Response to RC3

A Framework for Automated Supraglacial Lake Detection and Depth Retrieval in ICESat-2 Photon Data Across the Greenland and Antarctic Ice Sheets

Philipp Sebastian Arndt[1] and Helen Amanda Fricker[1]

[1] Scripps Polar Center, University of California San Diego, 8885 Biological Grade, La Jolla, CA 92037, USA

Correspondence:
Philipp Sebastian Arndt (parndt@ucsd.edu)

Discussion: https://doi.org/10.5194/egusphere-2024-1156

Comments from the reviewers are given in black.
Our responses are given in red.
Quotes from the submitted manuscript are given in bold red.
Proposed amendments or additions to the revised manuscript are given in blue in the Times New Roman font.

References that were already included in the original manuscript are cited in-text only, in the same format as in the submitted manuscript. New references are added to the end of this document in full.

Referee Comment 3 (Sammie Buzzard)

RC3: 'Comment on egusphere-2024-1156', Sammie Buzzard, 08 Jul 2024
Citation: https://doi.org/10.5194/egusphere-2024-1156-RC3

Summary

This paper presents two algorithms that together allow the retrieval of supraglacial lake depths using ICESat-2. This method is scalable beyond the case studies presented and therefore a useful contribution to the community that I would strongly recommend for publication.
Understanding surface melt, and lake depths is key for predicting ice shelf stability, and datasets of lake depths are limited. In situ data is scarce and remote sensing methods are necessary for providing validation and calibration for models as well as understanding the development of these lakes so this work will clearly be of use to the community.
While there are methods existing to determine lake depths remotely e.g. using Landsat-8, this to my knowledge is one of only a small number using ICESat-2, the advantage of this particular method appears to be scalability (although see my comments below on this).
The methodology and results presented are in my opinion sufficient for publication, and my recommendations for changes to the manuscript are mostly minor.

Thank you so much Sammie for taking the time to read and review our manuscript, and for your positive, thoughtful and constructive comments. In particular, we believe that your comments suggesting more detailed comparisons to similar studies as well as better explanations of the limitations of our method have helped to significantly improve the manuscript.

General Comments

1. It would be good to clarify the limitations of the algorithm e.g. max/min lake widths/ lengths/ depths detected clearly early on in the paper (and state how this compares to other methods).
   Based on this and some of the other referee comments we realize that we have not made it clear enough early on in our writing that ICESat-2 data coverage is sparse. The most important limitation of the algorithm that readers need to understand is that due to its coarse track spacing and 3-month return period ICESat-2 cannot sample all supraglacial lakes on the ice sheets – or even close to all of them. ICESat-2's unique selling point is its ability to accurately and directly measure water depths from space, but ICESat-2 does not provide a comprehensive map of the locations and spatial extents of melt lakes (plus, there are image-based methods out there that are much better-suited).

   - We propose to rewrite part of the introduction, to make it clear that the fundamental limitation of this method is that a data set based on ICESat-2 alone cannot find all the supraglacial lakes, and that we do not intend to claim that it can: "While ICESat-2 has the unique capability to make direct and accurate measurements of water depth from space, its fundamental limitation is spatial coverage. ICESat-2 data are limited to discrete, one-dimensional ground tracks that are coarsely spaced on the Earth's surface (~9.9 km between neighboring reference tracks and ~3.3 km between all neighboring beam pair tracks at 70°N/S) with a relatively long revisit time of three months. This means that no supraglacial lake depth data product derived from ICESat-2 alone is able to provide samples of all (or even nearly all) supraglacial lakes on the ice sheets: ICESat-2's track spacing means that the majority of lakes form in locations that ICESat-2 ground tracks never sample, and the three-month return period means that for a significant number of ground tracks ICESat-2 never passes over at the time at which melt lakes are visible. ICESat-2 is also unable to penetrate optically thick clouds, thus further limiting the amount of data available for water depth measurements. While ICESat-2 data alone cannot be used to continuously monitor melt lake volumes,..."
   - We will further emphasize track spacing in Sect. 3.5 (Study regions and time span) by adding the following: "Our two study areas cover latitudes from 68.2°N to 72.1°N in Greenland and latitudes from 68.4°S to 74.0°S in Antarctica, meaning that ICESat-2 track spacing is similar over the two regions: in Central West Greenland RGT spacing varies from ~8.8 km in the north to ~10.8 km in the south; over the Amery Catchment RGT spacing varies from ~7.9 km in the south to ~10.7 km in the north."

   The other limitations mentioned here ("max/min lake widths/ lengths/ depths"):
   - widths/lengths are the same in ICESat-2, since the ground track is one-dimensional
   - <u>max depths:</u> This is the limitation that everyone seems to be interested in. It is controlled by how far ICESat-2's green light can penetrate the water column under ideal conditions (very clear water, high bottom reflectivity). To clarify this early on, we will add the following to the introduction (line 41):
     "This allows ICESat-2 to measure water depths up to 41 m under ideal conditions (very clear water and high bottom reflectivity), with typical accuracies of about 0.5 m (Dietrich et al., 2024)."
   - <u>Min depths:</u> This is indirectly answered by the above. Since typical accuracies are around 0.5 m, it is difficult to detect lakes shallower than that because the lakebed return is hardly distinguishable from the lake surface return. As you noted, FLUID can not detect any ATL03 lake segments that are less than 0.35 m below the lake surface at their deepest along-track location (which is about 26 cm in refraction-corrected water depth), since all photons from the bottom return would be removed with the surface. However, in practice, we have not encountered any examples where a lakebed <0.35 m would be discernible in the photon data. Furthermore, such lake segments are often short and tend to be in locations with relatively flat surrounding surface topography, which makes it also harder to detect them in the first place. We

will make a short note on this in the section where this is mentioned (3.3.3 SuRRF step 2: fitting the lakebed, line 405):

"Note that this imposes a theoretical minimum depth threshold for detection on lake segments: ATL03 segments need to exhibit a bottom return signal at least 0.35 m below the lake surface (or 0.26 m in refraction-corrected water depth) at their deepest along-track point to be considered by SuRRF. However, in practice, such shallow lake segments do not have a discernible bathymetric signal since typical depth retrieval accuracies for ICESat-2 are on the order of 0.5 m (Dietrich et al., 2024)."

- Min lengths: Since FLUID assesses the flatness of major frames (~ 140 m in length), lakes with open-water surfaces of a shorter extent in along-track distance are not guaranteed to be detected. However, in practice, significantly shorter lake segments are regularly detected by FLUID. This is likely because the return signal from flat water surfaces tends to be much stronger than the return signal from surrounding ice surfaces, which makes even short, flat water surfaces dominate the overall distribution of photon elevations within a major frame. We will make a note about this in the relevant section (3.2.1 FLUID step 1: identification of flat water surfaces, line 210):

  "Since FLUID assesses the flatness of the surface of full major frames that cover an along-track distance of ~ 140 m, lake segments with shorter open-water surfaces are not guaranteed to be detected by FLUID. However, lake segments that are significantly shorter than 140 m are regularly detected by FLUID in practice. This is because the return signal from flat water surfaces is typically much stronger than the return signal from the surrounding ice surfaces, which makes even very short, flat water surfaces dominate the overall distribution of photon elevations within a major frame."
- Max lengths: There is no limitation to this. (While we have not detected any supraglacial lake segments with a water extent larger than 10 km, this is not a limitation of our algorithms. Supraglacial lakes just typically aren't that large. An exception is a large melt lake that sometimes forms on Amery Ice Shelf and can get up to 70 km long, but ICESat-2 ground tracks do not cross this lake at the right orientation for lake segments to become this long.)
- Comparison with other methods: Since we explain that in practice these limitations arise from the ICESat-2 data themselves, there is no reason to compare these specific limitations with regard to other ICESat-2-based methods.

2. This method appears to only be scalable if you live in the US based on the information in the paper. Some comments on this would be useful e.g. can your methodology transfer to a reader's local (i.e. institutional) supercomputer or does it need to be a National level facility? How possible would it be to do this?

   Our Python code works without the need for using any OSG services: it can be run on a local computer or on any computing cluster (locally or in the cloud) in the provided singularity container. We used the OSG Open Science Pool because it provided us with free computational infrastructure. We do not have the resources to scale this method up to ice-sheet-wide implementation on commercial cloud computing platforms at a feasible price point, and during initial development PA had no access to NASA-provided computational resources due to the COVID19 pandemic preventing in-person fingerprinting of non-US citizens. We believe that the OSG Open Science Pool is the most accessible option for large-scale implementation, but there are unfortunately some limitations to who can use it. As a non US-based researcher, one could gain access to the OSG Open Science Pool if collaborating with a US-based researcher or an institution that operates its own access point. Since we want our methods and data to be accessible and encourage others to use it in their own research, we will explain this in more detail in the Code and data availability section:

   "The FLUID/SuRRF source code is freely available at https://doi.org/10.5281/zenodo.10905941 (Arndt and Fricker, 2024a). To execute this code, users need to create a free NASA Earthdata login for ICESat-2 data access.

The source code contains a singularity container in which this version of FLUID/SuRRF can be executed. The main Python script detect_lakes.py can be run either locally on any individual ATL03 granule, or on many granules in parallel on any computing cluster that supports the specified computing environment or the use of singularity containers. In this study we present our implementation of FLUID/SuRRF on the OSG Open Science Pool because it provided us with free computational infrastructure. Due to funding mandates, free access to the OSG Open Science Pool is limited to researchers contributing to a US-based project at an academic, government, or non-profit organization, or researchers affiliated with any project or institution that operates its own local access point. This means that to implement FLUID/SuRRF on the OSG Open Science Pool as described here, you need to have at least one collaborator on your team to whom the above criteria apply. This collaborator can register your project with OSG on the Open Science Pool. Then, anyone contributing to the project can register for an account on OSG Connect to gain access to the Open Science Pool. For more information, see https://osg-htc.org/services/open_science_pool.html and https://osg-htc.org/about/organization. More information is also included in the README file."

Detailed Comments

- Line 44: It would be good to have a more detailed comparison with the Datta and Wouters and Leeuwen methodologies to explain what the differences are here. Given the method presented here is promoted as scalable would it be possible to do a direct comparison to their results?
  Datta and Wouters (2021) and Leeuwen (2023) demonstrate that ICESat-2 data can be used to improve depth extraction from imagery, which we use to motivate our study. To clarify this, and based on other referee's comments, we are adding the following to the introduction:
  "While ICESat-2 data alone cannot be used to continuously monitor melt lake volumes, several case studies have shown that ICESat-2 depth measurements can be used to constrain parameters in models that estimate lake volumes from satellite imagery (Datta and Wouters, 2021; Leeuwen, 2023; Lutz et al., 2024). For instance, Datta and Wouters demonstrated that it is possible to accurately extrapolate depths along ICESat-2's ground track segments to the full lake basins that these segments intersect. To be able to use ICESat-2 to improve depth estimates of supraglacial lakes in locations where (and at times when) ICESat-2 measurements are not directly available, it will be necessary to rely on statistical methods that can generalize the relationship between water depth and reflectance for a particular passive optical sensor under a wide variety of conditions and independently of the availability training data that is close-by in space and time (Hastie et al., 2009). For this to work effectively, the data that is used to train statistical learning models capable of multiple non-linear regression for representing a complex depth-reflectance relationship need to adequately cover the parameter space defined by the combination of predictors that are included (Markham and Rakes, 1998; Wang et al., 2022). Since ICESat-2 observations of melt lakes are relatively sparse, it is therefore crucial to to obtain as many ICESat-2 depth estimates as possible from different locations and times (and thus under a wide variety of environmental conditions) to be able to effectively use ICESat-2 to improve monitoring of meltwater volumes across the ice sheets. This suggests that large-scale extraction of accurate supraglacial lake depths from a wide range of ICESat-2 photon-level data in combination with concurrent optical satellite imagery can provide a labeled training data set enabling the application of machine learning methods (e.g., Leeuwen, 2023) capable of generating a well-constrained data-driven model for ice-sheet-wide lake volume estimation (Melling et al., 2023)."

  However, neither study published their depth estimates (Datta and Wouters published Matlab code, Leeuwen did not publish any data or code with their Master's thesis), so there is no easy way for us to compare results. Fricker et al. (2021) is a great resource, which compares different algorithms for supraglacial lake depth estimates from ICESat-2 including four estimates from Watta and an earlier version of SuRRF. We will point the reader to this paper for depth estimate comparisons, and also include the Watta results for the lakes shown in Fricker et al. (2021) in Fig. 12 (see below).

We will add the following comparison between SuRRF and alternative methods to the end of section 4.2.2 (line 612):

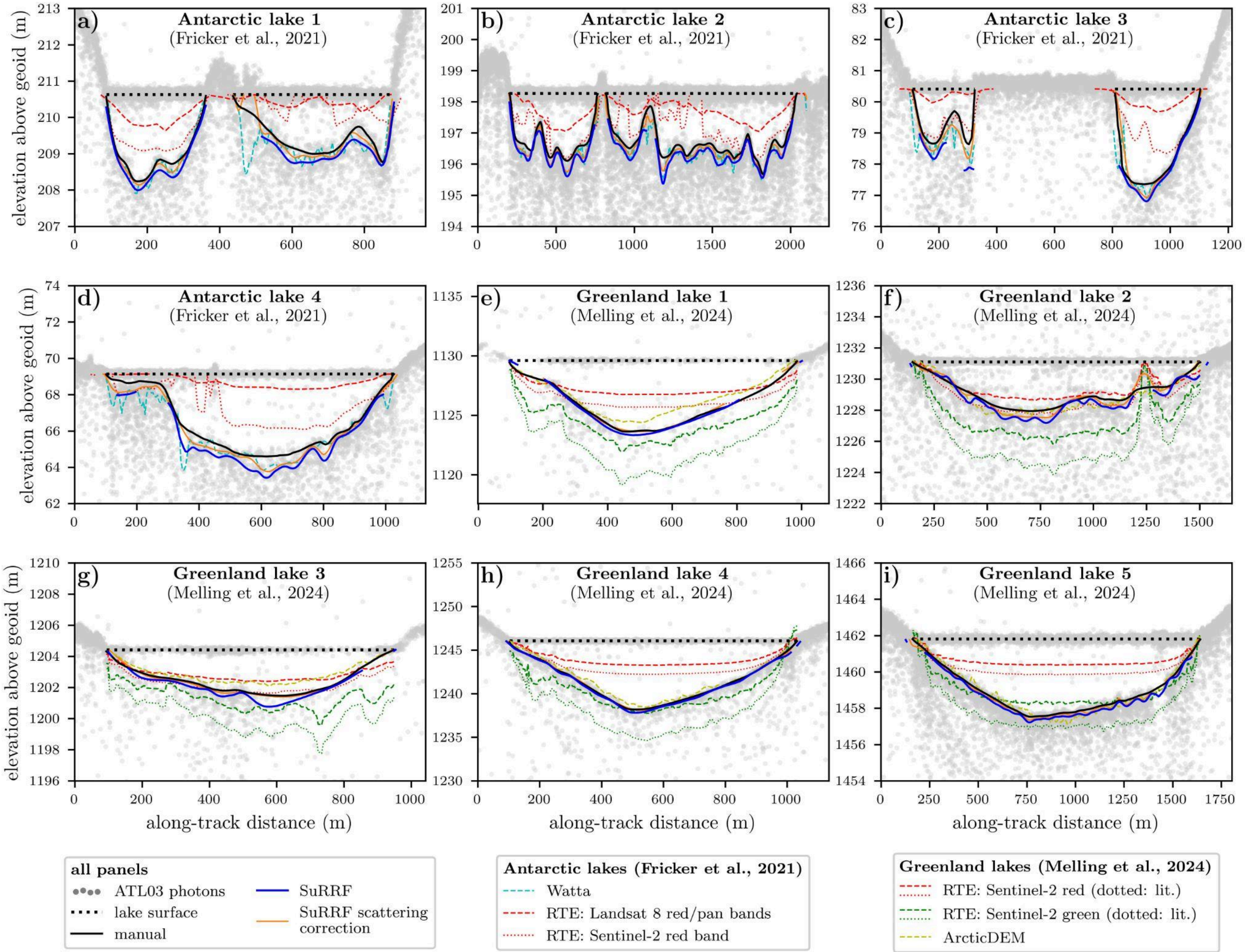


Figure 12. Comparison between SuRRF water depth estimates, manually annotated ICESat-2 depths and results from other methods for meltwater depth estimation. a)-d): ICESat-2 melt lake segments on the Amery Ice Shelf with manual annotations from Fricker et al. (2021). Other meltwater depth estimates that were reported by Fricker et al. are shown for the Watta algorithm (based on ICESat-2; Datta and Wouters, 2021), for the RTE method applied to the average of Landsat 8's red and panchromatic bands (Spergel et al., 2021) and to Sentinel-2's red band (Moussavi et al., 2020). e)-i): ICESat-2 melt lake segments in Southwest Greenland with manual annotations from Melling et al. (2024). Other meltwater depth estimates that were reported by Melling et al. are shown for the RTE method individually applied to both Sentinel-2's red and green bands, as well as estimates from post-drainage lakebed topography in ArcticDEM (based on Bowling et al., 2019).

“In addition to SuRRF and manual depth estimates, panels a)-d) of Fig. 12 show depth estimates for Antarctic lakes that were reported in Fricker et al. (2021), for the Watta algorithm (based on ICESat-2; Datta and Wouters, 2021) and for the RTE method applied to the average of Landsat 8’s red and panchromatic bands (Spergel et al., 2021) and to Sentinel-2’s red band (Moussavi et al., 2020). Both SuRRF and Watta track the general shape of lakebed returns in the ATL03 photon clouds well (Pearson’s correlation coefficients of R = 0.99 and 0.94, respectively), and and largely agree with manually determined along-track water depths (MAEs of 0.29 m and 0.30 m, respectively). In contrast to SuRRF, Watta appears to have a tendency to overfit where photon density near the lakebed is high with a large elevation spread, resulting in an unreasonably “wiggly” lakebed fit (e.g., Antarctic lake 1, 500-800 m). SuRRF's smoother fit under these conditions is likely due to the fact that it utilizes an adaptive kernel for its robust fit, whose width increases as the number of photons that narrowly cluster around the previous iteration's fit decreases (Sect. 3.3.1). In contrast to SuRRF, Watta also attempts to fit the lakebed across the entire lake basin, even where the lakebed is not visible or indistinguishable from noise, which can sometimes result in arbitrary, unrealistic depth estimates (e.g., Antarctic lake 1, around 450 m). Under such conditions SuRRF assigns a low confidence score to the lakebed fit and discards associated depth estimates to prevent arbitrary results. However, in some cases this results in SuRRF discarding depth estimates where Watta appears to fit the lakebed reasonably well (e.g., Antarctic lake 3, 250-300 m). The RTE approach based on Landsat 8's red/panchromatic band average consistently underestimates water depths, reporting a total amount of water that is ~ 73 % lower than the manual baseline. The RTE approach based on Sentinel-2's red band also underestimates water depths, and reports a total amount of water that is 34 % lower than the manual baseline.

Panels e)-i) of Fig. 12 also show non-ICESat-2 depth estimates for Greenland lakes that were reported in Melling et al. (2024), for the RTE method individually applied to both Sentinel-2's red and green bands, as well as estimates from post-drainage lakebed topography in ArcticDEM (based on Bowling et al. 2019). The RTE approach based on Sentinel-2's red band generally underestimates depths and reports a total amount of water that is 42 % lower than the manual baseline (similar to this method's performance over Antarctic lakes). In contrast, the RTE approach based on Sentinel-2's green band generally overestimates depths and reports a total amount of water that is 34 % larger than the manual baseline. However, Melling et al. also note that when using values of tuneable parameters that have been commonly used in the literature (Sneed and Hamilton, 2007; Georgiou et al., 2009; Pope et al., 2016), the RTE approach for Sentinel-2's green band overestimates lake depths even more, which results in reporting a total amount of water that is 84 % larger the manual baseline, with individual depths being overestimated by up to 153 %. This implies that RTE-based methods, while being popular for their simplicity, can potentially result in highly inaccurate meltwater volume estimates. The depth estimates derived from DEMs of emptied lake basins match the ICESat-2 manual baseline reasonably well, and when compared with it underestimate the total amount of water by 6 % with a MAE of 0.34 m. Since this method's performance is comparable to that of ICESat-2-based methods, this implies that DEM-based methods could be used supplement ICESat-2 depth measurements for labeling reflectance in passive optical imagery with supraglacial water depths, at least on the Greenland Ice Sheet where melt lakes on grounded ice drain regularly (Johansson et al., 2013)."

- Line 57: Most of the Greenland ablation zone rather than most of Greenland?
  We will include this, as suggested: "Across most of the Greenland Ice Sheet's ablation zone, ..."

- Line 60: There are enough notable examples of melt away from the grounding zone (over the ice shelves of Larsen B, George VI) or on grounded ice (e.g. Corr et al. found more than a quarter or meltwater features on grounded ice https://essd.copernicus.org/articles/14/209/2022/) that maybe 'mostly' could be changed?
  Thanks for pointing out ice shelves that have meltwater features away from the grounding zone (such as Larsen B and George VI). We will change our statement to include these examples. Besides this, we believe that our statement is quite accurate. While Corr et al. found more than a quarter of meltwater features on grounded ice, the vast majority of these were indeed located at low elevations near the grounding zones of ice shelves. Corr et al. actually is a great reference for this, so we will include a citation to their study. We will also change the terminology from **"grounding lines"** to "grounding zones", as suggested. Given the above, we propose to change our statement to: "...is mostly observed on the floating ice shelves and at low elevations near their grounding zones (Stokes et al. 2019; Corr et al., 2022)."

- Line 73: Acceleration isn't always the case. Some of these references are fairly old, Davison et al. have a nice review (https://doi.org/10.3389/feart.2019.00010), although there may be updated literature.
  To reflect this, we will change the phrasing to: "..., which has the potential to lubricate the bedrock and cause

acceleration of ice flow due to enhanced basal sliding." Davison et al. (2019) is a great reference for this and they explain very nicely how rapid lake drainage *usually* results in transient ice speed-up and net acceleration, so we will include a citation to this review.

- Line 75: I'm not sure this is strictly true it's been observed, more than it's been suggested as a possible mechanism.
  To reflect this, we will change the phrasing to: "... but recent observations suggest that this mechanism is also driving ice flow speed-ups on the Antarctic Ice Sheet."

- Line 97: If we're providing estimates we can also model this (e.g. https://tc.copernicus.org/articles/12/3565/2018/tc-12-3565-2018.html for individual lakes) but remote sensing is quicker/ more scalable but it's not technically true we have to rely on it.
  We will change **"estimates"** to "observations" to fix this, and simultaneously remind the reader that this section is about *observations* of supraglacial lakes.

- Line 152: Is it a good assumption that lakes are non-turbid? Certainly for sea ice ponds we can assume they are turbulent (of course they are shallower) but I wonder if e.g. it's a very windy day when the measurements are taken how much this impacts the flatness.
  The statement that we make is simply saying that the lakebed can be visible to ICESat-2 when water is non-turbid. If the water is too turbid, then ICESat-2 will not see a bathymetric signal. We are not making the assumption or claiming that all or most lakes are non-turbid. We would like to argue that windy days, lake flatness and turbulence are outside of the scope of an introductory sentence, which tells the reader that "if water is clear, ICESat-2 can look through it". (and we admit that we also don't quite understand ourselves how all these factors are related within the scope of line 152)
  To attempt to broadly answer the questions brought up here: it seems like most supraglacial lakes are sufficiently non-turbid for ICESat-2 to obtain a bathymetric signal, which makes sense because there are less sediments on the ice sheets than there are around inland water bodies, and algal growth in meltwater lakes seems to be fairly limited. The fact that ICESat-2 does not obtain a lakebed signal whenever turbidity is too high is actually helpful, since in this case the optical methods based on imagery that are commonly used will usually result in wrong depth estimates (e.g., Lutz et al., 2024).
  To discuss the effects of wind (also requested by RC2), we propose to re-write the part starting on line 227:
  "Afterpulses only become noticeable when the sensor is nearly or fully saturated, which means they often appear in ATL03 data over supraglacial lakes because smooth open water surfaces (i.e. the surface of stationary water bodies that are not affected by wind) can result in specular reflection. This suggests that the presence of wind ripples increases the likelihood of detecting a lake with a clear bathymetric signal in ATL03 data by preventing sensor saturation and afterpulsing (Lu et al., 2019; Tilling et al., 2020) and also explains why we observe afterpulsing more frequently near the (more wind-shielded) margins of melt lakes than over their (more wind-exposed) interior."

- Line 207: How were these numbers determined (empirical observation is a little broad e.g. how many lakes were examined?)
  We agree that "empirical observations" is a bit too unspecific. We used a trial-and-error approach based on a number of different hand-picked granules that featured a broad range of surface conditions, including some that contained many lakes and some that did not contain any lakes. For clarification, we propose to change this part to the following:
  "Based on these assumptions, and using a trial-and-error approach, we defined the following thresholds on the density ratios that need to hold for a major frame to pass the flatness check: $d_0/d_1 \geq 2$, $d_0/d_2 \geq 5$, $d_0/d_3 \geq 10$ and $d_0/d_4 \geq 100$. As part of this trial-and-error approach, we manually assessed the effects of

tweaking the above thresholds on a number of hand-picked granules, which we judged to be likely representative of various possible environments, to ensure adequate performance (i.e. granules without surface melt vs. pervasive surface melt, granules with smooth vs. rough background topography, granules containing ice-covered and partially ice-covered lakes, granules containing slush areas, granules containing exposed bedrock, partially cloudy granules, weak vs. strong beam data, night- vs. daytime acquisitions, etc.)."

- Line 259 onwards: What if the lake is e.g. 0.92m deep, would the bathymetric signal still get picked up? I think this is what you are saying in line 273 but you could clarify how you might discern the two, or if it is likely to be possible.
  We agree that we did not describe this clearly enough. Since we remove afterpulses in saturated pulses only, bathymetric signals will be preserved in unsaturated pulses – unless if all the pulses are saturated, in which case an actual flat bathymetric signal is practically indistinguishable from an afterpulse in the point cloud data. To clarify this, we propose to add the following to the end of the paragraph that starts on line 259:
  "Since this procedure removes photons in saturated pulses only, true bathymetric signals that overlap with the elevation of a known afterpulse are still retained as long as they appear in any unsaturated pulses. However, if all pulses within an along-track section of the data are saturated, any true bathymetric signals from a flat lakebed at the elevation of a known afterpulse will be removed from the data because they are practically indistinguishable from the afterpulses that we expect to see in the point cloud under such highly saturated conditions."

- Line 314: Can you determine why there is no signal from the lake bed here? Is it in the ice covered lake areas?
  It appears we did not include this in the text, but these two major frames overlap exactly with the partially ice-covered area. Thanks for catching that. We will add the following to the end of the paragraph:
  "These two major frames visibly overlap with the location of a thin partial ice cover near the lake's northern shore (Fig. 7, panel a), which explains why some of the lakebed is occluded in the photon cloud. While such areas, where part of the lakebed is occluded, may not pass the bathymetry check, they are later included in the data that make up a full ICESat-2 lake segment, as explained in the next section."

- Line 319: Not sure if the word 'each' here is a typo?
  We will delete **"each"**.

- Line 321: See my general comment about lake sizes, what does a 'few' mean here, please be more specific. Is this related to the 10 in equation 3? (Or if not where does that come from?)
  We agree that we should be more specific here. The phrase "along-track extent" was poorly chosen. The 10 in equation 3 is indeed what describes this: the maximum separation between two clusters to be merged is 10 major frames, so clusters that are separated by more than 1.4-1.6 km are not merged. To clarify this we propose to replace (ii) on line 321 with:
  "(ii) a ground track rarely crosses the same lake in two distinct locations that are separated by more than about 1.5 km."
  We will further explain equations (2) and (3) in plain language after they are defined:
  "Equation (2) states that neighboring clusters are only merged if their respective lake surface elevations are within 0.1 m of each other, and Eq. (3) further states that neighboring clusters are only merged if they are separated by ten major frames that did not pass the bathymetry check, or less (about 1.5 km). This means that if FLUID encounters the unlikely but possible scenario in which a ground track crosses two arms of the same lake, which are separated in along-track distance by more than ten major frames, then these two crossings are considered to be separate lake segments and returned as two separate files in the output data rather than being merged together into

one lake segment. If these two conditions do not result in…"

- Line 350: I'm not sure what you mean here by 'removing any lakes that fully overlap with another lake'. Does that not just make them the same lake (and is some information about that lake then lost)?
We agree that **"fully overlap"** sounds confusing. We propose to change this to:
"Since this expansion of the along-track ranges of lake segments can create lake segments that overlap, the set of buffered lake segments is corrected by separating partially overlapping lake segments at the midpoint of their along-track overlap and removing any lake segments that are fully contained within another lake segment."
[Explanation: Note that a lake segment being fully contained within another one means that if we remove it no information is lost, we just remove information that was duplicated. This does not happen often in practice, but imagine a large lake that has a physically thin but optically thick ice cover that stretches for more than 1.5 km. If ICESat-2 obtains a bathymetric signal on both sides of the ice cover, then FLUID will consider those two stretches of signal to be two separate lake segments (Eq. 2 above), but since the ice cover is physically thin (and therefore flat and very close in elevation to the water surface), FLUID will also expand the along-track ranges of both segments across the entire lake, including its ice cover. In this situation, there will be two lake segments that both have data for the entire lake, so one of them is a duplicate and needs to be removed. → We consider this to be such a fringe case that we do not think it needs to be explained in full in the manuscript.]

- Line 405: So is this a limitation on minimum lake depths that can be detected?
Yes, 0.35 m is a theoretical minimum, but in practice lakes < 0.5 m deep are not very reliably detected due to noise in the data, and the fact that such lake segments are often short and tend to be in locations with relatively flat surrounding surface topography. See our answer to the first general comment.

- Line 435: What is the situation where this could happen? Does this suggest a lack of confidence in surface retrievals or is the algorithm picking up something else (an ice lens?).
This happens anywhere where the fit to the photons that remain after removing the ones reflected from the water surface is above the water level of the lake. The most frequent example for this is the surface of the ice that surrounds a lake (e.g., Fig 7, panel d: in about the first and last 200 m in along-track distance, the red line is clearly above the elevation of the water surface), but it can also happen when there is a thick ice cover on a lake. SuRRF is usually quite confident in the fit in these situations, because surface returns above water tend to be much stronger than bathymetric returns. Either way, SuRRF is fitting something that is at an elevation higher than the surface of the water, so we can be generally confident that water depth is zero there.

- Section 4.1.2: Could you compare with the Moussavi et al dataset? That goes up to 2020 according to the data description: https://www.usap-dc.org/view/dataset/601401
The Moussavi et al. (2020) dataset includes only one of the four melt seasons considered in this manuscript. Furthermore, it is >70GB in size and has to be downloaded as a single file by contacting the data center and making an individualized manual request. As such, it is not very accessible and not very practical to use. While this dataset includes estimates of lake depths, in Section 4.1.2 we assess FLUID lake *detection* only (FLUID itself does not provide lake depths). This means we are concerned with whether the locations of detected lakes overlap with the locations detected in imagery. For this purpose, the Moussavi dataset would be overkill, and only give us information about one out of four melt seasons. Mapping the spatial extent of melt lakes based on imagery is quite simple, and Tuckett et al. (2021) made it even easier by sharing their methods. Due to all the reasons above, we believe that it makes much more sense in the context of this section to simply map the extent of supraglacial lakes using Google Earth Engine and compare their extents to the locations of FLUID lake segments.

To make this more intuitive to the invested reader, we will include high-resolution maps of the Landsat 8 lake extents and ICESat-2 lake segments for all the melt seasons included in the paper in the updated release of the supplementary material at https://zenodo.org/doi/10.5281/zenodo.10901826 (see also attached to the end of this document).
To compare our SuRRF *depth* estimates to RTE-based depth estimates from imagery that are similar to the Moussavi et al. (2020) dataset, we will include the RTE depth estimated reported in the papers that we use to assess depth estimation accuracy in section 4.2.2 (Fricker et al., 2021 and Melling et al, 2024). The Sentinel-2 based depth estimates in Fricker et al. (2021) were contributed by M. Moussavi based on the dataset that you mentioned. We also include Antarctic lake depth estimates based on Landsat 8 (contributed to Fricker et al. 2021 by J. Spergel and J. Kingslake and based on Spergel et al. 2021) and Greenland lake depths based on Sentinel-2 from Melling et al. 2024. (see comment above about comparison to alternative methods)

- Section 4.2.2: How many lakes are there actually in track 81? Is this something that could be determined manually to get an idea of false positives/ negatives from SuRFF?
There are eight lakes in track 81 GT2L. We described this in section 4.1.2:
"To evaluate whether FLUID detects all supraglacial lakes that have bathymetric data in this ATL03 ground track, we manually inspected the data for any evidence of meltwater and determined whether any such along-track contained a clearly discernible return from a lakebed. Out of 25 along-track segments with meltwater, we judged that only eight were supraglacial lakes with a clear bathymetric signal."
We did further examine same-day Sentinel-2 imagery to visually verify that ICESat-2's ground track does not overlap with any melt lakes that are not captured in the ATL03 data in track 81 GT2L. We did not mention this in the submitted manuscript, so we will add this to the revised manuscript:
"We further evaluated whether the ATL03 photon cloud misses any supraglacial lakes that track 81 GT2L crosses, by mapping it over a mosaic of Sentinel-2 scenes from the same day (same as used in Fricker et al. 2021). Based on visual inspection, the ICESat-2 ground track crossed supraglacial lakes that were clearly distinguishable in the imagery only in the same eight locations that we had also judged to be lake segments in the ATL03 data. Most other ICESat-2 segments that showed evidence of surface water in ATL03 also showed some evidence of meltwater in the imagery, and were associated with ice-filled crevasses, narrow melt channels, likely areas of slush, or melt lakes with an opaque ice cover (for which depth determination is not possible)."
In Section 4.2.2, we refer back to section 4.1.2 where it was explained.
"For the 16 potential melt lakes that FLUID identified in ICESat-2 track 81, GT2L on Jan 2nd, 2019 over the Amery catchment (Sect. 4.1.2 Fricker et al., 2021), SuRRF assigned a nonzero lake quality score only to the eight data segments which we had manually determined to be supraglacial lakes with a clear bathymetric signal."
To clarify that the eight lake segments mentioned in these sections are the only ones present in track 81, we will replace **"clear"** with "discernible" (because "clear" leaves it open to interpretation whether there could still be lakes segments with bathymetric signals in the track, whose bathymetric signals just aren't super obvious).

- Line 630: 'this' satellite rather than 'a' satellite. Datasets exist for other satellites.
ICESat-2 is the only satellite (to date, and likely for a while into the future) that can directly measure lake depths from space, so **"a satellite"** is correct. However, we never stated this explicitly in the manuscript, and should absolutely do so. To make this evident to the reader, we will state this early on in the introduction (line 37): "Launched in 2018, NASA's Ice, Cloud and land Elevation Satellite (ICESat-2) laser altimeter became the first (and thus far only) satellite capable of making direct, accurate water depth measurements from space, …"

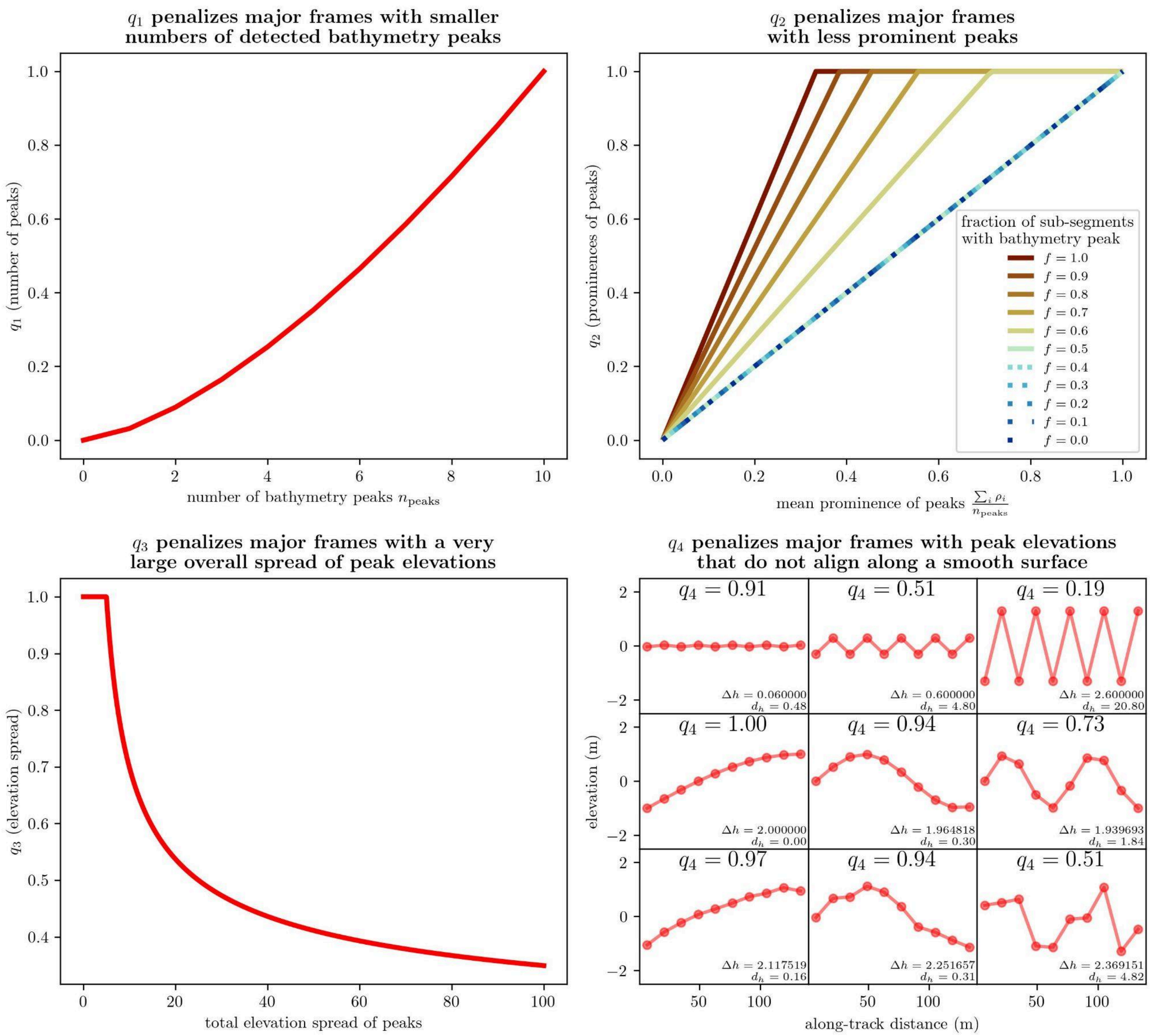
- Line 740: What were these equations based on? I understand you used trial and error but they are complex equations that must have had a starting point.

  We propose to add the following to the end of Appendix C to illustrate our thought process:

  "While the quality heuristics q_i were obtained by trial-and-error, the starting points for this approach were based on the following assumptions and observations:

  - The starting point for q_1 was the idea that major frames with a smaller number of detected bathymetry peaks are less likely to have a consistent signal from a lakebed. The way the equation for q_1 is designed, major frames with a very small fraction of detected bathymetry peaks (0-30% of sub-segments) are disproportionately penalized, since such small numbers of peaks are much more likely to be noise.
  - The starting point for q_2 was the idea that major frames with less prominent peaks represent either a weak, inconsistent bathymetric signal or noise. Here, q_2 is equal to the mean prominence of peaks when the fraction of detected bathymetry peaks $f < 0.5$. However, for fractions larger than 0.5, the assumption is that the bathymetric signal is consistent enough that even smaller mean peak prominence suggests that the bathymetric signal is clearly visible.
  - The starting point for q_3 was the observation that in most cases supraglacial lake segments with a usable bathymetric signal have fairly small lakebed slopes. Therefore, we do not expect the lakebed elevation of a major frame with a good signal to span a large elevation range. In contrast, we observed that major frames with detected potential bathymetric peaks that span a very large elevation range are often due to noise in the data. Based on this, we designed the equation for q_3 such that its value drops off once the total elevation range of detected bathymetry peaks within the major frame (of length ~140 m) exceeds 5 m. There are, however, some lake segments with a clear bathymetric signal that do exhibit a large range of lakebed elevations (often due to a single burst of noise being detected as a potential bathymetric peak). Therefore, we made sure that the value for q_3 is large enough for major frames to still pass the bathymetry check if Δh is very large but its other quality heuristics q_i are closer to 1.
  - The starting point for q_4 was the idea that in most cases bathymetry peak elevations will align along a smooth surface in the along-track distance direction rather than randomly fluctuate (which is usually the case for noise). Here, we penalize every "direction change" (i.e., wherever a peak has two neighbors and its elevation constitutes a local minimum or maximum). We allow for random fluctuations of up to 0.5 m per peak detection without a large penalty, since we observed that a vertical photon spread of up to about this value is quite possible even for lake segments with a somewhat "fuzzy", yet clearly distinguishable return signal from the lakebed.

A figure that illustrates these quality heuristics is available in the supporting information at https://doi.org/10.5281/zenodo.10901826 (Arndt and Fricker, 2024c). (and shown below for convenience)


q_1 penalizes major frames with smaller numbers of detected bathymetry peaks
q_1 (number of peaks)
number of bathymetry peaks n_{peaks}
q_2 penalizes major frames with less prominent peaks
q_2 (prominences of peaks)
mean prominence of peaks $\frac{\sum_i \rho_i}{n_{\text{peaks}}}$
fraction of sub-segments with bathymetry peak
$f = 1.0$
$f = 0.9$
$f = 0.8$
$f = 0.7$
$f = 0.6$
$f = 0.5$
$f = 0.4$
$f = 0.3$
$f = 0.2$
$f = 0.1$
$f = 0.0$
q_3 penalizes major frames with a very large overall spread of peak elevations
q_3 (elevation spread)
total elevation spread of peaks
q_4 penalizes major frames with peak elevations that do not align along a smooth surface
$q_4 = 0.91$ $\Delta h = 0.060000$ $d_h = 0.48$
$q_4 = 0.51$ $\Delta h = 0.600000$ $d_h = 4.80$
$q_4 = 0.19$ $\Delta h = 2.600000$ $d_h = 20.80$
$q_4 = 1.00$ $\Delta h = 2.000000$ $d_h = 0.00$
$q_4 = 0.94$ $\Delta h = 1.964818$ $d_h = 0.30$
$q_4 = 0.73$ $\Delta h = 1.939693$ $d_h = 1.84$
$q_4 = 0.97$ $\Delta h = 2.117519$ $d_h = 0.16$
$q_4 = 0.94$ $\Delta h = 2.251657$ $d_h = 0.31$
$q_4 = 0.51$ $\Delta h = 2.369151$ $d_h = 4.82$
elevation (m)
along-track distance (m)

References

- Dietrich, J. T., Rackley Reese, A., Gibbons, A., Magruder, L. A., & Parrish, C. E. (2024). Analysis of ICESat-2 data acquisition algorithm parameter enhancements to improve worldwide bathymetric coverage. Earth and Space Science, 11(2), e2023EA003270.

- Bowling, J. S., Livingstone, S. J., Sole, A. J., & Chu, W. (2019). Distribution and dynamics of Greenland subglacial lakes. Nature communications, 10(1), 2810.

- Davison, B. J., Sole, A. J., Livingstone, S. J., Cowton, T. R., & Nienow, P. W. (2019). The influence of hydrology on the dynamics of land-terminating sectors of the Greenland ice sheet. Frontiers in Earth Science, 7, 10.

- Georgiou, S., Shepherd, A., McMillan, M., & Nienow, P. (2009). Seasonal evolution of supraglacial lake volume from ASTER imagery. Annals of Glaciology, 50(52), 95-100.

- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

- Johansson, A. M., Jansson, P., & Brown, I. A. (2013). Spatial and temporal variations in lakes on the Greenland Ice Sheet. Journal of hydrology, 476, 314-320.

- Lu, X., Hu, Y., & Yang, Y. (2019, December). Ocean subsurface study from ICESat-2 mission. In 2019 photonics & electromagnetics research symposium-fall (PIERS-fall) (pp. 910-918). IEEE.

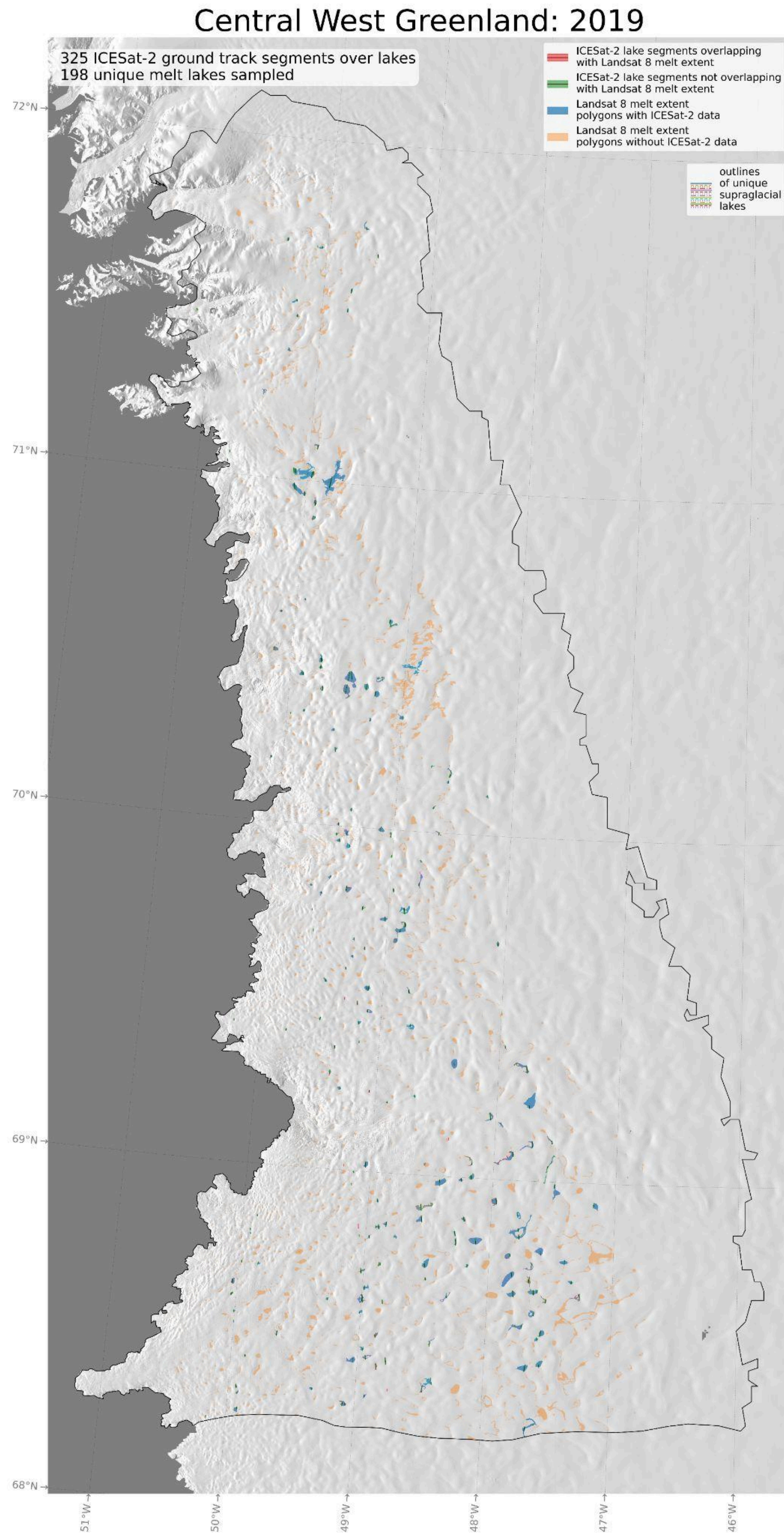
- Lutz, K., Bever, L., Sommer, C., Humbert, A., Scheinert, M., & Braun, M. (2024). Assessing supraglacial lake depth using ICESat-2, Sentinel-2, TanDEM-X, and in situ sonar measurements over Northeast Greenland. EGUsphere, 2024, 1-25.

- Markham, I. S., & Rakes, T. R. (1998). The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. Computers & operations research, 25(4), 251-263.

- Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., ... & Philip, S. Y. (2022). Generalizing to unseen domains: A survey on domain generalization. IEEE transactions on knowledge and data engineering, 35(8), 8052-8072.

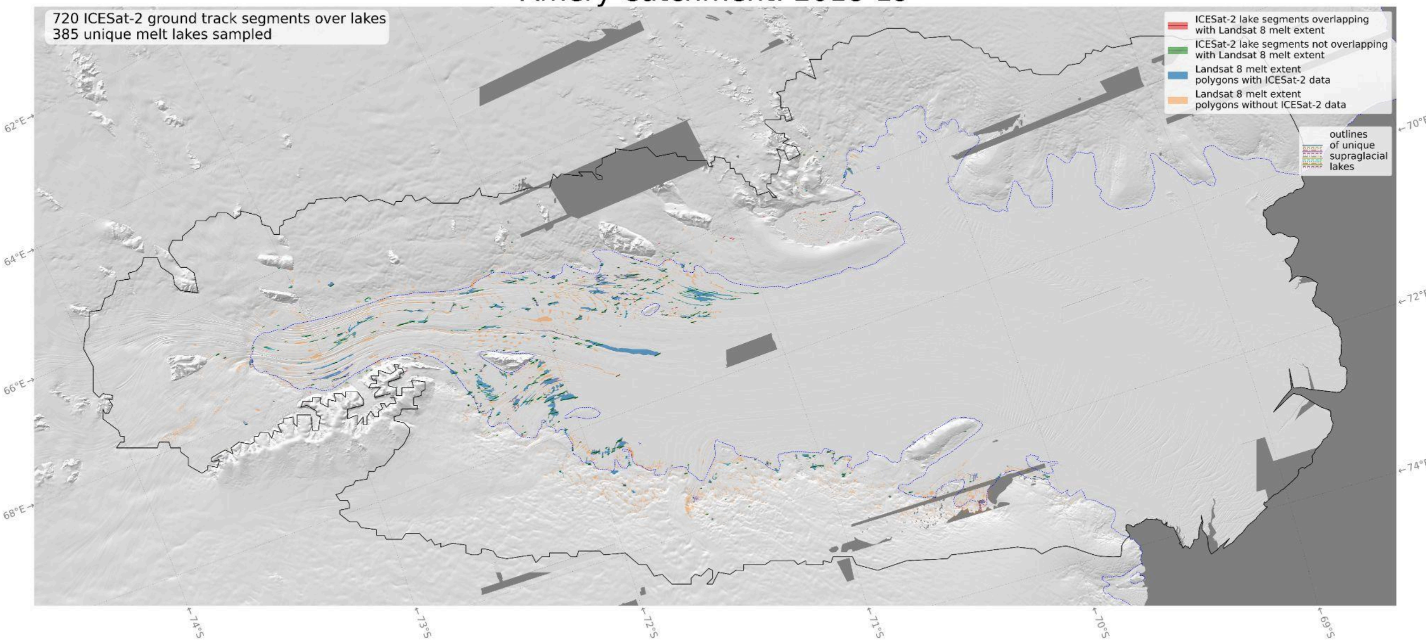
Maps of ICESat-2 lake segments and unique supraglacial lake extents

Reviewer comments about distinguishing “unique supraglacial lakes” from distinct “ICESat-2 melt lake segments” prompted us to estimate which distinct melt lakes are sampled by FLUID lake segments more than once. We report the number of unique supraglacial lakes in the text of the manuscript and Table 1, and explain how we estimated the given numbers. In a new version of the supplement at https://doi.org/10.5281/zenodo.10901826 (Arndt and Fricker, 2024c), we will additionally provide high-resolution maps of the ground tracks of FLUID lake segments over Landsat 8 maximum melt extents, with unique lakes identified by their borders. These maps can be used to better understand the spatial distribution of FLUID melt lake segments and how they relate to the overall distribution of pooled meltwater over our study regions. Due to their large scale, these maps cannot be reasonably included in the main manuscript. Lower-resolution reproductions of these maps are shown on the following pages.


Central West Greenland: 2019
325 ICESat-2 ground track segments over lakes
198 unique melt lakes sampled
ICESat-2 lake segments overlapping with Landsat 8 melt extent
ICESat-2 lake segments not overlapping with Landsat 8 melt extent
Landsat 8 melt extent polygons with ICESat-2 data
Landsat 8 melt extent polygons without ICESat-2 data
outlines of unique supraglacial lakes
72°N
71°N
70°N
69°N
68°N
51°W
50°W
49°W
48°W
47°W
46°W


Central West Greenland: 2020
175 ICESat-2 ground track segments over lakes
114 unique melt lakes sampled
ICESat-2 lake segments overlapping with Landsat 8 melt extent
ICESat-2 lake segments not overlapping with Landsat 8 melt extent
Landsat 8 melt extent polygons with ICESat-2 data
Landsat 8 melt extent polygons without ICESat-2 data
outlines of unique supraglacial lakes
72°N
71°N
70°N
69°N
68°N
51°W
50°W
49°W
48°W
47°W
46°W

Amery Catchment: 2018-19



Amery Catchment: 2020-21