Response to reviewers' comments on manuscript TC-2023-126: Estimating differential penetration of green (532 nm) laser light over sea ice with NASA's Airborne Topographic Mapper: observations and models

Editor: Huw Horgan

We use the following color and font coding scheme in our response to aid visual distinction for readers with color vision deficiency (CVD):

Referee's comments

Response: authors' response to comments.

Dear Dr. Horgan,

Our combined response to the two anonymous referees' excellent comments (RC1 and RC2) are below. We would like to thank both referees for their constructive and helpful feedback and the many suggestions to improve the paper. In our revised manuscript we will include a reference to the companion modeling paper that is now in review in The Cryosphere Discussions (https://doi.org/10.5194/tc-2023-147).

Best regards,

Michael Studinger

Response to the Referee 1 comments (RC1) manuscript TC-2023-126 https://doi.org/10.5194/tc-2023-126-RC1: Estimating differential penetration of green (532 nm) laser light over sea ice with NASA's Airborne Topographic Mapper: observations and models

We thank the referee for the positive general comments and helpful questions. We will revise our manuscript accordingly.

General questions:

In general, the work is concise and well-written. However, a reader not familiar with the field (e.g., I am a laser physicist, not particularly familiar with the intricacies of LIDAR altimetry) might benefit from

additional details and justifications that are potentially obvious to someone in the field. Hopefully, some of these points become clearer with the questions below.

> *<u>Response:</u> We have considered expanding the basic background of laser altimetry. We hope that some of our responses and suggested edits to the points below address some of that. The manuscript was prepared with the audience of The Cryosphere in mind. The ATM altimetry data from Operation IceBridge and the ICESat-2 data products are major data sets that are used by a large audience for a broad spectrum of cryospheric research. It is therefore reasonable to assume that a large part of the community has some familiarity with laser altimetry. We feel we have provided the appropriate level of detail and background in our manuscript that is suitable for the target audience of the journal. If the handling editor advises us to expand certain aspects, we will of course do so.*

Have laser wavelengths between 532 nm and 1064 nm been considered, such that return signals are stronger than at 1064 nm, while sub-surface scattering is reduced compared to 532 nm? Or are there currently no viable alternatives to what I assume are Q-switched Nd:YAG lasers?

> *<u>Response:</u> Our paper analyses existing data products for differential penetration over sea ice. A discussion of the pros and cons of 532 nm and 1064 nm lidars is a multi-faceted, complex topic that is beyond the scope of our manuscript. In lines 73 – 84 we provide an explanation of the challenges involved with dual-color lidars and a justification why we chose to focus our analysis on 532 nm lidar data only (lines 87 – 92). Generally speaking, the wavelength and type of laser (ATM has used short-pulse diode-pumped, frequency-doubled (not Q-switched) lasers) is not the only consideration which laser to fly on an aircraft. Detector considerations, in particular detectors that can detect and not distort high-bandwidth laser pulses in high-reflectivity snow background light that varies when overflying darker ice/rock often drive the selection of lasers. For example, the decision to use a 532 nm laser on ICESat-2 instead of 1064 nm was primarily driven by the sensitivity/bandwidth of available photon-counting detectors for the respective wavelengths. That said the harsh, high vibration environment with large temperature gradients on an aircraft imposes additional constraints on what instruments can be used. Sophisticated detectors that work in a controlled environment on an optical bench don't necessarily last long in a harsh aircraft environment. More exotic lasers with wavelengths between 532 and 1064 nm have been considered in the past but have not been seen as viable options.*

Since the manuscript deals with signal differences that are close to the precision limit, it might be worth discussing the technical implementation, as well as the post-processing (particularly the algorithms used to determine the slant range and elevation) explicitly, instead of referring to previously published work. This would make the paper more comprehensible and give the reader a chance to better understand the problem at hand, without having to search through additional literature.

> *<u>Response:</u> Our paper uses relative changes in pulse shape as the primary observation on purpose. Previous attempts have tried to use absolute elevations. The referee is correct that elevation biases from differential penetration are comparable in size to the instrument's precision, which is why previous attempts often lack conclusive evidence for green penetration. Learning from these mistakes we have made the case that relative changes in pulse shape observed over different types of sea ice are far more reliable than previous attempts. Therefore, we don't see any value in a detailed discussion of ATM processing here, which as the referee said has been previously*

published (https://doi.org/10.5194/tc-16-3649-2022). As stated in the paper, the strong spatial correlation between pulse width and ice type shown in Fig. 2 is evidence that the observed changes are not related to instrument noise or processing uncertainties (see Sections 4.2.1 and 4.2.2 of the manuscript).

Naively, I would expect the rising edge of the waveform to be most sensitive to timing changes (largest change in signal amplitude upon temporal shift). In addition, would an algorithm using the earliest return photons, i.e. those scattering directly from the surface, not be less dependent of waveform broadening effects? As mentioned above, an explicit discussion of signal post-processing, specifically why the slant ranges are determined via the centroid tracking algorithm will be beneficial to the manuscript, as it is immediately relevant to the problem you are trying to address.

Response: The referee is correct that elevation biases from green penetration are sensitive to the tracking algorithm use. We briefly discuss previous publications on this topic in Section 4.2.2. Our paper focuses on elevation biases in existing data products and is not trying to determine the best possible tracking method, which is well outside the scope of our paper. ***We will add a brief summary of the ATM centroid tracker at the end of the first paragraph in Section 3.2.1 that was previously published.*** *The author has created a Jupyter notebook that demonstrates the ATM centroid tracker in addition to the code and method that were previously published. We will reference that notebook in the revised paper: https://github.com/mstudinger/ATM-Centroid-Tracker and invite the referee to look at it.*

Centroid tracking is a widely accepted standard tracking method in modern laser altimetry and we don't see a reason why we should provide a detailed justification for using a community established tracking method in this paper. The paper that includes the ATM centroid tracking method was more focused on range determination and the handling editor agreed that this level of detailed belonged into the Appendix of that paper (https://doi.org/10.5194/tc-16-3649-2022)

Journals like The Cryosphere use so called similarity reports to detect previously published material in a submitted manuscript to eliminate re-publication. It is our understanding that it is common practice to reference previously published material that is relevant to the submitted manuscript and not repeat it.

The change from using a leading-edge threshold tracker to centroid tracking was motivated by evolving lidars, detectors, and data acquisition systems of the ATM systems over the years and is not relevant to this paper. We emphasize that the focus of the paper is to identify elevation biases in existing data products and not justify their methodology after the fact.

Regarding the previous: If I understand correctly (DOI: 10.1109/IGARSS.2011.6050002) using the rising edge for slant range determination is not invariant with signal integration / photon accumulation. However, for single passes over a water lead (the scenarios described in this manuscript) would it be feasible to use a constant integration length/time for all surface types involved? Would this allow using a threshold tracking algorithm and potentially provide bias free elevation measurements in the present case?

Response: The referee's suggestion is interesting but is beyond the scope of our manuscript. As mentioned by the referee in the next comment, threshold-based leading-edge trackers have in *general higher scattering in range determination in addition to more artifacts in what is called range walk (amplitude dependent variations in range). This is the reason why waveform enabled laser altimeters in general don't use threshold-based tracking methods anymore but more accurate Gaussian or centroid trackers. While leading-edge threshold trackers are less sensitive to subsurface scattering, we don't anticipate an overall improvement in range determination because of the issues we mentioned. Using different trackers over different surface types as the referee suggests will only introduce new biases and worsen the problem.*

Again regarding the previous: According to Yi et al. 2015, DOI: 10.1109/TGRS.2014.2339737 there is a 3 cm precision improvement when using Gaussian or centroid methods compared to thresholding. However, wouldn't the reduced precision be acceptable in light of 10s of centimeter bias over various ice types?

Response: *We appreciate the referee's time to study the paper by Yi et al. 2015. We again refer to our statements in Sections 4.2.1 and 4.2.2 of the manuscript that show that our observed biases are not a result of instrument precision. As a co-author of Yi et al. 2015 and PI of the ATM instrument team I believe that the results of that paper related to range precision are not applicable to the newer 1.3 ns lasers, because of differences in pulse length and fidelity of the laser pulses.* The data presented in our paper are all using centroids for range determination, not a leading-edge threshold.

It might be worth plotting return times on x-axis to allow for visual identification of centroid shift to longer return times (and resulting slant ranges). In addition, I would be interested in a visual comparison of the centroid shift with the shift in a threshold value, e.g. at 50% rise of the leading edge of the waveform.

Response: *The ATM lidars are conically scanning airborne lidars. As a result of that measurement geometry the slant range of each laser footprint depends on aircraft attitude and scan azimuth. Even over a relatively flat surface like the sea ice targets we discuss the slant ranges (or return times) are different from shot to shot and a comparison as suggested would more likely just cause confusion.*

Specific questions:

l. 250ff – Is the ice-type classification simply based on visual analysis of the natural-color images and if so, which parameters and features (brightness, visual layer overlap, …) are used? Have these features in the past been identified and characterized by ground-truth measurements?

Response: *The classification is based on interpreting features in natural-color images that are described in the World Meteorological Organization's (WMO) classification methodology we use (Section 3.1 Natural-color imagery ice type classification). The WMO classification is the accepted standard which for this reason we used in the paper. We will clarify the above sentence in the revised manuscript to make clear that only imagery and not lidar data were used for ice type classification.*

l. 265 – By "[…] classify laser footprints based on their visual appearance […]" do you mean based on their location with respect to the natural-color image? It is not clear, if the classification at this points

is solely based on comparison with the optical image, or if it already involves analysis of the LIDAR waveforms.

Response: See above. We will clarify the sentence in the revised manuscript.

Fig. 2 – Adding a panel showing the corrected LIDAR elevation measurements (result of this work) would be good.

Response: The purpose of our paper is to identify biased elevation measurements in existing data products and provide evidence that these biases are a result of differential green penetration rather than artifacts. Developing and validating a correction is well outside the scope of this paper. The last sentence of the abstract refers to the challenges: "The spatial correlation of observed differential penetration in ATM data with surface and ice type suggests that elevation biases could also have a seasonal component, increasing the challenge of applying a simple bias correction."

Fig. 2 – Can you explain why over the water lead many data points are missing in the center of the scanned track?

Response: The ATM lidars are conical scanning lidars with a 15° off-nadir angle (wide scan) and a 2.5° off-nadir angle (narrow scan). Over specular surfaces like smooth leads most of the energy is reflected away from the receiver for an average 15° angle of incidence, which results in gaps in returns over such areas in the wide scanner. For this reason, ATM has added a second lidar with a 2.5° scan angle to obtain returns over leads for freeboard estimates. Fig. 5b shows this effect. The wide scan (15°) is lacking returns over the lead, while the narrow scan (2.5°) almost continuously tracks the surface of the lead. Because the aircraft used for the campaign for Fig. 2 did not have the necessary space to install two lidars we used a rotated and slightly tilted version of the wide-scanner to increase the chances of getting returns over lead. As Fig. 2 shows the approach was successful but included expected losses over leads. The shot density in the ATM swath is higher near the edge and therefore the chances of getting a weak return increase compared to the center of the scan.

Fig. 2 d): Please indicate the meaning of the two white arrows in the image or in the caption.

Response: We will add a sentence to the figure caption in the revised manuscript. The two arrows refer to lines 291 – 292: "A pronounced change in pulse width related to the transition from single-layer to finger-rafted ice can be observed in area B (marked by two arrows in Fig. 2d)."

Figs. 2 and 3 – The shape of the symbols are hardly discernible.

Response: We have noticed a loss of resolution in the assembled PDF compared to our submitted document. Before submission, we experimented with larger symbol sizes, but there were two issues that let us go with the current symbol size: 1) Larger symbols have more overlap between symbols making it more difficult to discern the shape and 2) larger symbols cover more of the underlying optical image making it more difficult to see the correlation between ice type and lidar footprints. If the paper gets accepted, we assume that the journal will have a version of the figures with high enough resolution to discern the symbol shapes.

l. 265 – Laser footprint(s) sounds like a term for the spatial dimension and distribution of the laser beam on the surface. Maybe in this context LIDAR data points would be a preferred terminology?

<u>Response:</u> The referee is correct that a lidar footprint is the spatial distribution of laser energy on the surface. Within the laser altimetry community this is the accepted terminology, and we therefore prefer to use footprint instead of data point. We will add a sentence to the revised manuscript that better defines the term lidar footprint. We find the term "lidar data points" misleading since the shape of the waveform that we discuss in the paper is a result of how the spatial distribution of laser energy interacts with the surface and subsurface and not a point measurement.

Fig. 4 – Please consider adding standard deviations for the averaged waveforms.

<u>Response:</u> The initial version of this figure included a shaded area for the standard deviations of the waveforms. We think plotting the standard deviations of the water and ice waveforms distracts from the main message of the figure, which is to illustrate the changes in waveforms over sea ice and water relative to the penetration-free calibration target. The busy plots reduce the clarify of the figure, while not adding much relevant information in our view. We therefore decided not to plot the standard deviations to make the figure clearer and easier to understand. We prefer to keep the figure simple and clear, but have included the alternative figure for comparison:

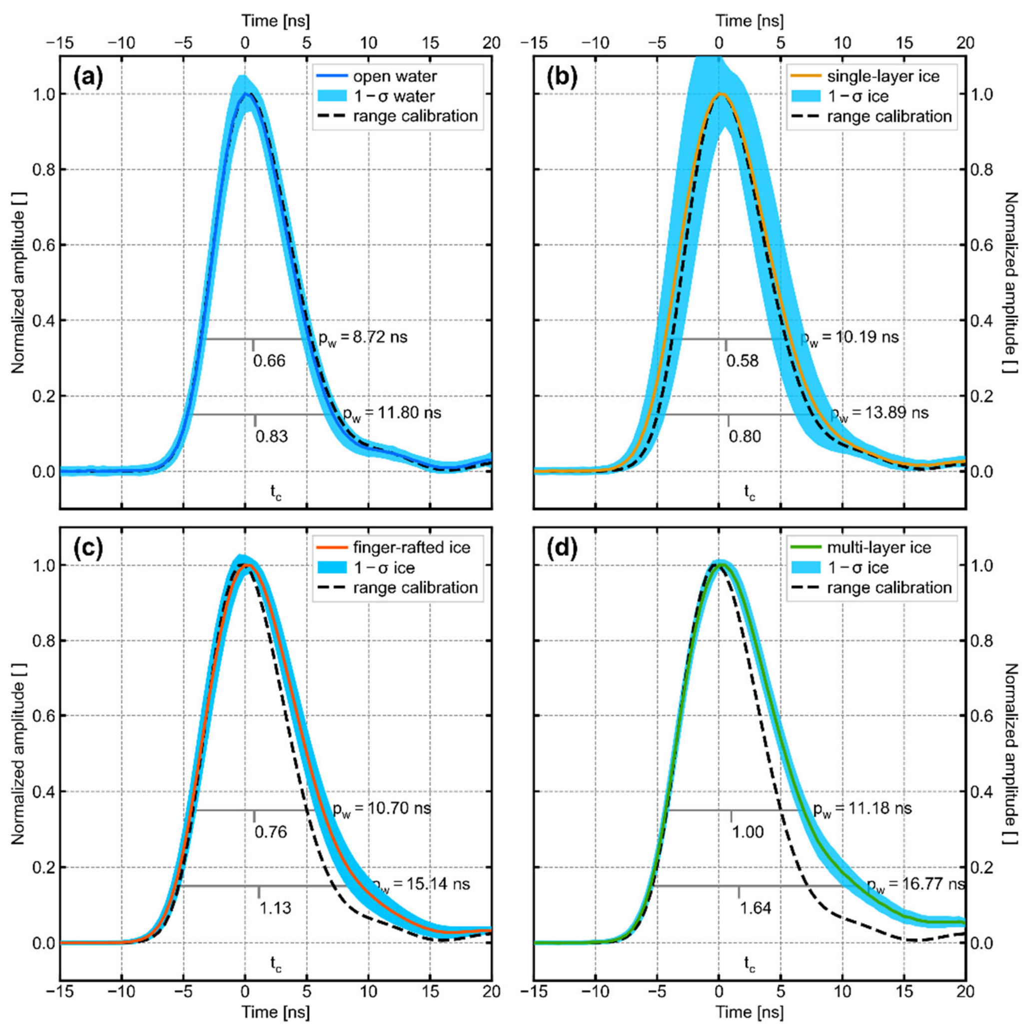


Fig. 9 – You show slant range differences of 0.28 m, however, the elevation bias for single layer thin ice is only 0.1 m. Is there a minimum distance that can be resolved in terms of surface and bottom return pulses, before the return pulses coalesce?

> *Response: The method we used to determine the 0.28 m range difference for Fig. 9a is described in Studinger et al., 2012 (https://doi.org/10.5194/tc-16-3649-2022). Using synthetic Gaussian waveforms, we estimated the minimum water depth to be on the order of 30 cm, but shallower depths can be resolved depending on overall return signal strength (see Fig. B2 in Studinger et al., 2022). We will add a sentence to the revised manuscript stating that 28 cm is near the detection threshold and refer the reader to Studinger et al., 2022 for more detail.*

Related to the previous: Could you distinguish broadening due to volume scattering from the scenario in which one pulse is reflected from the water surface and a second from the ice, if the ice were submerged by only a few centimeters?

> *Response: We have in fact tried that for the data shown in Fig. 7, but the results were inconclusive and therefore we have decided not to include them in the manuscript. We believe the bandwidth limitation of the 6 ns laser used in this case may have contributed to the poor results.*

l. 334f – Would broadening of the return waveform over water, possibly due to sub-surface scattering induced by turbidity or the presence of submerged particles, thwart the efforts to find a universal range bias correction, as the reference signal for zero elevation would change?

> *Response: Subsurface volume scattering from turbidity, sediments, or algae could have an effect similar to volume scattering in ice and potentially adding to the complexity of developing a bias correction. We can add a sentence in the discussion of the revised paper that mentions these additional complications. The vast majority of data over sea ice is collected away from the coast in very clear water and the effect should be rather rare. Algae blooms will have the additional complication of a seasonal component.*

l. 330 – I believe the reference should be to Fig. 4 a) and not to Fig. 4 b).

> *Response: Agreed. We will correct the reference to Fig. 4a in the revised manuscript.*

l. 331 – "The shift in centroid […] is negligible." If the shift in panel a) for water is negligible, then the shift in panel b) for single-layer ice (0.58 and 0.8 versus 0.66 and 0.83) also seems negligible.

> *Response: Agreed. We will rephrase this sentence in the revised manuscript. The changes in waveform of the single-layer ice are very subtle and the shift in the centroid position is sensitive to how the waveforms are aligned. We use the maximum as reference here. The changes in pulse width from open water to single-layer ice are however pronounced.*

l. 333f – "… most of the laser light is reflected away from the receiver …" I don't think this statement is correct, because at the mentioned incidence angles less than 10% of the light will be reflected (specular) by the surface. Most will be refracted and enter the water (and be absorbed in the absence of scattering). Either way, the return signal strength will be very low.

> *Response: Agreed. We thank the referee for pointing this out. We will replace the statement in the revised paper with a more neutral sentence that accounts for both effects without getting into the*

details of how much energy gets reflected and refracted, since as the referee correctly points out, the overall effect will be a low return signal strength. Suggested statement: "Depending on the angle of incidence, on a specular water surface some of the laser energy will be refracted away from the receiver and some will be refracted into the water, resulting in weak return signal strengths."

l. 339 – The main text does not discuss the data presented in Fig. 4 b).

Response: Agreed. We will include a short discussion of the data in panel Fig. 4b) in the revised manuscript along the lines described in our response to the referee's comment regarding our statement in line 331.

l. 368f – I believe the figure reference should again be to Fig. 4 a), since the sentence discussion the open water case.

Response: Agreed. We will correct the reference to Fig. 4a in the revised manuscript.

4.2.3 – You state that roughness and slope broaden the waveform symmetrically, while sub-surface scattering leads to asymmetric waveform broadening. Yet, the broadening for single-layer ice in Fig. 4 b) seems rather symmetric than asymmetric, when compared to the range calibration waveform. Can you comment on this?

Response: This is the same issue that we have mentioned in our response to the referee's comment regarding our statement in line 331. The way subtle changes in the shape of the waveform are sensitive to how the waveforms are aligned. The more pronounced observation here is the change in pulse width. We will add a brief discussion in the revised manuscript to clarify the observations.

l. 440ff – Could you confirm you hypothesis that part of the ice is flooded by calculating the NDWIice for the image in Fig. 7? Since even the shallow edges of melt ponds in Fig. 8 b) show a clear NDWIice signal, wouldn't this be applicable to the flooding case as well? I am assuming the flooding is only by a few centimeters.

Response: We have followed the suggestion of the referee and have calculated $NDWI_{ice}$ for the data in Fig. 7, but the results were inconclusive, and we have decided not to include them in the paper. The reason is that $NDWI_{ice}$ is not intended to be used for situations like the one shown in Fig. 7 nor is it suitable. A threshold based $NDWI_{ice}$ classification can be a powerful tool for identifying supraglacial lakes (https://doi.org/10.5194/tc-16-3649-2022) and sea ice melt ponds (Fig. 8) but we don't expect a discernible spectral contrast between ice submerged below a few centimeters of water and unsubmerged ice. A further challenge of using a threshold based $NDWI_{ice}$ classification for these kind of situations is the arbitrary choice of a threshold for instruments like DMS and CAMBOT which are not radiometrically calibrated but are passive instruments that use sunlight as the source of illumination (https://doi.org/10.5194/tc-16-3649-2022). In lines 445 to-446 we point out that the well-defined shadows of small pieces of ice embedded in the ice in the bright areas in Fig. 7b and c (marked with arrows) indicate that the surface of the ice is not submerged and above water level (Fig 7b). Also, in lines 447 – 449 we state: "The strongest evidence for flooding comes from a small, rafted piece of ice in Fig. 7c (marked with an arrow) that appears to bend the underlying ice with its load." We believe these two observations strongly support our interpretation that part of the ice in Fig. 7 is indeed flooded.

Fig. 10 – Consider adding a plot with corrected LIDAR elevations. Are you able to verify the corrected elevations via ground-truth measurements or other means?

Response: The goal of our paper was not to develop a correction for the biased measurements, but rather show and proof for the first time that these biases exist in available data products and are caused by differential penetration of green lidar light. The general lack of ground-truth measurements over our target areas is part of the reason why it has been so difficult in the past to identify these biases. The model estimates in Fig. 10 are aimed at helping to support the interpretation of our observations. There are issues still to be overcome related to turning them into corrections; these issues are described in the companion modeling paper that is now in review in The Cryosphere Discussions (https://doi.org/10.5194/tc-2023-147). We will refer to the modeling paper in the revised manuscript and will note that the correction discussed in the modeling paper is likely not adequate to correct sea-ice elevations.

Fig. 11 – Maybe you could include the waveform centroid in addition to the calculated bias. Am I assuming correctly that for the thin ice case, the centroid lies between 0 and the height bias value? In that case – since the centroid is a measure for elevation – showing the centroid values would highlight the main message of the manuscript.

Response: This is an excellent point. We have revised Fig. 11 to address comments from both referees. A detailed explanation is given in our response to RC2's comment re. Fig.11 on page 17 of this document. Instead of plotting the elevation bias from the model estimate we have switched to showing the elevation bias estimate determined from the centroid of the observed waveform relative to the model surface.

Reference MacGregor et al. 2021a has duplicate 2021b.

Response: We will remove the duplicate reference.

Reference Kurtz et al. 2013a and 2013b are duplicates.

Response: We will remove the duplicate reference.

Response to the Referee 2 comments (RC2) manuscript TC-2023-126 https://doi.org/10.5194/tc-2023-126-RC2:
Estimating differential penetration of green (532 nm) laser light over sea ice with NASA's Airborne Topographic Mapper: observations and models

We thank the referee for the many positive comments and detailed suggestions to improve the manuscript. We will revise our manuscript accordingly.

General response regarding use of color palettes: I fully share the concerns of the referee regarding color pallets that are accessible for people with a color vision deficiency (CVD). I also appreciate the efforts by Copernicus to publish color schemes that are more accessible for readers with CVD. My own color vision is significantly reduced by approximately 70-90% compared to an un-impaired color vision. My own color deficiency does not fall into any of the usual categories (e.g., red/green etc.). The color schemes I used for the manuscript minimize the range of colors that are indistinguishable for me and are color schemes that work best for my impaired color vision (including the much-criticized rainbow scale). I would prefer to use color schemes that work for me and have explained that in a previous paper published in The Cryosphere: https://tc.copernicus.org/preprints/tc-2022-78/tc-2022-78-AR1.pdf

Proposed solution: As a person with CVD I find the statements made by Crameri et al. (https://doi.org/10.1038/s41467-020-19160-7 and https://doi.org/10.5194/gmd-11-2541-2018) frustrating since there is no single solution that works for all types of color deficiencies. The examples they used in their papers are generally models that lack fine-scale detail. Data sets like that could easily be illustrated with a gray scale. Preparing the manuscript, I have tried (again) to use the Cameri et al. color map module for Python (https://zenodo.org/records/8409685) and found the results (again) unusable for my color deficiency.

Referee 2 and I obviously seem to have very different forms of CVD. To make some of the color figures in this paper accessible to more people I suggest using the color schemes currently in the paper and provide versions with CVD "friendly" color palettes following the referees suggestions in the Appendix. This will incur additional page charges that I am willing to support. We defer the decision to the handling editor, but personally I feel with digital publishing this should be the standard. We have provided two alternative versions of Fig. 8 below that incorporate the referee's suggestions. However, both versions have issues that we discuss below.

Review criteria:

Are the scientific methods and assumptions valid and clearly outlined?

> **RC2: Partly. The applied methods are valid and support the findings. But I miss a bit more details of the scattering model or explanations of the process why the scattering length is very long especially in thin sea ice and not in dry snow. Is this related to grain size or density? What are the main drivers of long scattering length? As already mentioned above I certainly miss an approach to correct the observed bias by using another technique to estimate the range from the returned pulses. Also the ICESat2 data analysis can be improved (see below).**

Response: We have responded to each of the topics (details of the scattering model, bias correction, ICESat-2 analysis) below.

Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

RC 2: Partly. I miss more details of the scattering model.

Response: We have responded to details of the scattering model below.

Does the abstract provide a concise and complete summary?

RC 2: Yes, I find it a bit too long and it would be good to mentioned that large range bias (or penetration) of tens of cm is found over thin sea ice and not over dry snow,

Response: We will add a statement regarding dry snow to the revised abstract. We defer the decision regarding abstract length to the handling editor and publisher.

Is the overall presentation well structured and clear?

Need to be improved. I find the paper sometimes very technical and full of details, which might be shortened and restructured. E.g. chapter 2.2 Natural-color optical imagery is very detailed (is this needed for understanding?) 3.2 Lidar data characteristics is placed under methods but would fit better to 2.1 or 4.2 here a discussion is already started in the results chapter

Response: We feel the level of technical detail in the paper is necessary. This is based on our experience with members of the altimetry community that initially believed the observed elevation biases were instrument or tracking artifact. The level of technical detail is necessary to rule this out. Since the lidar characteristics are directly related to the observations of pulse broadening we feel they belong into section 3.2 rather than 2.1. The discussion section of a research paper typically discusses the results in a broader context, which is the title and contents of our Section 7. Section 4.2 "Discussion of possible reasons for observed changes in elevation, pulse width, and pulse shape" is far more specific and belongs into Section 4 in our view. We can omit "discussion" from the Section 4.2 title and change it to "4.2 Possible reasons for observed changes in elevation, pulse width, and pulse shape" if that helps clarify the structure of the paper.

General:

Despite the paper is clear and of high quality I miss an assessment to improve the laser range detection to minimize the penetration effect of green laser light using different tracking approaches. This would be very beneficial and could guide an improvement of the ICESat2 range retrieval.

Response: The purpose of our paper is to identify biased elevation measurements in existing data products and provide strong evidence that these biases are a result of differential green penetration rather than artifacts. Evaluating the performance of different alternative tracking algorithms is well beyond the scope of our paper. Both referees believe that changing tracking algorithms could minimize or eliminate the range bias (see also our response to RC1 on page 3 of this document.

While different tracking methods have different biases and could potentially be used to minimize elevation biases as the referee pointed out, we don't believe a different tracking method could provide a satisfactory solution or correction to the problem, which is why it hasn't been implemented in ATM or ICESat-2 data products a long time ago. In our view a solution needs to eliminate biases using a validated methodology rather than just minimizing biases.

It should also be noted that tracking algorithms applied to a waveform recording lidar such as ATM cannot simply be transferred to a photon-counting lidar such as ICESat-2/ATLAS or be used to guide improvement of ICESat-2 surface tracking. There are fundamental differences between range determination for waveform recording lidars such as ATM and photon-counting lidars such as ICESat-2.

10.1109/IGARSS.2011.6050002 already showed that green laser tends to widen the pulse leading to an offset of elevation estimates especially for the centroid method. In this paper the findings of 10.1109/IGARSS.2011.6050002 are confirmed by more and better use cases, but the opportunity to improve the elevation product with the help of the scattering model is not taken.

Response: Our paper has a far more robust and comprehensive analysis of green penetration compared to the extended conference abstract referenced above. The purpose of our paper is to identify biased elevation measurements in existing data products and provide evidence that these biases are a result of differential green penetration rather than artifacts. Developing and validating a correction is well outside the scope of this paper. Our initial submission included a companion modeling paper that was removed by the editorial staff. The paper is now in review in The Cryosphere (https://doi.org/10.5194/tc-2023-147) and shows the complexity that is involved in developing and validating a bias correction for the case of land ice. Developing a bias correction for sea ice would be equally complex and require a separate paper.

Therefore, I would recommend, that the authors apply next to the centroid a threshold (e.g. TCOG or TFMRA of the leading edge) or a combination of a gaussian fit and threshold tracking for comparison. The best threshold (giving highest precision) could be first evaluated using simulated returns or by using a subset of open water lead returns. This could be compared to the precision of the centroid tracker. Then this threshold tracker can be applied to all data and compared to the centroid estimates over different sea ice regimes.

Response: We agree with the referee that the suggested analysis would indeed be very helpful for better understanding differential penetration biases in ICESat-2 data products assuming the referee refers to ICESat-2 data. However, our intention for showing Fig. 14 is simply to demonstrate that biased elevations likely exist in lower-level ICESat-2 products and probably deserve more attention. A meaningful comparison of different tracking algorithms needs to be thorough and therefore requires significant space in the paper. A major limitation for the suggested analysis is the lack of coincident optical imagery for ICESat-2 data, which is needed for reliable surface classification. Additionally, the approach used by the ICESat-2 data products is physics-based which is quite different and not easily compatible with threshold or Gaussian fit tracking methods. However, the suggested analysis could be a follow-up project for interested readers who want to build on our work. While we feel the suggested analysis is well beyond the scope of our paper, it is certainly worthwhile doing.

In addition, I would recommend that the authors try to understand the ICESat2 negative freeboard in more detail by using lower level ATL03 photon data set. Here, they also could apply the threshold tracker on the photon distribution or a gaussian fit of the photon distribution and evaluate if the freeboard can be corrected by maintaining the same accuracy. For the sea ice community this would be a very important finding.

> *Response: We agree with the referee's suggestion that the ATL03 photon data are key to further advance understanding of the negative freeboard in ICESat-2 data. We have attempted to use a Python version of the ATL07 surface tracker on negative freeboards and restricted the Gaussian width of the fit to assume a flat surface of < 2 cm. The results of our preliminary analysis were inconclusive and suggest that in order to improve the analysis a more advanced retrieval algorithm which incorporates the sub-surface scattering physics is required, which is beyond the scope of this paper. This and the previous suggestion by the referee would warrant a publication on its own in our view.*

Figures:

Please use CVD conform color palettes (https://www.nature.com/articles/s41467-020-19160-7?s=09). I can hardly see differences between green, orange or red colors in most of your figures!

> *Response: See our general response to color pallets in the introduction.*

Fig1, Fig2, Fig 7. Can you please enlarge the symbol size of the laser points. Hard to see.

> *Response: We have noticed a loss of resolution in the assembled PDF compared to our submitted document. Before submission, we experimented with larger symbol sizes, but there were two issues that let us go with the current symbol size: 1) Larger symbols have more overlap between symbols making it more difficult to discern the shape and 2) larger symbols cover more of the underlying optical image making it more difficult to see the correlation between ice type and lidar footprints. If the paper gets accepted, we assume that the journal will have a version of the figures with high enough resolution to discern the symbol shapes.*

I would recommend that you provide next to Fig2, Fig5, Fig7 where you show the elevation also a figure with the elevation distribution w.r.t to open water as histogram for the each of the different ice types for each of your selected sites.

> *Response: We don't think a histogram distribution is relevant or needed for our discussion. Fig. 3 already shows the standard deviations of the elevations for each surface type. The profile view in Fig. 5 is a visualization of the elevation distribution.*

Please also use in all figures, which show the elevation the same CVD friendly color scale and please always plot the elevation wrt. open water.

> *Response: See our general response to the use of color palettes and the two alternative versions of Fig. 8 discussed below. We refer to our response to Fig. 7 below why we need to use WGS84 as elevation reference for Fig. 7 rather than the averaged water surface.*

Fig 4 Please enlarge the font size of axis labeling and text

Response: We will enlarge the font size in the revised manuscript. We have noticed a loss of resolution in the assembled PDF compared to our submitted document that makes it more difficult to read the labels and text. If the paper gets accepted, we assume that the journal will have a version of the figures with high enough resolution to discern labels and text.

Fig3 and Fig4: I don't understand the connection of both figures or at least they are not showing what is explained in the text.

Response: We will make the wording clearer in the revised paper. In Fig. 3 the relative changes are reduced to a single parameter, the pulse width, which can be plotted on a map (Fig. 2d) or in a scatter plot (Fig. 3). Figure 4 examines the relative changes in waveform shape in detail using the entire waveform. This shows additional evidence for differential penetration in the asymmetric changes in the tail of the waveforms that can't be seen in Fig. 2 or 3.

In line 306 you write "The mean pulse broadening w.r.t. the mean pulse width over open water is 0.6 ns for the single-layer ice and 2.1 ns for the finger-rafted ice". Fig4 shows 8.72ns pulse width for open water and 10.19ns for single layer ice at 35% amplitude threshold. The difference is 1.47ns! The difference for finger rafted ice is 1.98ns. At the same time the centroid position for open water/calibrated range is 0.66ns. When this is taken as zero elevation then the difference for single layer would be: 0.58 – 0.66 = -0.08ns and for finger rafted 0.76 – 0.66 = 0.1ns.

With the velocity of light these transfer in a negative range bias of 2.3 cm for single layer and a positive range bias of 3 cm. However, this is not reflected in Fig3. What I'm doing wrong?

Response: The referee points out an apparent inconsistency that we need to address in the revised paper. In the revised paper we will provide more detail how the pulse widths and relative centroid locations were calculated to help the reader better understand what they reflect and how they can be interpreted.

Figs. 2 and 3: The pulse width is estimated from individual waveforms. First, the noise floor is estimated using the median of the first 21 samples of the waveform. Then the noise floor is subtracted from the waveform amplitudes and the 35% threshold of the maximum amplitude is calculated. Then, the time of the 35% threshold value is calculated using linear interpolation between the first sample below the amplitude threshold and the first sample above the amplitude threshold. This is done for both the leading edge and the tail of the waveform. The time difference between these two points is the pulse width at the desired threshold. The method has been described in a previous paper, which is referenced in our paper (https://doi.org/10.5194/tc-16-3649-2022), and the MATLAB® code (atm_centroid_tracker.m) that has been used to do this has also been previously published and is referenced in our paper (DOI 10.5281/zenodo.6341229).

*Fig. 4: The **averaged pulse widths** shown in Fig. 4 are calculated in a different way and therefore slight differences between Figs. 2 and 3 and Fig. 4 exist and should be expected. For Fig. 4 the amplitudes of the individual waveforms are first normalized and then aligned using the delay estimated from the maximum of a cross-correlation between signals (MATLAB® Signal Processing Toolbox's alignsignals function). Because the maximum of the cross-correlation does not always align with the maximum signal amplitude the maximum of the stacked and averaged waveforms is slightly below 1.0. In order to enable a consistent comparison between the calibration*

target/ground test waveforms and the airborne waveforms over ice and water, the averaged waveforms need to be normalized again. The pulse widths of the averaged waveforms at a desired threshold are then calculated as described in Figs. 2 and 3.

*The motivation to label the plots in Fig. 4 with **relative centroids** was to illustrate how the pulse broadening in the lower tail results in centroids shifting to longer ranges and therefore lower elevations. In order to illustrate changes in pulse shape between the calibration and ice targets we have aligned the averaged pulses using the maximum amplitude as a zero reference. The way of how to best align the calibration and airborne waveforms can certainly be debated. After some experimentation we felt that using the maximum amplitude as a reference point best illustrated that the changes in waveform shape over ice and water targets primarily occur in the tail of the waveform. Because of these changes in pulse shape, using a cross-correlation to align waveforms would be inappropriate. As a result, the centroid locations shown in Fig. 4 are relative to the aligned pulses and should not be interpreted as an absolute reference. If the referee finds this confusing, we suggest to remove the locations of the relative centroids in Fig. 4 and only presenting the pulse broadening, which is the primary observable used for our paper.*

***In conclusion:** Despite small, expected differences between the different ways to calculate pulse widths of individual and averaged waveforms we believe the results presented in the paper are consistent and confirm the robustness of our analysis. We are confident in our results and believe Figs. 2-4 in the paper are compelling illustrations of observed differential penetration. However, as mentioned above, because of the comments from the referee we realize that we need to expand on the description of how pulse widths and centroids are calculated and how they should be interpreted.*

Fig 7 Why you show the elevation wrt. WGS84? With different tidal states this is changing and leads to confusion! Please refer elevation to open water as you do in the other figures and please enlarge symbol size.

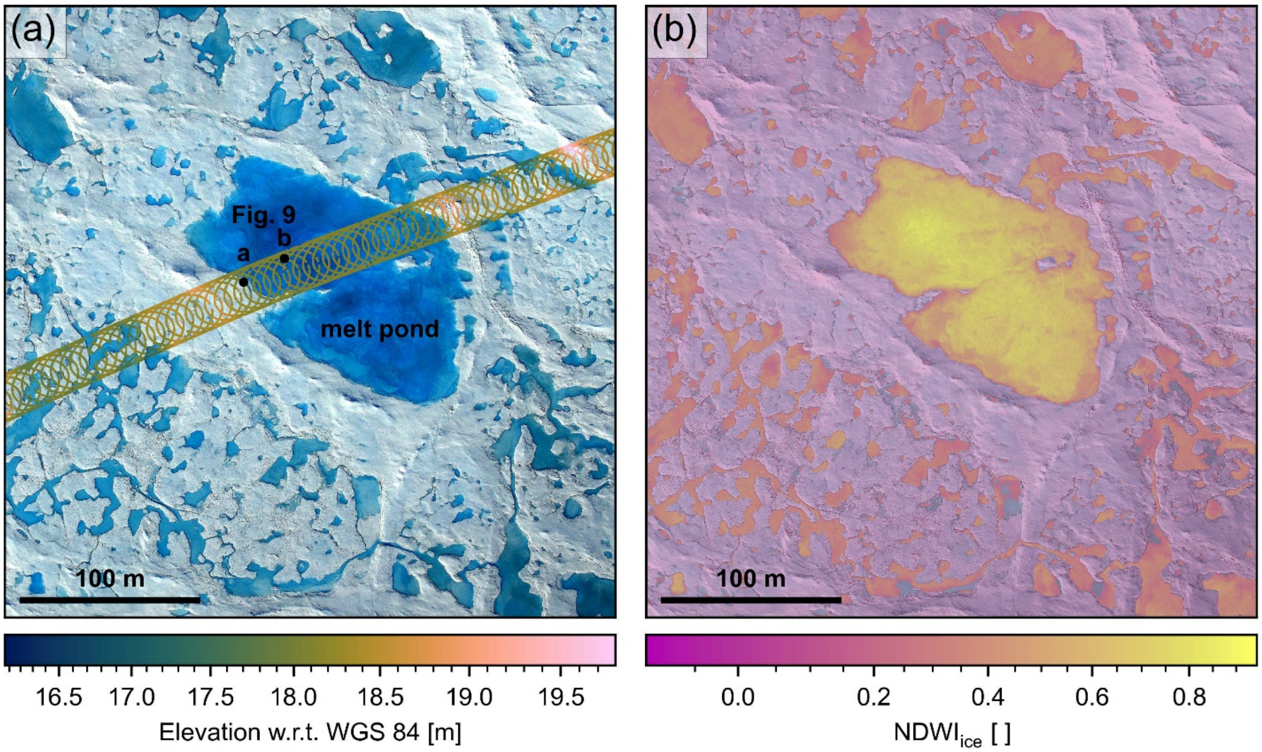
Response: *It is necessary to use the WGS84 as elevation reference here to support our statement in lines 450 to 451: "The change in surface elevation h_w of the open water between passes indicates gentle wave action that could result in flooding of the marginal ice areas." We have responded to symbol sizes in the comments to Fig1, Fig2, Fig 7 above.*

In line 443 you mention that the surface brightness of the ice north of the lead changed. However, to me it seems that also the ice above water (upper left area) is darker in panel (a). Can you please verify if you use the same gray scaling for each of the images?

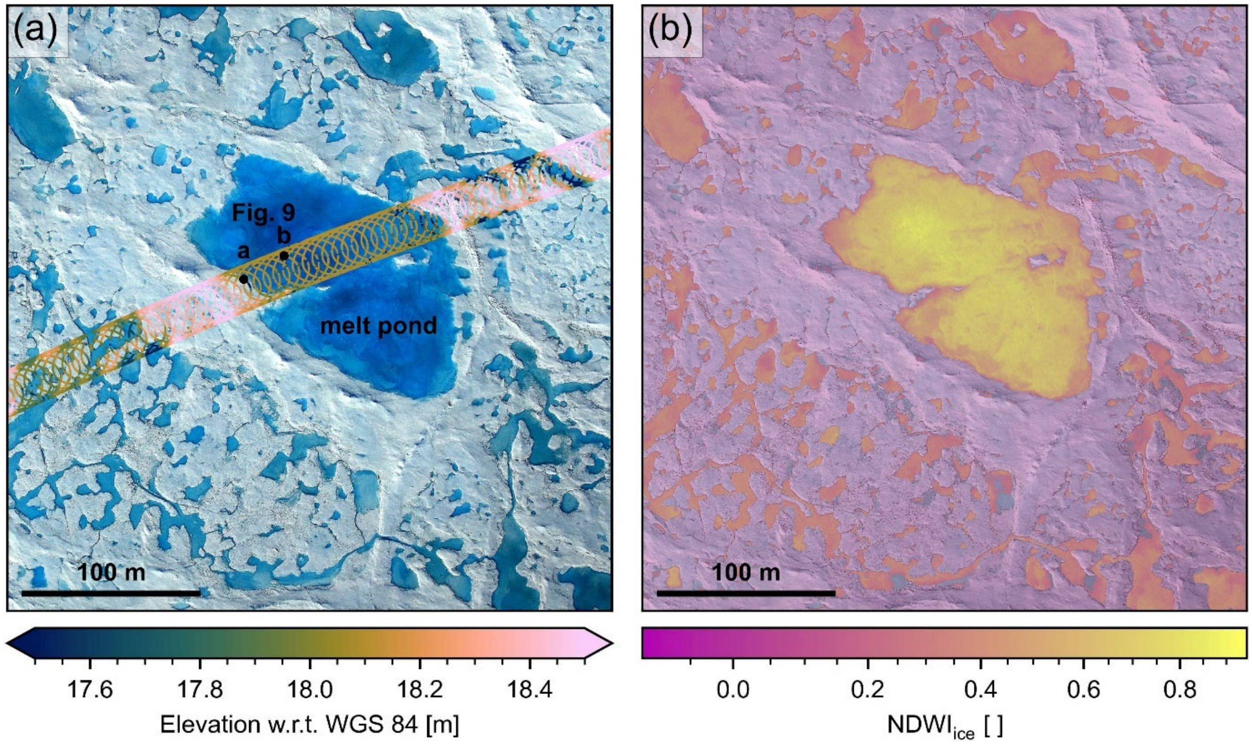
Response: *The backgrounds are image mosaics of natural-color (RGB) DMS images and there is not color or gray scale involved in plotting them. Neither DMS nor CAMBOT are radiometrically calibrated instruments (e.g., https://doi.org/10.5194/tc-16-3649-2022) and rely on illumination of the target by natural sun light. Lines 166 – 170: "DMS and CAMBOT are both passive instruments that depend on natural sunlight for illuminating the area within the field of view (FOV) for imaging. Illumination depends on sun angle and cloud cover and often varies considerably during a flight. During sea ice missions the CAMBOT operator adjusts exposure parameters such as shutter speed, aperture, and sensor sensitivity (ISO number) to minimize motion blur and optimize exposure for the dynamic range of the camera sensor (Studinger et al., 2022b)." For this reason, the color or* *brightness between frames should not be directly compared. Furthermore, unlike the CAMBOT system, the DMS system has a strong vignetting effect that darkens the frames toward the edge.*

Fig 8 Color scale! All looks the same for me. Maybe you narrow the min/max elevation as well. Maybe you zoom in even more so that you really focus on the lake.

Response: *We have followed the referee's suggestion and created two versions of Fig. 8 that use the Cameri "batlow" color palette for panel a) and the "buda" palette for panel b).*



I personally don't see an improvement in panel a) using the batlow color palette. I can't distinguish colors between 17 and 19 meters. Part of the problem is the linear color scale over the entire data range, which we will discuss in version 2 below. Panel b) shows an aspect of color palettes that is completely ignored in the Cameri et al. papers: plotting a transparent color map over a background image. We feel the version in our manuscript show more details and brings out the water surfaces more clearly.



For the second version we followed the referee's suggestion and clipped the color range between 17.5 and 18.5 meters. It brings out more detail but introduces a new problem that also lacks consideration in the Cameri et al. papers: the higher elevations over the pressure ridges are very difficult to distinguish from the background image. We have tried several of the Cameri color pallets which all have that problem because they are very similar towards the maximum.

Fig9. Please enlarge font size. In the figure caption you talk about slant range. Why not add the slant range as additional x-axis label?

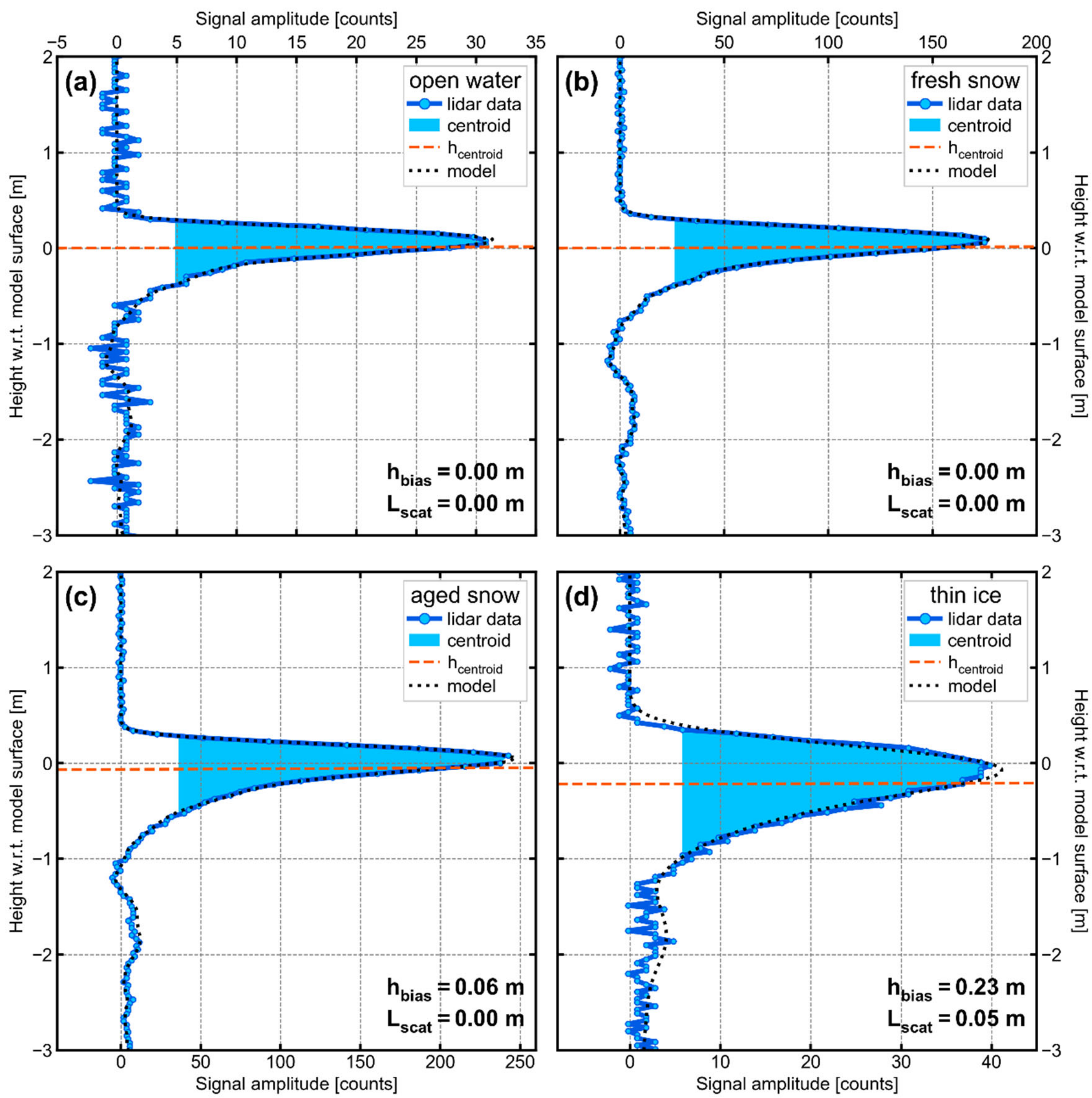
Response: Agreed. We will increase the font size in the revised paper. We feel adding a second x-axis to the plot introduces unnecessary complexity. There is only a single slant range value for each panel that does not justify adding a second x-axis in your view.

Fig.11 I don't understand the positioning of the hbias in 11(d). Is this the estimated centroid which gives the range? When looking at the waveform it seems that the hbias when this is representing the centroid should be closer to zero. In the text Line 340 you write that only small changes in the pulse shape are visible in the leading edge but figure 11d shows a clear widening when compared to 11(a). Can you comment on this?

Response: Based on the comments from both referees regarding Fig. 11 we have revised the figure and will expand our explanation in the revised manuscript. See also our response to RC1's comment on Fig. 11 on page 9 of this document.

The h_{bias} values in the initial submission are estimated using the model surface. Since RC1 suggested adding the centroid and RC2 also refers to the centroid we switched to plotting the range bias estimate determined from the centroid of the observed waveform relative to the model surface. The red dashed line in the revised Fig. 11 below is now the location of the centroid of the *observed waveform relative to the model surface elevation (we have also changed the x-axis labels to "Height w.r.t.* ***model*** *surface [m]" to make that clearer). The h_{bias} values now shown are the centroid-based range bias from the observed waveforms (red-dashed lines) corrected for the angle of incident and therefore reflect the estimated elevation bias relative to the model surface. We would like to point out that the h_{bias} values based on the model estimate shown in the initial submission, and the h_{bias} values from the observed waveform centroids shown in the revised figure match within a few millimeters. We believe this confirms that our analysis and estimates are consistent and reliable.*

Line 340 describes the waveform shapes in Fig. 4. The general statement that pulse broadening occurs primarily in the tail is true for all cases presented in the paper. We don't see the effect that the referee describes. The pulse in Fig. 11d is indeed significantly broader than in Fig. 11a as expected, but we don't see the broadening primarily happening in the leading edge as described by the referee.



L510 ff: Based on your modeling approach. Could you please briefly summarize what parameter drives the most significant change in scattering length. Is this density, grain size or temperature? At which grain sizes you see a change? Is this a linear or abrupt change? As shown for dry snow you expect little penetration or at least little effect of subsurface scattering on range estimates. Is this a valid assumption for green laser penetration over dry snow in general (e.g. is this valid over the whole Greenland and Antarctic ice sheet)? Maybe you can add a line in chapter 7 when you discuss the broader context of green laser light penetration in snow and ice as this is also important for the land ice community when they compare ICESat2 with radar altimetry.

Response: All other things being equal, the scattering length should be roughly proportional to the grain size and inversely proportional to the density, but because density varies only over a factor of a few between snow and ice while grain size varies over a few orders of magnitude, the largest factor driving variability in the scattering length in most cases is grain size. We will add a line to section 7 discussing this general issue, and pointing to the companion paper where the problem is discussed in more detail (https://doi.org/10.5194/tc-2023-147).

L695 Here it might be worth to accumulate all ATM returns to match the footprint size of ICESat2 and check if the differential penetration is still existing in the ATM data.

Response: Differential penetration is a physical interaction of the laser light with the surface and subsurface and therefore does not average out or disappear with averaging. Fig. 4 shows averaged waveforms that have a significant bias caused by differential penetration.