will add new figures to display the wall-to-wall results in the revised manuscript.

*The model validation suggests that almost all the performance is due to Landsat, whereas Sentinel-1 does not offer much except the potential to densify in time - which however is not actually done, since the Landsat segmentation acts as a hard constraint: the algorithm does not appear to allow SAR to add lake pixels.*

We agree that the potential LLOW areas strictly constrain the potential of SAR images in the current algorithm. We will draw a buffer around the potential

LLOW areas (Wangchuk et al., 2019) to allow water to extend beyond previous masks.

*Finding of a "decreasing trend in LLOW area" for 2017-2021 is rather trivial and expected. It would be more interesting to interpret the measured areas beyond just that obvious trend.*

We thank you for the insightful comment. The reliance on available satellite imagery inherently limits the scope of our trend analysis. Despite our efforts to utilize the most comprehensive and up-to-date datasets, the frequency and resolution of available images restricted our ability to identify and interpret more complex trends in LLOW dynamics throughout the study period. In particular, the temporal resolution of the imagery posed challenges in capturing short-term fluctuations or subtle changes in the LLOW area, which could potentially reveal more about the underlying processes affecting these water bodies.

***Presentation***

*Throughout, the text could be made shorter and more concise. E.g.,*

- *lines 215-220 are unnecessary, everything that is said there is already implied by the use of U-net*

We will delete this section of text.

- *228-240 is a verbose, meandering way to simply say “we manually chose a global threshold by inspecting histograms”*

We will shorten the description of threshold selection process.

- *250-264 says little more than that the definition of a land-locked lake is a water region surrounded by a rock region.*
- *etc.*

In this paragraph, we describe the process of BFS in detail and explain how a land-locked lake is identified. We will shorten this paragraph and simplify other wordy sections of the paper. Thanks for your suggestion.

*The introduction is verbose and not very focussed, touching on all sorts of studies about land-locked lakes that have no relation or importance for what the paper then does.*

Thank you very much for your comment on the introduction part. We will shorten the introduction, and focus on the lake identification.

*The analysis in lines 420-440 is rather hand-wavy, I was not able to see what purpose it actually serves. It gives me the impression that the authors just performed a random analysis that was easily doable, to send a message that the maps could potentially serve some useful purpose.*

We appreciate your comments. The purpose of this section was to elucidate the relationship between Positive Degree Days (PDD) and the area change of landlocked lake open water (LLOW) in different areas. We recognize the reviewer's concern that the analysis may appear cursory or lacking in clear purpose. To address this, we offer the following clarifications and enhancements to underscore the relevance and rigor of our analysis:

The primary purpose of the analysis was to demonstrate the influence of PDD on changes in LLOW areas with time. By showing a strong correlation (average $R^2$ value of approximately 0.9) during the growth phase of LLOW areas, we aimed to validate PDD as a key factor explaining most of the variability in LLOW area.

We detailed an exception in 2018 for LHs during an unusual cooling event to show that while PDD is a strong indicator of LLOW area changes, other climatic and geographic factors can also influence the outcomes. This exception was included to prevent oversimplification and to encourage a nuanced understanding of LLOW area dynamics.

*There are remaining language issues, both in terms of English grammar (random example: "due to non-uniform of field surveys") and in terms of technical expressions (e.g., "gradient disappearance" instead of "gradient vanishing").*

We appreciate the reviewer's comments.

We will thoroughly check the whole manuscript. Also, we will revise the points raised by you, following your suggestions.

***Technical Questions***

*The computational procedure is not entirely clear. It is one way of cobbling together a segmentation pipeline, but there is no clear explanation why that specific design was chosen. Whereas there are obvious concerns about it, e.g.,*

- *the potential benefit of Sentinel-1 for the land-cover map are not exploited*

The mask process between Sentinel-1 and land-cover map can be modified using the buffer method mentioned above. We will revise the workflow and results in the revised manuscript.

- *possible correlations between optical and SAR are lost*

Ensuring that optical and SAR image dates are close to each other is challenging in identification of LLOW in Antarctica. Thus, we aggregate all Landsat images into one potential LLOW area instead of the fusion of each Landsat and Sentinel-1 image.

- *the fusion seems to not leverage away the (pseudo-)probabilistic segmentation scores*

Our mask method follows the priority principle for potential LLOW areas identification. Any water pixel not within the potential LLOW areas is not

LLOW. We didn't consider probabilistic scores in the fusion process. The combination of buffered masks and probabilistic scores might improve our model and allow SAR images to add pixels. We will try this method and test the performance in the revised version.

*What is meant by saying 300x300 is the "common" patch size for U-net models? There isn't a single, canonical patch size for training those models, and at test time they are anyways fully convolutional and not tied to a specific patch size.*

This is an inaccurate description. We will correct it. We had used this parameter and the word "common" because we referred to previous publications and code. In addition, we agree that fully convolutional networks can deal with patches with various sizes. However, when we standardized the size of training patches, it did perform relatively well compared with the method without resampling.

*I don't understand the upsampling of the input for the U-net. No information is added by this and the effective receptive field / context window inside the network is actually reduced. So it would seem that one can reach at least equal performance, with lower computational effort, by properly training the U-net to handle the smaller images. Please explain.*

The 1024*1024 patches consume more computational resources than 300*300 patches during the U-Net inference. However, as described in the paper, the upsampling is used to magnify the small open water area. Some open water areas only consist of a few pixels in the images. Magnifying these areas through upsampling contributed to the identification of small LLOW.

*Certain augmentations should be ablated and empirically justified. Conceptually it seems problematic to apply transformations like rotating or vertical flipping, as this leads to illumination directions that are implausible in real Landsat images, especially at Antarctic altitudes with low sun elevation.*

Thanks for your suggestion. We didn’t consider the impact of image augmentations on the direction of illumination and shadows. Flipping and rotation disrupt the regular shadow features. These shadow features might benefit the training process. We will try a new augmentation strategy and compare the model performance.

*Using a threshold for segmentation is of course entirely correct and sensible, if it works. But the justification that single-channel input will lead to “instability” of U-net makes no sense. Countless applications use U-nets with various single-channel inputs (SAR, panchromatic, depth,...).*

Thanks for this comment. This is an inaccurate description. We will correct it. We agree that many studies had used U-Net models with single-channel inputs. But in our process of training U-Net model for SAR images, the U-Net produced several instable results. For example, in snow-cover areas, the snow with relatively low backscatters were identified as water by mistake.

*The fusion step is unclear. First you argue for using SAR, and for combining it with optical data, to obtain better temporal resolution. But then, a consensus over at least 2 Landsat acquisitions is required for a potential LLOW pixel, meaning that the shortest possible resolution of everything that follows is the interval between three Landsat overpasses (if a pixel flips from ice to water between two consecutive images, you need to wait for a third image to confirm, so over the entire period you cannot say whether the pixel remained the same or thawed and froze again).*

We used all land-cover maps during five years to generate one potential LLOW area map for one region. Among several dozen images, if a pixel is identified as LLOW in at least two images, then that pixel is considered a potential LLOW. This algorithm aims to preserve as much potential space as possible.

*In line 290, it remains unclear how the authors "disregard" underestimated lakes. To do that one must identify them first - but the algorithm, by definition, does not know where it made an underestimation error.*

We manually deleted those images which underestimated the LLOW. In the melting phase, the backscatters of water may suddenly increase due to strong wind or other reasons (Fig. 1). We will check the images where the lake areas suddenly decrease in the time series curve. If backscatters of water increase abnormally, we will manually remove these images.

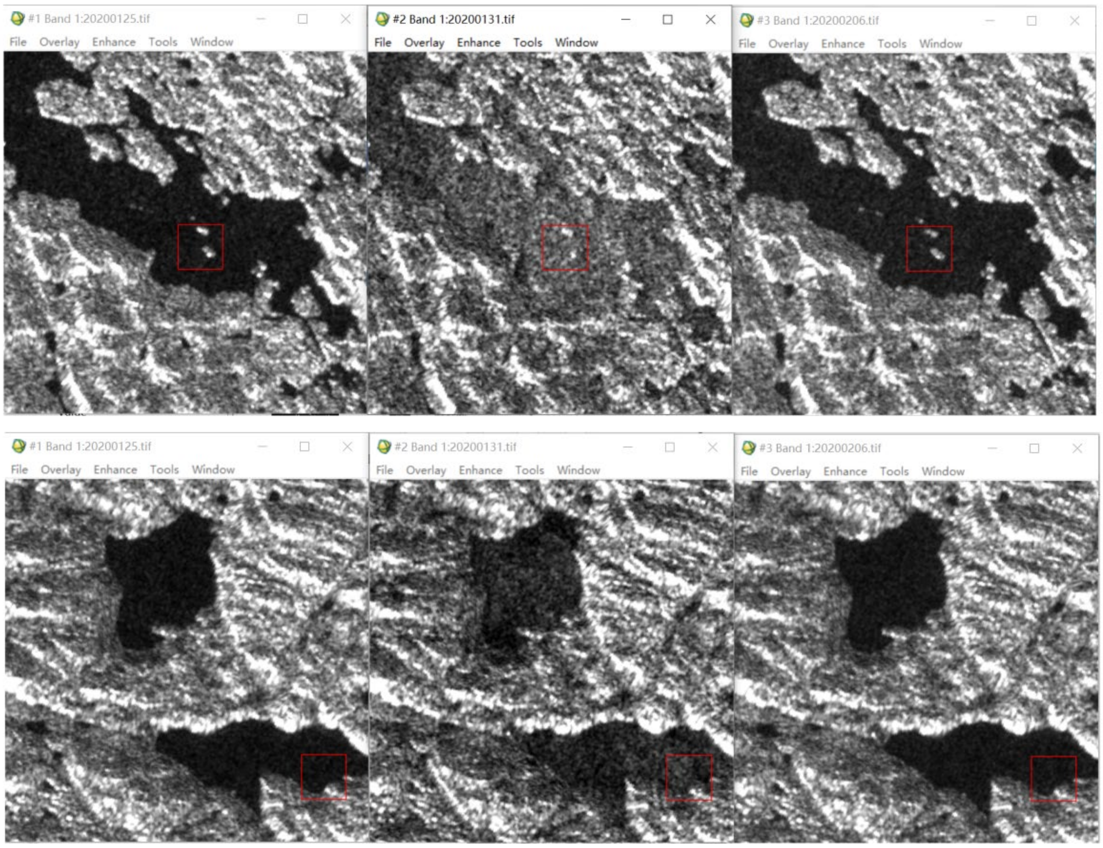


Figure 1. The significant backscattering changes of lakes during 25 January, 31 January and 6 February in the VHs.

*I do not understand why only a tiny set of 17k pixels were "annotated for U-net", but 225k pixels were "annotated for LLOW identification". 17k seems an overly small training set: assuming an average lake size of, say, 600x600m that would be fewer than 50 lakes. Why would one do that if apparently another 225k annotated pixels are available?*

The labels annotated for U-Net and LLOW identification are different. The labels for U-Net consist of water, rock and ice, while the labels for LLOW only consist of LLOW and non-LLOW. Thus, we can't apply LLOW data to U-Net testing.

*Cohen's kappa as a segmentation metric is discouraged (cf. [Pontius and Millones, 2011]). I would recommend to follow best practice and show confusion matrices, F1 scores, IoU scores.*

Thanks for your suggestion. We will add new accuracy validation experiments on F1 scores and IoU scores.

*The PDD metric (equation 4) is defined in a strange manner. According to the definition that metric is exactly the same for a cold spell with two weeks of constantly at 0 degrees, or of constantly -25 degrees. Surely that would make a difference for the ice cover of the lakes? Wouldn't it be more natural to look at the number of consecutive PDD, or to integrate the average temperature including negative values?*

Thank you for the insightful comments and suggestions. The use of the PDD metric in our study is grounded in its common application within glaciological and climatological research as a simplified measure to estimate the melting potential of ice and snow. The rationale for focusing exclusively on temperatures above 0°C is to directly quantify the thermal energy available for melting ice, which is a critical factor in the dynamics of lake ice cover.

However, you raised an important point about the continuity of temperature conditions during the melting phase. It is essential to highlight that temperature patterns preceding the accumulation of positive degree-days (PDDs) typically show a gradual increase towards 0°C. This pattern means that, even before reaching the threshold for PDD calculation, the thermal conditions are already near the melting point. Thus, a cold spell with two weeks of constantly at -25 degrees during the melting phase would be unusual.

To assess the influence of higher temperatures more accurately on the expansion of open water areas, we utilized the average daily temperature, inclusive of negative values, to calculate PDDs. This method acknowledged that any day with an average temperature above 0°C contributes to the melting process, providing a nuanced measure of thermal energy input relevant to ice melt. However, days with an average temperature below 0°C were considered to have a negligible impact on ice melting and were thus not included in our PDD calculations.

*For the decline (Section 5.2), why use the minimum temperature? To my knowledge, and also more in line with the PDD metric used earlier, a more* *common indicator is the number of consecutive negative degree days, at least in studies of lake ice in Canada, the Alps, etc.*

We appreciate your comment. In our study, we opted to use the average temperature rather than the minimum temperature for analyzing the decline phase of lake ice cover. The mention of minimum temperature in our text relates specifically to the context of defining the range for our correlation analysis between LLOW area and temperature metrics. This was articulated in the manuscript to delineate the bounds for calculating the $R^2$ value based on a linear fit, aiming to refine the focus of our correlation analysis.

We acknowledge the common practice of using consecutive negative degree days (NDDs) in lake ice studies across regions such as Canada and the Alps. This practice serves as a direct measure of the cold spell's duration and intensity, which influences ice formation and maintenance.

The choice to exclude NDDs as the main measure in this stage of our research is based on previous analysis results, which showed that although NDDs play a significant role in initiating ice cover formation, their connection to the shrinking size of LLOW area does not follow a linear pattern (Graf and Tomczyk, 2018). Our observations suggest that, once the lake ice cover is developed, its dynamics are more closely correlated with average temperature variations.

***Minor Comments***

*Why do ice-covered lakes "magnify the warming trend"? I would think they might rather dampen it?*

We appreciate the reviewer. Ice and snow typically have high albedo, reflecting a significant portion of incoming solar radiation, thus potentially having a cooling effect on the local environment. However, the context of our discussion focuses on the transition period from ice-covered to ice-free conditions and the consequent alterations in surface albedo and energy absorption.

The mechanism by which ice-covered lakes can amplify warming trends is linked to the concept of the ice-albedo feedback. During the winter and early spring, ice-covered lakes indeed reflect a large portion of solar radiation due to their high albedo. However, as temperatures rise, ice begins to melt, reducing the surface area covered by ice and exposing the underlying water. Liquid water has a significantly lower albedo than ice, indicating that the lake absorbs more solar radiation instead of reflecting it. This increased absorption of solar energy by the water surface not only contributes to further warming and melting of the remaining ice but also leads to a rise in water temperature, which can enhance local warming. This mechanism aligns with observations and modeling studies that have documented the significant role of ice-albedo feedback in accelerating warming, particularly in polar and high-latitude regions where ice and snow cover are integral to the climate system.

*It is a strange claim that U-net "requires less training datasets and time" than other neural networks. That depends on who you compare to, of course there are designs that are faster than U-net (e.g., those created for mobile or embedded devices). Moreover, "U-net" isn't a specific architecture but a whole family of networks with certain characteristics - essentially, symmetric hourglass encoder-decoder structure with dense skip connections. So some "U-nets" are a lot slower and more data-hungry than others.*

Thanks for your suggestion. We will revise this description following your comments. The U-Net model was originally designed to work with fewer training datasets for biomedical image segmentation, but the landscape has evolved considerably since then. Many new architectures have emerged, designed for accelerated training and optimized for smaller scales. In addition, parameters of U-Net also determine the computational cost.

Reference

Dirscherl, M., Dietz, A. J., Kneisel, C., and Kuenzer, C.: A Novel Method for Automated Supraglacial Lake Mapping in Antarctica Using Sentinel-1 SAR Imagery and Deep Learning, Remote Sens., 13,https://doi.org/10.3390/rs13020197, 2021a.

Dirscherl, M. C., Dietz, A. J., and Kuenzer, C.: Seasonal evolution of Antarctic supraglacial lakes in 2015–2021 and links to environmental controls, The Cryosphere, 15, 5205-5226,https://doi.org/10.5194/tc-15-5205-2021, 2021b.

Li, W., Lhermitte, S., and López-Dekker, P.: The potential of synthetic aperture radar interferometry for assessing meltwater lake dynamics on Antarctic ice shelves, The Cryosphere, 15, 5309-5322,https://doi.org/10.5194/tc-15-5309-2021, 2021.

Wangchuk, S., Bolch, T., and Zawadzki, J.: Towards automated mapping and monitoring of potentially dangerous glacial lakes in Bhutan Himalaya using Sentinel-1 Synthetic Aperture Radar data, Int. J. Remote Sens., 40, 4642-4667,https://doi.org/10.1080/01431161.2019.1569789, 2019.