Response to Reviewer #2 – Devon Dunmire, submitted 20 June 2024

General response

Thank you for your positive feedback; we are glad that the benefit of our research is well received. Additionally, we appreciate the thorough and constructive comments given in your review. We generally agree with your comments and have gladly integrated the majority of them into our manuscript. More specific responses to your comments can be found below.

Specific responses

General comments

We have made efforts to improve the readability, flow and redundancy of the writing where possible, especially in the examples you noted. We have also paid attention to the consistency of abbreviation usage and have tried to adapt a more consistent tense in the methods section. Overall, we have also tried to change as many phrases as possible to a more active, first-person form where sensible.

Abstract

We have integrated the comments mentioned over L14 – 28, in addition to making the tense more consistent and the final sentence more concrete.

Introduction

We have integrated the comments mentioned over L34 – 75, with the exception of introducing sonar data in the introduction, as we think the explanation given later in the methods section is sufficient for the reader's understanding.

Data and Methods

Any comments not specifically addressed below have been integrated into the Data and Methods section.

"In general, I think the organization of this section should be re-worked. The subsections switch between data, methods, and study region. Maybe it would make sense to break up the Data and Methods into separate sections? I think a more organized approach would start by introducing all the data used and then move on to the 4 different methods."

We had also originally thought to organize it in that way, but after writing it, realized that the sonar and ICESat-2 sections flowed much better when the data was described directly with the methodology.

"How are the lakes in this study delineated? Is this done manually, or have you used a pre-existing algorithm?"
They are automatically delineated using a deep learning method developed in Lutz et al. (2023). This was briefly mentioned in L214 for the TanDEM-X method and L226-227 for the method comparison section, but we agree that it was not made clear that this deep learning method was used throughout the entire method comparison. We have added another sentence to Section 2.6 to clarify that.

"L78-80: I would consider removing these lines. They feel out of place considering they introduce the four methods and then the next subsection immediately covers the data."

We understand how it seems a little jarring introducing the four methods and then going into the Sentinel-2 description; however, we think clearly presenting the four methods (and defining the names we will call them throughout the paper) is beneficial to have at the beginning of the chapter.

"L120: Surely these lakes are not the only lakes where an ICESat-2 path crossed a filled SGL? Did you look at all GrIS regions or just NE and SW?"
There are certainly more lake crossings than the ones we found. We searched for the lake crossings manually and while we disregarded some lake crossings due to low quality surface and bed distinction, there are surely some that were simply missed by us. Due to the high prevalence of SGLs in NE and SW Greenland, we limited our search to those regions; however, one would be able to find crossings in other regions around Greenland as well. A few words were added to this section to hopefully make it clearer that these are just the crossings used in this study – not all possible ones.

"L149: The sentences beginning with "Figure 2(b)…" seem out of place in this ICESat-2 lake cross tracking retrieval section, as these sentences refer only to Sentinel-2 data."
It is true that it does not fit perfectly with the section topic, but to create another section just for a couple sentences seems unnecessary as well. We have made it into a new paragraph and added a sentence before it that highlights the relevance of the next sentences in the context of the section.

"L175 – How is this error estimated? Is this from a different paper?"
I realized upon re-reading it, that the units on the error were incorrectly written. I have now corrected it to 0.20 m instead of cm. To address your point concerning the estimation of this error, we assume an error of 20 cm based simply on experience using the sonar tool and visual estimation of the clarity of the bed and surface boundaries. We have now added this clarification into the manuscript.

"L175 – The naming convention is unclear to me. How is it based on the location of the topographical depressions? As there are only 19 lakes in the manuscript here, perhaps it would be clearer to rename the lakes to something simpler (thus also removing the need for both numbers and letters in the naming)."
We have identified 1035 topographical depressions in our study area in Northeast Greenland and have labelled each one with an ID number. Since we only measured lakes with the sonar boat in Northeast Greenland, we used our depression numbering system to identify them. We have now added a sentence to this paragraph clarifying the system. However, when addressing the lakes with ICESat-2 crossings, we have now changed the identification of these lakes in Figure 1 and Table A1 to more generalized ID numbers (instead of the ICESat-2 beam IDs) since we do not have a system for Southwest Greenland.

"L208 – Why are the DEMs used from after lake drainage? Would it make more sense to use DEMs from before lake-filling as this would more accurately represent the non-lake surface? After drainage, the DEM surface may also include ice fractures or ridges that result from the drainage, which would therefore provide an inaccurate representation of the actual lake depth."
While there certainly could be changes in the ice surface after a drainage due to fracture and uplift, there could also be problems using a DEM from early in the season. The lake could have become buried under a layer of ice over winter, which would create a completely different DEM of the ice surface if the DEM is taken before the surface melts. Furthermore, with a longer time interval between the DEM creation and the lake level comparison, the ice could have flowed substantially, allowing for crevasses or other features to be moved into or out of the lake region. Thus, while there are potentially issues with taking it before or after, we chose to use post-drainage DEMs in this study.

"L218 – How is the surface elevation RMSE calculated? Is it compared with the Copernicus DEM? Or is this really the standard deviation of the lake edge pixel elevation?"
You are right – it is indeed the standard deviation of the lake edge pixel elevations. We have updated that in the manuscript.

Results

Any comments not specifically addressed below have been integrated into the Results section.

"L260-265 – I don't fully follow this section. For example: "while the data points in the Southwest function…". Are these 'data points' from actual ICESat-2 data? The use of 'in the Southwest function' confuses me a bit."

Yes, the data points are referring to the data in Fig. 5 (a) – (c), where the depths from ICESat-2 and the reflectance values from Sentinel-2 are plotted against each other for the different bands. For the green band, we fit a curve to the data found in the Northeast and to the data found in the Southwest, which we then refer to as the Northeast function and Southwest function. They are defined in Equations (3) and (4).

"L271 – Out of curiosity, if Lake 469 was included in the analysis, would it still reasonably fit the same curve in Fig. 5d?"

The curve including the data from Lake 469 resulted in the depths estimations being noticeably deeper.

"L280 – It is mentioned that the RMSE ranges from 0.27 to 0.94 m, but for which reflectance values is the RMSE relatively low or high?"

It varies quite a lot over the reflectance values, as it depends on how many data points there are within each bin and how large the spread of values within the bin is. So it is not as simple as saying the RMSE is larger for shallower depths, for example.

"Section 3.3 – A figure that shows depth error (compared with DEM) with lake depth for each non-DEM method would be helpful to reference throughout this section. For example, I'm thinking of something like this below:"

Thank you for this suggestion. We have now created a new figure similar to the example you drew, showing the average error of each method across the DEM depths. We have put this figure in the appendix and reference it throughout Section 3.3.

"L286 – What are the errors associate with using the DEM results as a reference? How accurate are the DEM results?"

The vertical offset of the TanDEM-X DEMs after co-registration, which is the remaining mean vertical deviation between each TanDEM-X acquisition and the Copernicus reference DEM, on stable non-glacierized terrain is 0.17m (SD = 2.40m) and 0.07m (SD = 3.58m) for the 2021-07-23 and 2021-08-13 DEMs, respectively. However, we deliberately did not include this estimate in the manuscript as we do not directly compare surface elevations of the TanDEM-X DEMs to other elevation datasets (e.g. Copernicus DEM or ICESat-2 altimetry). Instead, we extract the relative elevation difference between the lake shore and bed exclusively from each TanDEM-X DEM. Therefore, this relative elevation difference is not related to the accuracy of co-registration but to the creation of the differential interferograms and subsequent phase unwrapping. To avoid biases in the SAR elevation difference due to artefacts during the DEM creation, such as phase jumps, we investigated the distribution of the interferometric coherence, which indicated a loss of coherence for some parts of the lake bed. We removed the respective areas at the lake bottom as shown in Figure 6 to avoid a biased elevation difference estimate due to doubtful DEM values. For the remaining lake bed and shore areas the presence of interferometric distortions is unlikely but unfortunately we cannot quantify a meaningful error budget for the outlined relative elevation difference.

"L288 – The sentence: "The sonar equation produces the largest errors in the shallower regions", reads to me as: compared to the other methods, sonar is the worst in shallow water. Instead, I guess it is meant that sonar does worse in shallow water compared to its performance in deeper water. Consider rewording this for clarity."

We realized this statement is actually misleading and have thus removed it from the manuscript.

"L303 – What does 'significantly' mean here? 1m? 5m?"
The difference in estimation ranges up to 10 m in some areas. This can be better seen in Figure B1 in the appendix.

"L312, Figure 7 – How are these uncertainties determined?"
These uncertainties were determined based on different factors depending on the method and the associated instruments. For the sonar method, we estimated a geolocation error based on the variability between neighboring pixels, an error for wave-rocking of the boat based on crossing points of the sonar tracks, the RMSE of the data spread to the fit curve and the delineation of the lake bed in the sonar reading tool. Similarly, for the ICESat-2 method, we estimated a geolocation error, the RMSE of the data spread to the fit curve, and delineation errors of the lake bed and surface. Finally, for the RTM method, we estimated the error based on the sensitivity analysis shown in Fig. B3, where the change in the parameters of g, A_d width, and R_{inf} over a defined depth change were combined into a single uncertainty. The uncertainties were then applied for each pixel of lake water over the time series and summed per day. More detail to this has now been added to the beginning of Section 3.4.

Discussion
Any comments not specifically addressed below have been integrated into the Discussion section.

"Again, a reference to Melling et al 2024 and a discussion of the results in the context of this previous work is missing here."
We fully agree with including the citation of Melling et al. (2024), and see how the results align with and support our analysis. This citation was simply overlooked since it was published close to the finalization/internal review of our own manuscript. Reference to this work has now been added throughout the manuscript.

"L341 – "depths between 10 and 25 m." Are these maximum depths?"
Yes, this was referring to the maximum depths. A few words have been added to the sentence to clarify that.

"L341 – "Moreover, in the interannual comparison…" Which method are these statistics based on?"
They were based on the sonar equation. Since that was indeed not stated, it has now been added to that sentence.

"L352 – What we care about at the end of the day is the actual volume of water stored in SGLs. The red band underestimates deep depths and the green band overestimates shallow depths; but, we can't truly know which method is better suited without a similar comparison between red-band derived depths and DEM-derived depths. I'm not fully convinced that the question of which band is better suited can be answered with this work as it only includes 5 lakes in NE Greenland. How do we know that these 5 lakes are generally representative of GrIS SGLs?"
It is true that five lakes is not a large sample size to make definitive statements on. However, looking at Figure B3, we can see that the average depth of a substantial portion of the lakes found in NE Greenland is above 4 m, where 300 – 550 lakes are assessed for each year. As this is just the average depth of each lake, even lakes with average depths lower than that naturally also have depths reaching higher values. Even though these depths are only estimated using the sonar equation and not DEMs, we have seen with the 5 test lakes that these two methods tend to be in agreement under 8 meters or so. This is a strong indicator that lakes tend to regularly reach depths out of the scope of the red band. That is why we stated in L344 that "…the use of the green band seems to be a more suitable choice for estimating deeper lake depths…". Of course, if it is known that a lake only contains depths up to 3 or 4 meters, then the red band would be a suitable choice. That is also why we state in L449 in the conclusion "to improve the methodology overall, combining estimations from red, green, and blue

bands into a single algorithm could potentially overcome the attenuation limitations of each band, allowing for more accurate estimations in shallow water with the red band…"

"L358 – Can you quantify how many (or what %) of points used in the regression are deep (> 7m) or shallow (< 0.5m)."

For the ICESat-2 data from the northeastern lakes, there are no points above 7 m. For the shallow depths, 24/246 of the data points are below 0.5 m, so 9.76%. For the sonar data, there are only four data points above 7 m (1.64%) and eight data points below 0.5m (3.28%). We have added these statistics to the manuscript.

"L359 – It is not immediately clear to me why not having many very deep or shallow points for the regression would lead the ICESat-2 equation to overestimate depths."

We have removed this phrase from the manuscript.

"L364 – "… the sonar equation regression never reaching a value below 0.5 m…" Can you place a boundary condition on the regression equation such that the line has to reach 0m in this range of reflectance values? From Figure 5 it appears that there are several points that are near 0m depth. Can you force your regression to cross the y-axis somewhere between these reflection values so that the equation is physically bound?"

One could force the regression to reach 0 m; however, we chose not to do this here for a couple of reasons. Firstly, there are only a few points near 0 m and they are spread out over half of the reflectance value range, since the reflectance value in shallow water varies quite drastically due to the influence on the surrounding ice and environmental conditions. Because of this, it would be difficult to choose a reflectance value based on the data presented. Similarly, we wanted the methods to be purely based on the observed data to limit external bias. In future research, it could be valuable to force the equation to reach zero when taking other sources of information into account.

"L387 – "… imperfect lake masks…" Did you consider remasking the lakes to be sure that no water pixels were included in your calculation of Ad?"

This comment was meant in a more general sense, especially when applied to dense time series. Of course for the evaluation of our five test lakes, the lake masks were checked to make sure there were no large errors. That being said, it is not always clear exactly where the edge is. In an automated time series analysis like the one shown in Fig. 7, however, manual evaluation of individual lakes is not feasible. In such cases, if the lake masks are not perfect, the calculation of A_d could be affected.

"L394 – Since you discuss % changes in volume of the lake, I think it would be helpful to provide the parameter changes as % changes as well. E.g: what % change is a 0.01 m-1 change in g?"

The change in volume was represented in percentage because a value given in km^3 would not be as directly understandable for the reader since they are not familiar with the overall volume of the lake. A percentage change or, as we additionally wrote, a change in average depth in meters helps the readers put it into context more easily. This, however, would not be the case if the change in *g* were represented as a percentage, since the readers would be more familiar with the actual values being used for this parameter. Thus, we think it best to leave the representations as they are.

"I would like to see a slightly expanded sensitivity analysis. Fig. B3 only has 3 points per panel. I think this analysis would be improved if there were more values tested. Also, in L394, it is mentioned that a 0.01 m-1 change in g results in a 7.4% change in the lake volume, but none of the points in Figure B3 represent a 0.01 m-1 change in g."

The three values for each parameter were chosen specifically as being the realistic range of values from which one could include in their calculations, with the central value being the one used in our study. Thus, expanding the range further beyond these values would not bring any concrete benefit. Additionally, the parameters *g* and R_{inf} have seemingly linear relationships to depth, so adding more values would not make the trend any more apparent than it already is. Although the parameter A_d

width does not have a directly linear relationship, we also do not think that adding more values between 10 and 60 m would substantially increase the amount of information able to be obtained from the graph. In the end, the focus of this study was intended to show an example of how much these parameters can vary on one lake. To address your last point, after finding the amount the volume changed over the span of *g* values investigated, we scaled it so that the volume change is represented over a simple unit of *g* instead of the difference of values shown in Fig. B4.

Figures

Any comments not specifically addressed below have been integrated into the figures and/or their captions.

"Figure 1 – It would be helpful to also plot the lake locations in 1b. Also, in the caption header the future tense "will be" is used. Please change this to "was". Finally, the titles for the different lakes are not intuitive. I assume the date is in the form dd-mm-yy. I think it would be better to label the images with the date of the optical image (in a more intuitive format), as the beam information is already located in Table A1."

We have now added the lake locations to subset 1b as well as the glacier labels for better orientation. We have also changed the lake labels to be more easily understandable, i.e. NE1, SW1, CW1 and so on to refer to the region in which they are found.

"Figure 5 – I think it would make more sense to order the plots by band wavelength : R (a), B (b), G (c). The legend for the NE and SW Greenland points should be in panel (a). Did you use the Red and Blue bands with Sonar? If so, it may be interesting to see these bands as panels as well. In the caption, please specify SGL depth (L247). How does the fitted curve in d compare with the fitted curve in c?"

We plotted the bands in that order so that the green band for the ICESat-2 data would be directly next to the green band for the sonar data, allowing for a side-by-side comparison. We did not use the red and blue bands from the sonar data, having seen how ideal the green band was for our purposes in the ICESat-2 data.

"Figure 6 – Should mention the image date for each lake, either in the image or the caption. A separate (maybe in Appendix) figure showing the depth difference between each method and the DEM method would be helpful to more clearly visualize the differences, specifically as this is discussed in section 3.3."

We agree and have added the image dates for each lake to the figure. Since we have added the figure highlighting the difference in depth estimation between the DEM and other methods as you suggested in Section 3.3, we think adding another figure visualizing the differences in depth estimation would be superfluous.

"Figure 8 – Include a map of where these lakes are located."

As these images are used simply as visual examples of observed phenomena, we do not think a detailed map of these specific lakes is necessary.

"Figure B3 - Include panel labels (a), (b), (c). It may be better to represent the axis as % changes. So, % change in the parameter values and corresponding % change in calculated lake volume."

We have now included the panel labels of (a), (b), and (c). We think it is better to not have the data represented in percentages because the specific values of the parameters could be of interest to those looking to use the method themselves.