tc-2023-147

General Comments

This manuscript presents a study aimed at predicting biases in ICESat-2 surface elevations due to the subsurface scattering of photons. The authors base their approach on the comparison of airborne laser waveforms acquired with the Airborne Topographic Mapping (ATM) system and grain size estimates from the Airborne Visible/Infrared Imaging Spectrometer – Next Generation (AVIRIS-NG) instrument and the Ocean and Land Colour Instrument (OLCI) on-board Sentinel-3. A subsurface scattering model is employed to derive grain size estimates from the ATM waveforms, which are then compared to the AVIRIS-NG and OLCI observations. Finally, these results are to predict possible biases in ICESat-2 elevations due to estimated grain sizes.

Overall, I found the manuscript very confusing and difficult to read. Many of the comments below will speak more to this point, but there is general a lack of specificity and clarity in how the manuscript is written. I think much of the confusion with regards to the scope of the manuscript comes from how it is initially presented as "Understanding biases in ICESat-2 data …" when there is no actual ICESat-2 data used. With how often ICESat-2 is mentioned in the Abstract and Introduction, the authors build an expectation that at some point what is discussed will be applied back those data, but it never actually happens. Also, as is stated on line 586, to date no biases in ICESat-2 due to subsurface scattering have been observed, which brings into question the overall relevance of the study. Are the authors proposing a solution to a problem that doesn't exist? The broad-scale importance of this work is not clear.

I think the manuscript would benefit from a thorough revision focused on improving context and clarity. I believe the foundational elements (i.e., approach, data analysis, results) of the study are interesting and worthy of publication, but how they have been packaged makes the full arc of the study difficult to follow. I also found the manuscript to be under-referenced. The authors often make direct statements without providing any supporting evidence.

I recommend the authors consider the following points while revising their manuscript. I realize there are a lot of them, but I want to be detailed and clear with regards to where I had trouble following the manuscript to help the authors with their revisions.

Specific Comments

- Abstract: Please provide more context for the last sentence. Why is it important to correct for subsurface scattering biases in ICESat-2 data? Within the context of the overall ICESat-2 height uncertainties, how big of a problem is it if these biases are left unaccounted for?

- Line 31: Please consider including the corresponding wavelength for ICESat to support the comparison to ICESat-2's 532 nm green laser.

- Line 47: Here weaker scattering is implied as being associated with broader returns and larger delays, but wouldn't weaker scattering be equivalent to stronger absorption in the medium? In the second and last sentence of the paragraph, stronger absorption is associated with tighter photon distributions.

- Line 50: Someone approaching the manuscript from an ICESat-2 (i.e., individual photon) perspective may be unfamiliar with the concept of a laser "waveform". I'd suggest the authors early in the manuscript define what a waveform is. Here a "return photon timing distribution" provides an excellent opportunity.

- Lines 70-72: The process of using a model for the shape of ATM waveforms to predict grain size and comparing that to airborne and satellite measurements is well laid out. What I continually struggle with is the subsequent extension to ICESat-2. What does it mean to "confirm" that predicted biases in ICESat-2 are the same as those for the ATM case when there are no actual ICESat-2 biases to compare against? Is the authors goal to show 1) that their scattering model is valid when using airborne (ATM and AVIRIS-NG) data, 2) can be applied to Sentinel-3 gain sizes to predict a generic laser bias, and 3) that generic bias compares favorably against ATM data, so that, in principle, one could possibly apply it to ICESat-2? So much of the emphasis in the manuscript is placed on applicability to ICESat-2 but by the end, that link seems like more of an inference/prediction/extrapolation than something that is concretely demonstrated (i.e., confirmed). Furthermore, I suggest the authors be more specific with what they mean by the "sampling" of altimetry measurements and how that affects calibration. What type of "sampling" are the authors referring to?

- Section 2: I know maps are presented in Figure 5, but I would recommend the authors consider including a composite overview map when describing the datasets to help situate the reader.

- Line 96: For those unfamiliar with the ATM data, please provide some specific detail as to what wide-swath and narrow-swath means (e.g., cross-track look angle, width on ground, etc.).

- Line 105-108: Here the authors refer to some analysis they have done that underlies the choice of impulse response function but isn't presented in the manuscript. I would recommend the authors consider including this so the reader can understand why these choices were made.

- Section 2.2: The authors include a lot of detail regarding the ATM system but almost none for the AVIRIS-NG system. How does this system operate? What does it measure? How big is the field-of-view? How closely did the Basler follow the aircraft with the ATM? I believe these details will help to provide context and clarity for the reader.

- Line 119: Please provide some example numbers for how much wider the AVIRIS-NG swath is compared to the ATM swaths.

- Line 122: The authors refer to AVIRIS-NG data with grain sizes inconsistent with the expected surface conditions that were not included. I would recommend being more specific regarding what evidence exists to suggest that these measurements are erroneous. Otherwise, it could leave the impression that the included data were cherry-picked.

- Line 138: I find this description of the "look-up table" and especially the "time steps" confusing. Can the authors simplify to something along the lines of "<time-specific> grain sizes are interpolated from the daily maps."? It is also not clear the degree of interpolation the authors are performing using the daily maps. Are the authors interpolating at sub-daily intervals or just over missing time periods due to clouds?

- Line 145: What assumption are the authors referring to here and why does in only apply to "most" of the data used in the study? What about the remainder?

- Section 3: I think there is something missing from this section, specifically how τ^* related to r_{eff}. r_{eff} does not appear in any of the provided equations so it makes it very hard to follow the development of the waveform model as well as understand Figure 1b.

- Figure 1a: I am confused by Figure 1a. What is the point of introducing the constant velocity (orange) line? It seems to be to try and communicate the how sensitive τ^* is to the effective velocity of the medium, but in doing so does it not introduce a physically unrealistic scenario? That is, one where density is fixed as it relates to velocity but still allowed to vary with respect to the optical bulk scattering properties. How do the authors justify ignoring the density sensitivity in one variable in the τ^* equation, while keeping it in the remainder?

- Line 185: Could the authors provide some justification for their choice of 400 kg/m^3 as a nominal surface density value, when in-situ measurements across Greenland suggest a nominal surface density of 315 kg/m^3 (Fausto et al., 2018; https://doi.org/10.3389/feart.2018.00051)?

- Line 186: r_{eff} does not appear in Equation 1. Please clarify as to where this is coming from.

- Equation 3: I recommend the authors provide more clarification on the difference between Equations 3a and 3b. It seems 3a is something akin to an "ideal" model and 3b is how the authors approximate it in this study. If that is the case, is it necessary to have both? What does 3a add that can't be communicated by 3b?

- Equation 3: I recommend providing an equation or a more in-depth discussion for the *G*-term in Equation 3b and how it has been selected. How is the surface roughness parameterized within this term? Is it an analytical or empirical representation of roughness? Has this type of parametrization been used previously? Over what roughness range is the model valid? What does it mean to have surface roughness expressed in time, as is done in Line 214? Without this detailed information, the reader cannot make any inference on whether the specific roughness model included in the model is applicable in this application.

- Equation 4: Please provide some explanation for what the *A* and *N* terms in this equation are and what they mean.

- Lines 257-259: I don't know what this sentence is trying to communicate. I don't understand how the test data have been generated or what "… half the spacing between the searched

values." is supposed to mean. I recommend the authors revise this sentence to make it clearer exactly what they have done.

- Line 270: What is the noise value the authors are referring to here? Please be more explicit in what it is and how it comes about.

- Lines 270-278: Much of this paragraph is dedicated to describing the effects ATM amplitude had (past-tense) on the uncertainty in the estimated grain size. The issue I find however, is that the ATM data results have not been covered yet. What do the authors expect the reader to take-away from this paragraph when they have not seen the grain size estimates from the ATM data yet? Furthermore, what does it mean for a surface to be "dark" (Line 275) with respect to laser altimetry? I suggest the authors elaborate or clarify this point. Finally, looking back on this paragraph after reading through the full manuscript, I find it odd that the discussion of precision or uncertainty was not carried through to the actual data analysis. Can the authors quantify the uncertainty in the grain size estimates produced from the analysis of the ATM data that they mention here?

- Line 290: Please consider elaborating further on why negative numbers affect the waveform-median retracker. Is there no other type of retracker that is not sensitive to sensitive to negative numbers that could be used? Also, I would recommend the authors consider including more detail on the two types of retrackers applied to the simulated ICESat-2 data (windowed mean and windowed median) for those who are not familiar with these specific ICESat-2 details. How do they work? The authors are assuming the reader is familiar with this nuanced part of ICESat-2 operations.

- Line 293: What is the IRF function for ICESat-2 and where do the authors get it from? I recommend the authors provide more elaboration on this point. Substantial space was given to establishing the ATM IRF in Section 2, while here the ICESat-2 IRF is almost glossed-over.

- Section 3.4: I think a key point I'm struggling to understand is the link between waveform model of Equation 3 and what is measured by a photon-counting laser altimeter such as ICESat-2. It underlies this entire section and the Section beginning on line 452 but it is not clear to me how this works. Perhaps including an example diagram of how the authors extract the more ICESat-2 relevant parameters from the modeled waveforms would help with this?

- Figure 5: I suggest the authors consider including specific dates in the titles of the subfigures instead of the generic "spring" and "summer" labels. The authors refer to earlier and later spring campaigns in the main body (Line 326) and including actual dates in the figure would make this clearer. Also, some of the detailed patterns the authors discuss in the main body (e.g., Line 331) are very difficult to see in Figure 5. I suggest the authors consider including a zoomed in version of the larger maps that highlight exactly what they are talking about.

- Line 326: When speaking to specific sub-figures, I recommend the authors actually refer to them (i.e., Figure 5b). This helps to make things much clearer for the reader.

- Figure 5 and Figure 6a: Please include coordinates for all maps.

- Figure 6a: The Blues colormap used to represent estimates grain size on top of the Landsat image is very difficult to distinguish. I recommend the authors consider a different colormap

that stands out more from the background. Another option would be to segment the grain size estimates based on the ranges presented in Lines 250-251.

- Line 350: Here the authors refer to the lower portion of the Leidy glacier as having experienced extensive melting. I recommend the authors include some justification for this as the distance between the larger grain size lower portion and the finer grain size upper portions is not much. Has extensive melting been so highly concentrated in only the lower section?

- Figure 7a: Please include a colorbar as is done for Figures 9 and 10.

- Figure 7b: Is there a specific reason as to why the distributions are presented on a log-normal scale? What are the units for the spreads provided in the legend? It seems odd to plot the data on a log-normal scale (especially something like a ratio) and then use the standard deviation. I recommend the authors explain why they expect the ratio between the wide and narrow swath ATM grain sizes to be logarithmically distributed.

- Line 363: To me "around" does not reflect the situation presented in Figure 7a. It appears as if the wide swath grain sizes are consistently larger than the those from the narrow swath. Perhaps it would be more representative to use a term like "near"?

- Figure 8: Please provide coordinates for the plot. Also, the authors mainly discuss the upper half of this plot. So would it prudent to zoom into the upper half in order to see the spatial patterns the authors discuss more clearly?

- Line 398: Here the authors refer to the comparison of the ATM and AVIRIS-NG gain sizes in Figure 9. I recommend the authors clarify which ATM dataset is being compared. Is it the narrow swath, the wide swath, both?

- Line 401: Here the authors refer to ATM grain size being equal to or smaller than the AVIRIS-NG grain sizes when the pattern look more like a loss in sensitivity in the AVIRIS-NG results at small grain sizes (the authors also state this on Line 494). Is this the most appropriate way to describe the results, as being equal to or smaller than? It implies that there is closer agreement between the datasets, whereas it appears more likely to be a numerical artefact.

- Figures 7a, 9, 10, and 11: All these plots show the same type of comparison between grain sizes. As such, I'd recommend the authors consider standardizing them such that the axis limits and colorbars are the same.

- Figures 9-11: I recommend the authors elaborate on why different bounds were chosen for the dashed white lines in these plots (i.e., factor of 2 in Figure 9 and factor of 3 in Figures 10 and 11). Also, please ensure the bounds are plotted properly as they look very similar between the factor of 2 and factor of 3 cases.

- Line 426: Here the authors state that comparison of grain size estimates "… roughly follow the 1:1 line.". Would it be possible to strengthen this argument using a quantitative value such as the correlation coefficient or some other metric?

- Lines 442-444: I find this sentence very confusing, and I don't know what the authors are trying to communicate. Why is it important to use OLCI grain sizes to calculate ATM biases and vice versa? What is the specific satellite bias the authors are presenting in Figure 12a? In

the caption to Figure 12a, what does it mean to predict biases based on OLCI grain sizes that are a function of ATM grain sizes? How are the OCLI grain sizes a function of the ATM grain sizes?

- Figure 12: Why are these bias curves different from those presented in Figure 4? In Figure 12b a 0.12 m ATM bias corresponds to a grain size larger than 1000 μm whereas in Figure 4 it occurs at less than 1000 μm.

- Line 449: The ATM biases in Figure 12b don't really appear to be uniform at around 0.02 m for all grain sizes below 250 μm as stated in the text. For instance, at 1 μm the bias is effectively 0. Is there a lower limit the authors would apply to their best feasible correction or do they mean to imply the existence of a constant 0.02 m bias in the ATM data for small grain sizes?

- Line 459: Could the authors elaborate on what they mean by "the robust spread of the distribution" as it is not a familiar metric. Is it similar to the interquartile range or mean absolute deviation? Also, the reason for using this metric as opposed to something like a standard deviation isn't provided until line 471. I recommend including the rationale for choosing this type of deviation metric when it is first introduced.

- Line 471: Could the authors please elaborate why after establishing the use of the robust spread metric, they suddenly switch to using standard deviation? What does presenting both the robust spread and standard deviation add to the manuscript?

- Figure 13c: The y-axis of this plot implies that at least 1000 samples are found in every bias bin. Is this true? Also, I recommend not using 'K' to express one thousand and either write the complete number or label the y-axis as being in thousands.

- Discussion: Throughout the Discussion the authors continually refer to data or results that have been previously presented. To make it easier for the reader I suggest the authors include pointers to the specific figures they are referring to.

- Line 487: Here the authors suggest look angle could be a reason for the larger grain sizes in the wide swath ATM grain size results. What evidence to the authors present that supports this inference? There is no assessment of grain size versus look angle, so it is difficult to assess the validity of this explanation.

- Line 500: The authors state the grain size relationship between the various grain size estimates is not as consistent as they would have hoped for. Could the authors quantify what the consistency is or what they hoped the agreement between the datasets would have been? This sentence is a little jarring because in the sentence right before the authors state the relationships are consistent but then they say the consistency isn't what they were hoping for and that for a substantial portion of data points there is no clear relationship. What is the reader supposed to take away from this?

- Line 574: Here the authors conclude that the ATM grain size estimates are not strongly sensitive to acquisition geometry. This stands in contrast to line 487 where the authors state that larger grain sizes in the wide swath ATM results are due to look angle. Which is it? And again, there is no support for such a statement in the manuscript.

Technical Comments

- Line 15: “form” to “from”

- Lines 32-35: This is a very long and meandering sentence that is difficult to follow in its entirety. I would recommend partitioning it to more concise statements. Also please include references to support the specific points the authors are making (e.g., ICESat-2 vs ICESat precision and efficiency, weak absorption of green light by ice).

- Lines 36-40: Again, a long and meandering the sentence. It begins talking about glaciers but then halfway through ice shelves and sea ice is introduced, both time- and space-varying biases are discussed but in different contexts. Please consider partitioning the sentence into more distinct statements.

- Lines 41-49: The authors use terms such as “escaping”, “leaving”, “returning”, and “scattering” all to describe a laser signal that is reflected from the surface back towards the detector. I would recommend choosing one and using it consistently. Also, for a paragraph outlining the fundamental physical processes underlining this study, the lack of references supporting them is surprising. I recommend including references supporting the physical phenomena discussed.

- Line 71: Double use of “waveform”

- Line 79: “return-pulse shape” vs “recorded pulse” vs “waveform” all seem to refer to the same concept, so I would recommend the authors choose one term and use it consistently.

- Line 78-79: Please provide a reference(s) to support this type of direct statement.

- Line 79-83: Please provide a reference(s) supporting the statement on how surface and subsurface effects manifest in different altimetry systems and how easy they can be measured.

- Line 84: Please provide a reference(s) supporting the ATM heritage and evolution.

- Line 91-95: Another winding sentence. I recommend partitioning between the discussion of LVIS and SIMPL.

- Line 146: Please be consistent in the use of “AVIRIS” or “AVIRIS-NG”.

- Line 156-157: The authors refer to multiple studies with forward laser waveform models, specifically those making use of diffusion theory to predict the waveform shape, but only one reference is provided. If there is more than one study using a similar model, please include them as well.

- Line 160: Please provide a reference supporting the statement that the diffusion-based approach can produce unphysical results.

- Line 187: Here it is stated that a grain size of 200 μm is used in the derivation of Figure 1 when the caption states 1000 μm. Please clarify which is the correct value.

- Line 240: Is the colon in this line is meant to be a period?

- Lines 255-268: There is only one sentence in this paragraph that doesn't begin with either "To" or "For" and that's because it starts with "Figure". I recommend the authors avoid such repetitive writing as it is makes the paragraph difficult to read.

- Line 280: ICESAT-2 to ICESat-2

- Line 289: Please be consistent throughout the manuscript on how Equations are referenced. In this line the authors use parentheses, *(6)*, while in other instances *Eq.*, *Eqn.* and *Equation* have also all been used.

- Line 299: Please be consistent in using ICESat-2 or IS2. In this paragraph and Figure 4 legend, the authors flip back and forth.

- Line 316: The authors have two Section 3's. I imagine this Section and all sub-sections should be renumbered to Section 4.

- Line 365: Here and throughout the manuscript the authors use multiple versions of grain size (grain-size, grainsize, grain size). I recommend the authors use the more common "grain size" consistently throughout the manuscript.

- Line 366: The "For" at the end of the line should not be capitalized.

- Line 370: I believe the authors are referring to Figure 7a here instead of 6a?

- Figure 10: I assume the comma at the end of the caption is meant to be a period?

- Line 452: Please be consistent through the manuscript with how subfigures are referred to. Here the authors use "panel B" but they have also used "XB" such as on Line 425.

- Line 467: I recommend the authors be consistent with their units. Here they use both centimeter and meter units when referring to the same thing.

- Lines 593-598: This sentence is too long. I recommend partitioning it into smaller, more direct statements.

- Line 603-607: This is almost an exact repeat of what is stated earlier in the same paragraph (lines 586-588). Please carefully review the document to check for redundancies.