Understanding biases in ICESat-2 data due to subsurface scattering using Airborne Topographic Mapper waveform data

Benjamin E. Smith[1], Michael Studinger[2], Tyler Sutterley[1], Zachary Fair[2], and Thomas Neumann[2]

[1]University of Washington Applied Physics Laboratory Polar Science Center, Seattle, WA, 98122, USA
[2]Cryospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA

Correspondence to: Benjamin Smith (besmith@uw.edu)

Abstract.

The process of laser light reflecting from surfaces made of scattering materials that do not strongly absorb at the wavelength of the laser can involve reflections from hundreds or thousands of individual grains, which can introduce delays in the time between light entering and leaving the surface. These time-of-flight biases depend on the grain size and density of the medium, and so can result in spatially and temporally varying surface height biases estimated from laser altimeters, such as NASA's ICESat-2 (Ice Cloud, and land Elevation Satellite-2) mission. Modelling suggests that ICESat-2 might experience a bias difference as large as 0.1-0.2 m between coarse-grained melting snow and fine-grained wintertime snow (Smith et al., 2018), which exceeds the mission's requirement to measure seasonal height differences to an accuracy better than 0.1 m (Markus et al., 2017). In this study, we investigate these biases using a model of subsurface scattering, laser altimetry measurements form NASA's ATM (Airborne Topographic Mapping) system, and grain-size estimates based on optical imagery of the ice sheet. We demonstrate that distortions in the shapes of waveforms measured using ATM are related to the optical grain size of the surface estimated using optical reflectance measurements, and show that they can be used to estimate an effective grain radius for the surface. Using this effective grain radius as a proxy for the severity of subsurface scattering, we use our model with grain-size estimates from optical imagery to simulate corrections for biases in ICESat-2 data due to subsurface scattering, and demonstrate that on the basis of large-scale averages, the corrections calculated based on the satellite optical imagery match the biases in the data. This work demonstrates that waveform-based altimetry data can measure the optical properties of granular surfaces, and that corrections based on optical grain-size estimates can to correct for subsurface-scattering biases in ICESat-2 data.

Formatted: Right: -0.03"

Deleted:

Deleted:

Deleted: .

Deleted:)

Deleted: grain-size estimates

Deleted: argue

Deleted: grain-size estimates

Deleted: has the potential to

Deleted: grain-size estimates

Commented [BS1]: *R1:- 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?*

Deleted: have the potential

1 Introduction.

Laser altimetry techniques allow efficient measurement of precise snow-surface elevations for ice sheets and glaciers, both from satellites (Abdalati et al., 2010) and aircraft (MacGregor et al., 2021). Repeated measurements over glaciers and ice sheets allow the detection of surface elevation changes that show the effects of surface-mass-balance and ice-dynamic processes (Smith et al., 2020), while measurements over floating ice are used to estimate sea ice thickness (Petty et al., 2022) and to infer melt rates beneath ice shelves (Sutterley et al., 2019). These techniques rely on the altimeter's ability to measure the range to the ice or snow surface with high precision. Since its launch in late 2018, ICESat-2 has been making high-precision measurements of ice-sheet and glacier elevation. Unlike the near-infrared (1064-nm) laser used by its predecessor, ICESat, ICESat-2's laser transmits and receives green light, with a wavelength of 532 nm. The shorter wavelength allows ICESat-2 to use highly sensitive detectors to measure the arrival time of individual return photons, improving its overall precision and resolution relative to that of ICESat (Brunt et al., 2021; Markus et al., 2017). At the same time, the choice of a green laser introduces potential biases in its altimetry measurements because ice absorbs green light weakly (Warren and Brandt, 2008), allowing photons to scatter over relatively long distances within the snow before returning to the surface and, potentially, the satellite. These biases can interfere with ICESat-2's primary mission goals of precisely measuring elevation changes over glaciers, ice sheets, and sea ice (Markus et al., 2017) because time varying biases in ICESat-2 measurements could produce spurious signals that might be interpreted as ice-sheet mass changes (Smith et al., 2018). Likewise, spatially varying biases in ICESat-2 measurements over sea ice might falsely be interpreted as variability in freeboard and thus ice thickness (Harding et al., 2011; Smith et al., 2018).

The problem of biases in altimetry data that result from subsurface multiple scattering in snow and ice has been described in previous studies (Harding et al., 2011; Smith et al., 2018). Light is reflected from snow surfaces primarily by multiple scattering, where each photon scatters off many snow grains before escaping the snowpack (Wiscombe and Warren, 1980; Warren, 1982). When light scatters from granular materials that absorb light strongly, only those photons that have scattered a small number of times escape the surface. By contrast, light scattering from weakly absorbing granular materials may enter the surface and scatter from tens or hundreds of grains before escaping again. The extra distance travelled during these subsurface scattering events delays the return of the photons to the surface, so light escaping the surface includes photons that have travelled a distribution of long and short paths. A lidar system measuring the range to a weakly absorbing surface will measure returning photons that have a longer mean travel time and a broader distribution of return times than it would from a non-scattering or strongly absorbing surface. The mean delay of the photons and the shape of the returning pulse (i.e. the measured waveform in an analog lidar, or the distribution of photon timing in a photon-counting lidar) depend on the scattering properties of the material, with lower densities and

Deleted: (Abdalati et al., 2010)

Deleted: (MacGregor et al., 2021)

Deleted: atmospheric

Deleted: (Smith et al., 2020)

Deleted: (Petty et al., 2022)

Deleted: (Sutterley et al., 2019)

Commented [BS2]: *R1- Line 31: Please consider including the corresponding wavelength for ICESat to support the comparison to ICESat-2's 532 nm green laser.*

Commented [BS3R2]: Done.

Deleted: efficient

Deleted: ncreasing

Deleted: efficiency

Commented [BS4]: R2- L33: There are a lot of "efficients" in the first paragraph. It would be useful to clarify why these detectors are so "efficient". Are they sensitive? Low SNR? Energy efficient?

Commented [BS5R4]: We now specify "highly sensitive"

Deleted: , but it leads to

Deleted: s

Deleted: ice shelves

Commented [BS6]: R2 — L32-25: Long, wordy sentence, consider splitting.

Commented [BS7R6]: Also:
R2: L36: Just glaciers? Or ice sheets as well? It seems like these two terms are being used interchangeably which is at odds with the first couple of sentences.

Commented [BS8R6]: Wordy sentence: Now split into t ... [2]

Deleted: These biases are relevant to interpretations of ICE ... [1]

Deleted: (Harding et al., 2011; Smith et al., 2018). ¶

Commented [BS9]: R2 — L32-25: Long, wordy sentenc ... [3]

Commented [BS10R9]: Also: ... [4]

Commented [BS11R9]: Added "ice sheets" ... [5]

Deleted: When

Commented [BS12]: R2: L41-49: Even though this is a v ... [6]

Commented [BS13R12]: Added references to two previ ... [7]

Deleted: and

Deleted: leaving

Deleted: a distribution of photon delays

Deleted: , which both delays the mean time of the returning ... [8]

coarser grain sizes corresponding to deeper penetration of photons into the snow, broader returns, and longer delay times (Fair 2024). Light absorption within the scattering medium can also influence time distribution of returning photons, with stronger absorption producing narrower distributions and smaller net delays because photons are more often absorbed by the medium before they can accumulate long delays. The distribution in time of reflected energy thus can provide information about the optical properties of snow and ice surfaces. The dependence of return photon timing distribution on ice optical properties has also been explored in recent studies (Smith et al., 2018; Allgaier and Smith, 2021; Hu et al., 2022), including in one study where researchers have used predictions from a scattering model to interpret measurements from a hand-carried system to estimate snow and ice optical properties, using a pulsed laser and a detector pressed against the ice surface, separated by a few centimeters (Allgaier et al., 2022). Although other researchers have noted the potential for these theories to be applied to laser remote-sensing measurements, only a few studies have attempted to infer snow and firn properties based on remotely sensed lidar scattering measurements (Hu et al., 2022; Lu et al., 2022; Harding et al., 2011). More recently, a study using ATM measurements from Northeast Greenland (Fair 2024) demonstrated an association between apparent elevation differences between green and near-infrared laser-altimetry measurements and grain-size variations. A second study (Studinger et al., 2023) demonstrated that subsurface scattering of green laser light is associated with negative biases in estimated sea-ice surface elevations, in some cases leading to floating-ice elevations that are apparently below the water surface.

In this study, we investigate the scattering properties of Greenland snow and ice surfaces using altimeter waveform shapes, with the goal of developing a correction for the biases that subsurface scattering can introduce into ICESat-2 data. Although this study is motivated by the need to understand biases in ICESat-2 measurements related to subsurface scattering of green light, data from ICESat-2 are rarely suitable for investigation of subsurface scattering biases, because over rough and sloping surfaces, ICESat-2's 11-m footprint leads to a significant random component in the timing of returned photons, which tends to obscure small changes in the timing distribution associated with subsurface scattering. Slope and roughness tend to be largest in low-elevation regions of Greenland (Nolin and Payne, 2007), which are the same regions where we expect to see the largest subsurface scattering biases. Instead, we use waveform measurements from the ATM airborne laser-altimetry system to test a previously developed model of subsurface scattering (Smith et al., 2018) based on a comparison between the shapes of the returned pulses and pulse shapes expected based on the model. We demonstrate that when we adjust the grain size and surface roughness in the model to match modelled waveforms to measured waveforms we can recover an estimate of the near-surface optical grain size. We test the grain-size estimates from waveform matching by comparing them against grain-size estimates derived from airborne and satellite reflectance measurements. Although this comparison does not suggest a 1:1 linear relationship between waveform-derived grain sizes and reflectance-derived grain sizes,

Deleted: weaker scattering

Commented [BS14]: *R1 - 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.*

Commented [BS15R14]: We added a clause to the last sentence to explain why more absorption is associated with narrower returns:
Light absorption within the scattering medium can also influence time distribution of returning photons, with stronger absorption producing narrower distributions and smaller net delays because photons are often absorbed by the medium before they can accumulate long delays.

Commented [NT(616R14]: Does including 'weaker scattering' here really add anything. I read it as confusing - it's a whole lotta scattering that leads to the broader returns and longer delay times. So to me that's stronger scattering. But that the hell do I know (I come from Waunakee)

Deleted: …. ¶ ... [9]

Deleted: previous

Commented [BS17]: *R1- Line 50: Someone approaching the manuscript from an ICESat-2 (i.e., individual photon) perspective may be unfamiliar with the concept of a laser* ... [10]

Commented [BS18R17]: We now say: (i.e. the measured waveform in an analog lidar, or the distribution of photon ti ... [11]

Deleted: (Smith et al., 2018; Allgaier and Smith, 2021; Hu et al., 2022)…, and ... [12]

Deleted: (Allgaier et al., 2022)…. Although other studies researchers studies …ave noted the potential for remote-sen ... [13]

Deleted: (Hu et al., 2022; Lu et al., 2022; Harding et al., 2011)…. Two studies submitted contemporaneously with ours…ore r ... [14]

Commented [BS19]: L56-59: I would need access to these manuscripts to judge overlap and novelty of this paper. ... [15]

Deleted: Unpublished manuscripts provided as attachments) explore, respectively,

Deleted: the…estimation of snow grain size based on…ssociation between apparent elevation differences between green and n ... [16]

Formatted: Font: 8 pt

Deleted: , and the biases associated with subsurface scattering of green laser light in altimetry measurements over sea ice. Th ... [17]

Formatted: Indent: First line: 0.5"

Moved up [1]: In this study, we investigate the scattering properties of Greenland snow and ice surfaces with the goal of

Deleted: ¶

Deleted: ta. …useUsing…waveform measurements from an …he ATM airborne laser-altimetry system over the Greenland ic ... [18]

Deleted: (Smith et al., 2018)… based on a comparison between the shapes of the returned pulses and timing distributions… ... [19]

we use ICESat-2 biases calculated for the ATM grain-size estimates as a proxy for direct measurements of ICESat-2 biases to calibrate a correction based on reflectance-derived grain sizes, and demonstrate that the calibrated correction can produce elevation estimates that, averaged over a range of Greenland terrain and surface conditions, are unbiased. Although the results of this study fall short of a correction that could eliminate grain-size-driven biases in ICESat-2 data, we provide a description of some of the advances in satellite remote sensing that would be needed to more adequately address this problem.

2. Data

This study is based on waveform data from the ATM lidar systems, grain-size estimates based on the airborne AVIRIS-NG (Airborne Visible/Infrared Imaging Spectrometer, Next-Generation) spectroradiometer, and grain-size estimates based on the spaceborne OLCI (Ocean and Land Colour Instrument) instrument. A summary of measurement locations for the airborne data is presented in section 3.

2.1 Altimetric waveforms from the Airborne Topographic Mapping lidar systems.

ATM (the Airborne Topographic Mapping system) make laser-altimetry measurements using a conically scanning laser that maps elevations beneath an airplane over a swath 40-500 m wide. ATM has made measurements over the Greenland and Antarctic ice sheets since 1993, with an evolving configuration of lasers and measurement strategies that have gradually improved measurement precision and reliability (MacGregor et al., 2021; Krabill et al., 2002). Since 2017, the system has used green (532-nm) lasers with a 1.3-ns pulse duration (full width at half maximum) and a receiver with a bandwidth of around 1 GHz. At a nominal flight elevation of 500 m above ground level the size of the lidar footprint on the surface is ~0.70 m diameter. ATM's configurations include a narrow-swath scanner whose 5º full scan angle makes measurements over a ~40-m swath on the ground at a flight elevation of 500 m, and a wide-swath scanner whose 30º scan angle produces a ~460-m swath.

Many lidars, including both photon-counting instruments such as that used by ICESat-2 and analog instruments such as ATM, can measure the time distribution of light that has reflected off their targets. Photon-counting altimeters measure the distribution of photon-return times directly, while analog lidars measure a time series of voltages that are approximately proportional to the rate at which photons are incident on the detector. Ideally, each of these types of measurement would give a good approximation of the time distribution of photons reflected from the ice, and a waveform measured by an analog lidar would be equivalent to a histogram in time of photons detected by a photon-counting lidar. In practice, the characteristics of the altimeter and the characteristics of the surface measured both play a role in the degree to which subsurface-scattering effects can be distinguished in the recorded waveform. In our model (see section 3), the waveform

Deleted:

Deleted: We then demonstrate the use of grain-size estimates derived from satellite imager

Commented [BS20]: *R1- 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?*

Deleted: y

Deleted: to estimate range delays for ICESat-2, confirm that these delays are consistent with delays derived based on airborne waveform waveforms and discuss how sampling in high-res ... [20]

Commented [BS21]: *R1 - Section 2: I know maps are presented in Figure 5, but I would recommend the authors consider* ... [21]

Commented [BS22R21]: We now point the reader to section 3 for measurement locations. ... [22]

Deleted: waveform data

Deleted: grain-size estimates

Deleted: from

Deleted: an airborne spectrometer

Deleted:

Deleted: satellite-based grain-size estimates

Formatted: Normal

Commented [BS23]: R2- L84: Define acronym on first use of term "ATM" (L75) rather than here. ... [23]

Commented [BS24R23]: The acronym is defined in the abstract (~line 15), and is not needed here.

Deleted: T

Deleted: return-pulse shape

Deleted: Because

Deleted: recorded pulse

for a laser altimeter corresponds to the temporal convolution of the distribution of photon delays, the impulse response function (IRF) of the recording system, the range to the surface, and the shape of the transmitted pulse, so effects of subsurface scattering become easier to measure for narrower transmitted pulses, higher bandwidth recording systems, flatter surfaces, and smaller beam divergence values. The recent (post-2017) versions of the ATM transceiver offer good potential to measure scattering effects, because the temporal resolution of the system (corresponding to the receiver sampling interval and the pulse duration) is not large compared with the path delays predicted for green light reflecting from snow surfaces (Smith et al., 2018). Similar measurements have been made using the Land, Vegetation, and Ice Sensor (LVIS) (Hofton et al., 2008), but because of that sensor's longer pulse duration and infrared wavelength, we expect its waveform shapes to have only limited sensitivity to snow conditions. Photon-counting lidar measurements by the Slope Imaging Multi-polarization Photon-counting Lidar (SIMPL) (Yu et al., 2016; Harding et al., 2011) offer some of the advantages of ATM data, but used a photon-counting detection strategy that is not compatible with the processing methodology used in this study.

Table 1. Dates and instruments for ATM measurements.

Campaign	**Instrument**	**Dates processed**
Summer, 2017	narrow-swath	July 7 – July 24
Spring, 2018	narrow-swath	3 March – 1 May
Spring, 2019	narrow-swath	3 April – 14 May
Summer, 2019	narrow-swath , wide-swath	4 September – 11 September

ATM waveform measurements in this study come from data collected in Greenland in the 2017 summer campaign, the 2018 spring campaign, and the 2019 spring and summer campaigns. Most of the data that we processed (summarized in table 1) were collected using the ATM narrow-swath scanner, but we also processed wide-swath data from the 2019 summer campaign. For both scanners, the laser's incidence angle on a flat surface is approximately half the full scan angle, thus 15º for the wide swath and 2.5º for the narrow swath. Waveform data from these campaigns are distributed in the ILNSAW1B and ILATMW1B products (Studinger, 2018a, b), which provide digitized transmitted and received waveforms associated with each transmitted pulse. The waveforms have a temporal sampling of 0.25 ns, and are quantized at 8 bits, to produce digital values between 0 and 255. A variable neutral density filter in front of the receiver determined the optical throughput of the system, and was set to avoid digitizer saturation over snow surfaces. We considered using the near-infrared waveform data collected during the 2019 Greenland campaign, but found that the

Deleted: expected

Deleted: the

Deleted: ¶
The Airborne Topographic Mapping System instrument suite (ATM) has made altimetry measurements over the Greenland and Antarctic ice sheets since 1993, with an evolving configuration of lasers and measurement strategies that have gradually improved measurement precision and reliability. Since 2017, the system has used green (532-nm) lasers with a 1.3-ns pulse duration (full width at half maximum) and a receiver with a bandwidth of around 1 GHz. At a nominal flight elevation of 460 m above ground level the size of the lidar footprint on the surface is ~0.64 m. It is this more recent set of data that offers the best potential to measure the optical properties of snow surfaces, because

Deleted: (Hofton et al., 2008)

Deleted: provide

Deleted: of waveform shapes

Deleted: ;

Deleted: p

Deleted: (Yu et al., 2016; Harding et al., 2011)

Deleted: software

Formatted: Caption, Keep with next

Formatted: Font: Bold

Formatted: Font: Bold

Deleted: ¶
Waveform measurements in this study come from the ATM wide-swath and narrow-swath waveform products

Commented [BS29]: *R1 - 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.).*

Commented [BS30R29]: We now provide a summary of swath widths and scan angles for the two systems.

Deleted: (Studinger, 2018a, b)

Deleted: in each airborne campaign

Deleted: During flights, a

Deleted: is adjusted to compensate

Deleted: for varying conditions and

signal-to-noise ratio of these data was much lower than that of the green data, and that over coarse-grained surfaces, the infrared return was often absent even when the green waveform showed a clear return. Therefore, to obtain a consistent set of measurements, we focuse our study on the green waveforms.

At the start of each ATM measurement campaign, waveforms were recorded with the laser aimed at a fixed, flat panel of fine-grained white material (Spectralon®) (Studinger et al., 2022a). We take these measurements to represent the system IRF $I(t)$ for the whole campaign. Although ATM instruments record both the received and transmitted waveforms, we found that the recorded transmitted waveforms were not a good representation of the system impulse response (see supplemental material section S1). Because of this, we disregard the measured transmitted pulse shapes, and instead assume that the system IRF is consistent with the most recent calibration measurement available. The wide-swath and narrow-swath ATM instruments produce very similar measurements, but use separate transmitters, optics, and receivers; for this reason, we use separate calibrations for the two systems for each campaign (Studinger et al., 2022b).

2.2 Grain-size estimates from the AVIRIS-NG airborne spectrometer

To help evaluate whether the ATM-derived waveforms were consistent with the returns we would expect from known surface conditions, we used data collected using AVIRIS-NG, on a separate aircraft that followed the aircraft carrying ATM on five subsequent days in the autumn of 2019. AVIRIS-NG measures radiances at 425 different wavelengths between 380 and 2510 nm on a detector array that produces images with 598 across-track samples (Thompson et al., 2018); its ~7.5 km altitude during the 2019 survey produced images on a ~4-5 km-wide swath, with ~6-7 m pixel sizes (Nolin and Dozier, 2000). These measurements were processed to estimate grain sizes using a technique that uses the strength of an absorption feature in the reflectance spectrum of snow at 1.03 μm as an indicator of snow grain size (Nolin and Dozier, 2000). We rejected one of the data files (the single file collected on 9 September, 2019, and the only file with extensive coverage of sea ice) because while the image appears to resolve a melting surface including a variety of sea-ice features including melt ponds and leads, the range of retrieved grain sizes span a small range (90% of values between 164 and 287 μm). The reason why this file should contain anomalous values is not clear, although we note that the sun was lower in the sky than it was for any other file (79º solar zenith angle, as compared to ~70-72º for other files in the campaign), which we hypothesize might result in lower-quality grain-size retrievals. The remaining 26 data files cover two coast-parallel lines and a few coast-perpendicular lines in northwest Greenland, spanning a range of grain size conditions from large-grained melting surfaces near the coast to fine-grained surfaces inland, and 17 of these overlapped with available ATM waveform files. Most (~80%) overlapping measurements within a 5-day window were collected within three hours of one another, and to limit how much the surface might have changed between one set of measurements and the other, we compare measurements

Deleted: . Therefore the amplitude of the recorded pulse does not have a consistent relation with the intensity of the received signal.

Formatted: Indent: First line: 0.5"

Deleted: (Studinger et al., 2022a)

Deleted: The

Commented [BS31]: L103: Instruments? Surely there is only one LiDAR or were LVIS and SIMPL also onboard? It may be useful to name the lidar sensor given that ATM is defined as a suite of instruments (or revise L84).

Commented [BS32R31]: We now specify "ATM instruments"

Deleted: .

Deleted: The recorded transmited waveform is delayed through a multimode fiber-optic cable to eliminate backscattered photons from the scan mirror and the aircraft's optical window. This delayed waveform was intended to record the shape of the transmitted pulse, but our analysis suggests that dispersion in the delay fiber introduces significant temporal blurring of this waveform, so that for smooth, flat surfaces with minimal subsurface scattering, the measured waveforms are more consistent with the calibration waveform than they are with the waveforms recorded during flight.

Commented [BS35]: L110-112: Might be useful to clarify the difference in swath widths here or L96 when it is first mentioned. Or are these two different sensors? Either way I think some general tightening of terminology is needed in this section

Commented [BS36R35]: This is now covered by our improved description of the two systems.

Deleted: (Studinger et al., 2022b)

Commented [BS37]: *R1 - Section 2.2: The authors include a lot of detail regarding the ATM system but almost none* ... [24]

Deleted: ¶

Deleted: **Grain-size estimates**

Deleted: verify that

Formatted: Space Before: 12 pt

Commented [BS38]: R2- L114: "Verify" seems a bit st ... [25]

Commented [BS39R38]: We now say "evaluate whether"

Deleted: NG (the Airborne Visible/Infrared Imaging ... [26]

Deleted: Basler

Deleted: These measurements were processed to estimate ... [27]

Deleted: (Nolin and Dozier, 2000). These data provide gr ... [28]

Deleted: The quality of the grain-size retrievals depends s ... [29]

Formatted: No underline

Formatted: No underline

Deleted: grain-size

Deleted: (

Deleted:)

between the two systems only if the differences between timestamps for the data files are less than 200 minutes.

2.3 Grain-size estimates from OLCI reflectance measurements

To demonstrate potential corrections for ICESat-2 height biases, we use a set of satellite measurements (Vandecrux et al., 2022b) derived from the OLCI instrument onboard the European Space Agency's Sentinel-3A satellite. OLCI provides surface-reflectance information for 21 spectral bands over a 1270-km wide swath with sub-kilometer resolution, giving sub-daily revisit times for Greenland during summer months. Images that were determined to be cloud free were converted to grain-size estimates by comparing estimated surface reflectances at 685 nm (far red, band 17) and 1020 nm (near infrared, band 21) with the output of a radiative-transfer model (Kokhanovsky et al., 2019). The result is a set of daily maps of Greenland, posted at 1 km, giving an estimate of the surface optical grain size for cloud-free areas of the ice sheet (Vandecrux et al., 2022a). Validation of these maps (Vandecrux et al., 2022b) against ground-based grain-size estimates derived from the infrared (1310 nm) reflectance of surface-snow samples collected at EastGRIP in northeast Greenland found that the OLCI-based estimates were systematically larger than ground-based estimates, but showed the expected decreases during snowfall events and increases during melt events. We compare ATM and AVIRIS-NG grain-size estimates with the OLCI-based estimates by bilinear interpolation into each daily grid; if measurements are marked as invalid in an OCLI map because of the presence of clouds, we derive an estimate based on the previous day's map under the assumption that the grain size had not changed substantially between the two days, and if the previous day's estimate is invalid, we reject the data point.

3. Methods

Work in this study is based on a model of how the measured time distribution of light reflected from a scattering surface depends on the properties of the surface and on the properties of the transmitted waveform (Smith et al., 2018). We partially validate this model by comparing its results with measured waveforms, and by tuning the parameters in the model, we estimate surface grain sizes in Greenland, and use these grain-size values as a proxy for the degree of subsurface scattering to help predict subsurface scattering delays in ICESat-2 data.

3.1 Modeling return time distributions

We model light scattering in snow and firn based on a Monte-Carlo radiative transfer model for near-surface scattering combined with an analytical extrapolation of the shape of the return for photons with long scattering

Deleted: .

Deleted: Grain-size estimates

Deleted: satellite

Deleted: (Vandecrux et al., 2022b)

Deleted: Ocean and Land Colour Instrument (

Deleted:)

Deleted: grain-size estimates

Deleted: (Kokhanovsky et al., 2019)

Deleted: (Vandecrux et al., 2022a)

Commented [BS42]: L134: Consider revising because the way it's written makes it sound like Gallet et al. (2009) validated snow-grain sizes from OLCI which was launched in 2016

Commented [BS43R42]: We deleted the reference to Gallet et al. That reference is contained in the Vandecrux reference, which should be adequate documentation of the comparison.

Deleted: in-situ datasets

Deleted: C

Deleted: estimates derived from the infrared (1310 nm) reflectance …(Gallet et al., 2009) of surface-snow samples collected (Gallet et al., 2009) of surface-snow samples collected at EastGRIP in northeast Greenland,

Deleted: (Gallet et al., 2009) of surface-snow samples collected at in northeast Greenland,

Formatted: Not Highlight

Deleted: ,

Deleted:

Deleted: (Vandecrux et al., 2022b)

Deleted: (Vandecrux et al., 2022b)

Commented [BS47]: *R1 - Section 3: I think there is something missing from this section, specifically how τ* related to r"##. r"## 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.*

Commented [BS48R47]: We have clarified this section by modifying equation 3 to include the relationship between T* and Reff

Deleted: The daily grids include labels that mark pixels for which the grain size could not be retrieved, generally because of the presence of clouds. We assembled these maps into a look-up table for time-varying grain size by creating a daily grid for Greenland, and for each time step, updating the grain-size estimate for the previous time step with valid data from the current time step, while keeping track of the difference between the time for each grid cell and the most recent valid observation for that grid cell. Val … [30]

Deleted: (Smith et al., 2018)

Deleted: grain-size values

delays (Smith et al., 2018). This model is similar to that used in other studies (Allgaier and Smith, 2021;Hu 2022), except that we use a Monte-Carlo model to predict the return photon distribution at short delay times and diffusion theory at longer delay times, where the other studies use diffusion theory at all times. The choice to use diffusion theory is appropriate when the detector and the laser source are not coincident (i.e. when all photons measured have travelled an appreciable horizontal distance through the scattering medium) but less so for the backscatter geometry used here, because diffusion theory can produce unphysical results for very short time delays (Flock et al., 1989). For measurements in which there is a horizontal offset of more than a few times the scattering length between the source and the detector, these short delays are not observed, whereas in the backscatter geometry of an altimetry measurement, many photons are likely to return after only a few scattering events. By directly modelling the time of flight for the incident beam and the first few scattering events, our Monte Carlo model avoids this problem.

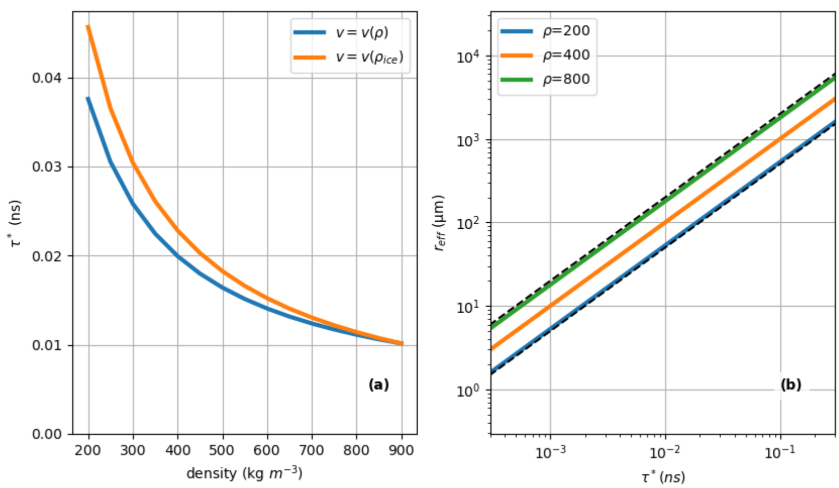

Figure 1. Relation between scattering time, density, and effective grain size. Panel A) shows the relation between scattering time and density for a constant grain size of 200 µm, using a mixing law to calculate the velocity, and using a constant velocity appropriate to solid ice. Panel B) shows the relationship between scattering time and grain size, for three different densities. The dashed black lines show double and half the effective radius for $\rho = 400\ kg\ m^{-3}$.

Returns from our model can be described as:

Deleted: (Smith et al., 2018)

Deleted:)

Deleted: (

Deleted: (Allgaier and Smith, 2021)

Deleted: ,

Deleted: , and the horizonal displacement

Deleted: of

Deleted: means that

Deleted: 10

Commented [BS49]: *R1 - 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?*

Commented [BS50R49]: We discuss this in the text. As you note, the orange line is there to demonstrate that the main sensitivity of T* to density comes about because of the distance between scattering events, not because of the effective velocity. As you observe, the orange line represents an unphysical situation, but since it is there to illustrate a sensitivity of the model, we do not see this as a problem.

$$SRF_m(t) = S\left(\frac{t}{\tau^*(r_{eff}, \rho)}\right)\exp\left(-k_{abs}(r_{eff}, \rho)v_{eff}(\rho)t\right) \qquad 1$$

where

$$\tau^* = \left(v_{eff}(\rho)k_{scat}(r_{eff}, \rho)\left(1 - g(r_{eff})\right)\right)^{-1}$$

Here v_{eff} is the effective velocity of light traveling through the scattering medium, which depends on the density; k_{scat} and k_{abs} are the bulk scattering and absorption coefficients of the medium; g is the asymmetry parameter of scattering in the medium; and S is a scattering function that gives the distribution of return times from a non-absorbing scattering half space, in units of the average time between scattering events in the half space. The quantity τ^* describes the time required for light to travel between two scattering events, where we have approximated the anisotropic scattering characteristics of light interacting with large particles by multiplying the scattering coefficient by a factor $(1 - g)$ (Smith et al., 2018). We estimate the optical bulk scattering properties based on a Mie-theory calculation treating ice grains as independent spheres of ice surrounded by air (Gardner and Sharp, 2010), which gives estimates of k_{scat} and k_{abs}, and g as a function of wavelength, grain size, and density. We approximate the velocity of light in firn for density ρ:

$$v_{eff} = c\left(\frac{\rho}{\rho_{ice}}n_{ice} + \frac{\rho_{ice} - \rho}{\rho_{ice}}n_{air}\right)^{-1} \qquad 2$$

where c is the speed of light in a vacuum, ρ_{ice} is the density of ice, n_{ice} is the real part of the refractive index of ice calculated from a published compilation (Warren and Brandt, 2008), and $n_{air} = 1$.

To reduce our description of scattering to a single parameter, we use a nominal density value of 400 kg m^{-3}, and a corresponding velocity value of 0.27 m ns^{-1}, which lets us express Eq. 1 solely in terms of k_{abs} and r_{eff}. Although the choice of 400 kg m^{-3} is somewhat arbitrary, it strikes a balance between the smaller, 270-350 kg m^{-3}, densities typical of Greenland snow (Fausto et al., 2018) and the larger, 410-910 kg m^{-3} densities observed in ablation-zone surfaces (Cooper et al., 2018). Figure 1A shows t^* as a function of density for a grain size of 200 µm, plotted once using the relationship between velocity and density from Eq. 2, and once using a constant velocity value appropriate for solid ice. Over this range of densities, t^* varies by about a factor of 4, while the difference in t^* associated with the velocity model is at most about 20%. This shows that most of the variability in scattering time is associated with the distance between scattering events (determined by the density and the grain size), not with the velocity of light in the medium (determined by the density alone). Figure 1B shows grain size that would be inferred for a given t^* value, for our nominal density value (400 kg m^{-3}), and for densities corresponding to light, fresh snow (200 kg m^{-3}) and to nearly solid ice (800 kg m^{-3}). Over the three-order-of magnitude range of r_{eff} considered here, the range of r_{eff} at any given value of t^* between the nominal and the extreme values of density is just less than a factor of two, which demonstrates

Deleted: (Smith et al., 2018)

Deleted: (Gardner and Sharp, 2010)

Deleted: (Warren and Brandt, 2008)

Commented [BS51]: *R1 - Line 185: Could the authors provide some justification for their choice of 400 kg/m3 as a nominal surface density value, when in-situ measurements across Greenland suggest a nominal surface density of 315 kg/m3 (Fausto et al., 2018; https://doi.org/10.3389/feart.2018.00051)?*

Commented [BS52R51]: We agree that this choice is arbitrary, and added a statement to that effect to the text, quoting the density values from Fausto et. Al.:
Although the choice of 400 kg m-3 is somewhat arbitrary, it strikes a balance between the smaller, 270-350 kg m-3, densities typical of Greenland snow (Fausto et al., 2018) and the larger, 410-910 kg m-3 densities observed in ablation-zone surfaces (Cooper et al., 2018).

Deleted: eqn

Formatted: Superscript

Commented [BS53]: *R1 - Line 186: r"## does not appear in Equation 1. Please clarify as to where this is coming from.*

Commented [BS54R53]: We modified Equation 1 to make clear that the scattering coefficients depend on grain size and density, and the asymmetry factor depends on grain size.

Deleted:

that while there is some uncertainty in the relationship between t^* and r_{eff} when the density is unknown, a measured value of t^* can constrain the surface grain size to around a factor of two.

3.2 Modelling expected waveform shapes

The return waveform measured by an altimeter depends on the scattering properties of the surface, on the shape of the surface, and on the IRF of the system making the measurements. We calculate model surface-return shapes as:

$$W_{model}(t - t_{surf}, r_o, \sigma) = I_{est}(t) \otimes SRF_m(t; r_o) \otimes G(t, \sigma) \quad (3)$$

Here $W(t - t_{surf})$ is the received waveform, where t is time and t_{surf} is the round-trip travel time to the surface, and $\otimes$ represents a temporal convolution. $SRF_m(t; r_o)$ is calculated from Eq. 1, $I_{est}(t)$ is an estimate of the system IRF. We approximate the distribution of photon delays due to slope and surface roughness weighted by the illumination pattern of the laser as a Gaussian function, $G(t, \sigma)$. Our approximation of the effects of slope and roughness follows studies that modelled satellite laser altimetry waveform shapes (Yi et al., 2005; Smith et al., 2019a). If we assume that the illumination pattern is represented by a two-dimensional Gaussian function with standard deviation σ_b and the illuminated surface is represented well by a rough plane whose normal makes and angle φ with the beam direction, and that the roughness produces a Gaussian distribution of elevations relative to the plane with standard deviation σ_r, then the standard deviation of the Gaussian function, σ, should be equal to $\frac{2}{c}(\sigma_r^2 + \sigma_b^2 \tan^2(\phi))^{1/2}$ (Yi et al., 2005; Smith et al., 2019b). This means that more strongly sloping surfaces should produce broader returns, and that returns from the wide-swath ATM instrument should be broader than those from the narrow-swath instrument. ATM's 0.7-m footprint implies that over a flat surface smooth surface, $\sigma \approx 1$ ns for the wide (15-degree incidence angle) swath , or 0.2 ns for the narrow (2.5-degree incidence angle) swath.

Deleted: Combining these gives

Commented [BS55]: *R1 - 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?*

Commented [BS56R55]: We have removed equation 3a, leaving only our approximation of the surface return (formerly 3b).

Deleted: $W(t - t_{surf}) = I(t) \otimes SRF(t) \otimes \delta z\left(\frac{ct}{2}\right)$ ... [31]

Deleted: b

Deleted: *ct2*

Deleted: *0,*

Formatted Table

Deleted: Equation 3a describes our assumed model of how the received waveform relates to *I(t)*, the system IRF of the altimeter, including the transmitted pulse and the receiving optics and electronics, to *SRF*(t), the rate of photons returning from the surface, and to δz, the distribution of surface heights within the altimeter's footprint. Equation 3b describes our model of the return, where

Deleted: ,

Deleted: and $G(t, 0, \sigma)$ is a Gaussian function representing

Commented [BS57]: *R1 - 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.*

Commented [BS58R57]: We expanded on our discussion of slope and roughness, to include citations to two studies that treat roughness this way, and include a calculation of how the ATM scan angle affects surface return broadening angle.

Deleted: combination of surface roughness and return broadening due to surface slope.

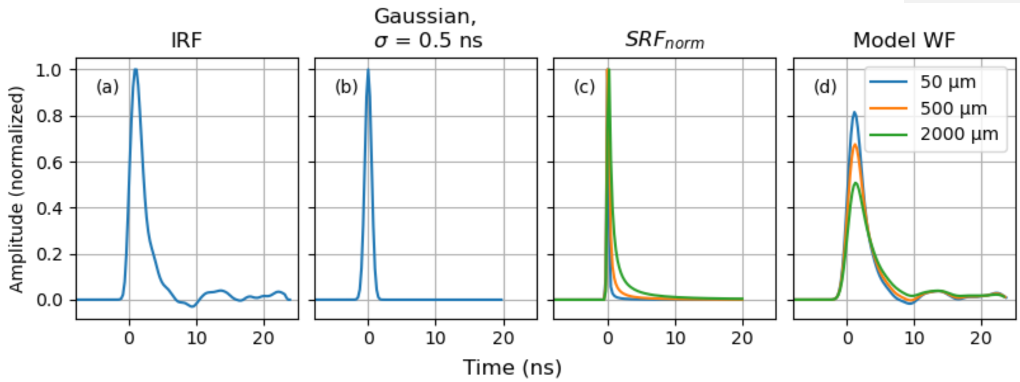

Figure 2. Components of the waveform model. The ATM IRF (a) is convolved with a Gaussian function representing surface roughness (b) and the surface response function (c) to produce the model waveform (d). Three SRFs and corresponding waveforms are shown in (c) and (d), for r_{eff}=50, 500, and 2000 µm. Curves in (a-c) are normalized to unit amplitude, curves in (d) are based on an IRF with with unit amplitude.

Figure 2 shows the components of Eq. 3, and resulting waveforms, based on the system IRF measured using a calibration target with no significant subsurface scattering on 9 March 2018, for a surface roughness equivalent to 0.5 ns (*i.e.* 7.5 cm), and for three snow grain sizes: 50, 500, and 2000 µm. The modeled waveforms show that for increasingly large grain sizes, the peak amplitude of the waveform becomes smaller and the waveform becomes broader, with the trailing edge of the waveform being blurred much more than the leading edge. The measured *I(t)* has a distinctive droop (negative excursion) just after the end of the main pulse, which is reflected in the predicted waveforms, although for larger grain sizes it no longer extends below zero. We were initially uncertain that the droop in the *I(t)* was due to a process that would be modeled correctly by Eq. 3, but the consistency between modeled and recovered waveforms (see section 4.1) suggests that the process that leads to the droop is a linear effect, likely in the receiver electronics. We speculate that it is due to bandwidth limitations in the receiver, perhaps due to an impedance mismatch at the input of the digitizer, but do not have strong evidence about its origin.

3.3 Matching modelled waveform shapes to measured waveforms

For each measured waveform, we identify the first sample at which the waveform exceeded 50% of its maximum amplitude and assume that all samples more than 3 ns before this sample contains only a uniform background offset and noise, whose values we calculate as the mean and standard deviation (N_{est}) of the sample values in this region. We then correct each waveform by subtracting this background offset.

Deleted:
Deleted: A
Deleted: B
Deleted: C
Deleted: D
Deleted: C
Deleted: D
Deleted: .
Deleted: equation
Deleted: b
Commented [BS59]: L213: It would be useful to briefly remind readers how the IRF was measured here.
Commented [BS60R59]: Added: "measured using a calibration target with no significant subsurface scattering on 9 March 2018"
Deleted: eqn
Deleted: the next
Deleted: ied
Deleted: ed
Deleted: ed
Deleted: DC o
Deleted: ed
Deleted: ed
Deleted: DC

To match waveforms with model results, we minimize the misfit between the DC-corrected and modelled waveforms:

$$R^2(r_{eff}, \sigma, t_0) = \sum \left[\frac{P_m(t_i) - A\, W(t_i - t_0, r_o, \sigma)}{N} \right]^2 \qquad 4$$

Here $P_m(t_i)$ is the waveform sampled at times t_i, corrected for the background offset, and W is the modelled waveform, A is a scaling term relating the amplitude of the modelled waveform to that of the measured waveform, and N is the number of samples in the waveform.

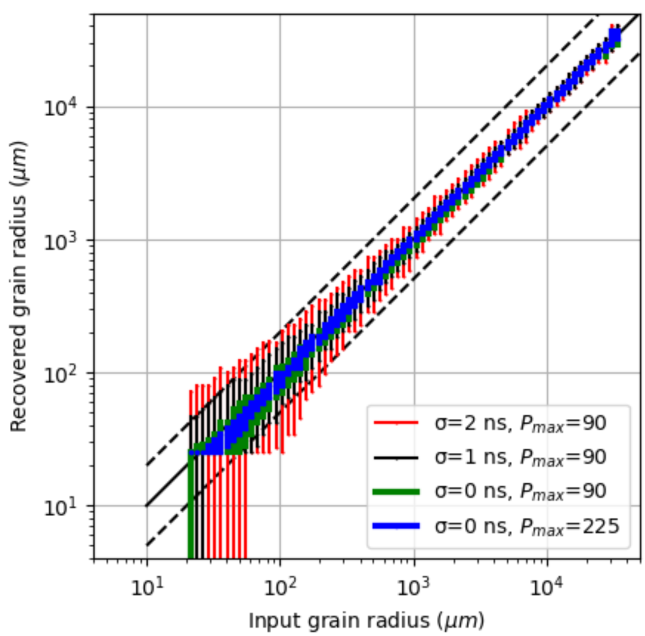

Figure 3. Fitting test data. Vertical bars show the range of recovered grain sizes for each input grain size value, three low -amplitude, rough-surface case with P_{max}=90 and variable pulse broadening (σ=2, 1, and 0 ns), and a high-amplitude case with P_{max}=255 and no pulse broadening (σ=0 ns). Bars indicate the 5th and 95th percentiles of the recovered grain sizes; bars extending off the bottom of the plot for the smallest grain sizes and the low-amplitude case indicate that for more than 5% of the waveforms, the best fit was with the non-scattering model

Deleted:

Deleted: d

Commented [BS61]: *R1 - Equation 4: Please provide some explanation for what the A and N terms in this equation are and what they mean*

Commented [BS62R61]: Added: ", *A* is a scaling term relating the amplitude of the modelled waveform to that of the measured waveform, and *N* is the number of samples in the waveform. "

Deleted: background rate

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: for a

Deleted: A

Deleted: , σ=2 ns

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

Deleted: ,

Deleted: smooth-surface case,

Deleted: A

Deleted: ,

Formatted: Font: Italic

waveform. **The solid line indicates a 1:1 relationship between the input and recovered grain sizes, and the dashed lines indicate the 0.5:1 and 2:1 relationships.**

We find optimal values for our adjustable parameters using a three-stage golden-section search (Press et al., 2007) in σ, r_{eff}, and t_0. The search algorithm consists of an outer search over a range of r_{eff} values, with an inner search over σ values, and within that a second inner search over t_0 values. Within the search over t_0, the amplitude values are found with a least-squares regression between each model waveform and the measured waveform. The searches use a tolerance in σ of 0.25 ns and a logarithmic tolerance in r_{eff} of 10%. After each golden-section search has converged, a final parabolic-search step is used to further refine the estimated σ, and r_{eff} values. The convolutions in Eq. 3 are computationally costly, so we keep track of all waveforms we have calculated, and, whenever possible, use pre-computed waveforms in the misfit calculations. Using the golden-section search rather than a derivative-based searching strategy (*e.g.* a steepest-descent or conjugate-gradient search) lets the fitting algorithm search a consistent set of parameters as it encounters waveforms that are similar to waveforms that it has previously matched, which greatly reduces the time required to fit a collection of waveforms, many of which are similar to one another. We further reduce our computational times by fitting only every fourth waveform for data from the narrow-swath scanner, and every second waveform from the wide-swath scanner. For most purposes in this study, we further reduce the spatial resolution of the recovered grain-size estimates using a 10-meter block-median filter, in which we identify the pulse containing the median grain size value within each 10x10 m block sampled by each survey and report its location and grain size.

To evaluate the resolution and accuracy of this fitting procedure, we generated a set of test waveforms based on $I_{est}(t)$, for a range of grain sizes, pulse amplitudes, and broadening values. We assessed the sampling distribution of the recovered grain-size estimates by generating 256 different waveforms for each combination of parameters, normalizing each to a specified peak amplitude (P_{max}), adding random (normal-distribution) values with a standard deviation of two digitizer counts to each sample, and applying our fitting algorithm to each. Our fitting algorithm selects grain sizes based on a set of pre-computed waveforms generated for grain-size values separated by 10%, so to demonstrate the worst-case performance of our algorithm, we generated the test data based on grain sizes that were half-way between the grain-size values used by the algorithm. Figure 3 shows the relationship between the specified and recovered grain size for small amplitudes and a range of broadening values (P_{max} = 90, σ=0, 1, and 2 ns), and for large amplitudes and small broadening values (P_{max}=225, σ=0 ns). For the high-amplitude waveforms with little broadening (P_{max}=255, σ=0 ns), the fitting procedure consistently recovers grain sizes as small as 20 µm, converging to either the next larger or the next smaller grain size value among the searched values (separated by 10%) with a moderate preference for the next smaller value, giving recovered values whose distribution width (5th to 95th percentile) is on the

- **Deleted:** (Press et al., 2007)
- **Deleted:** :
- **Deleted:** n
- **Deleted:** b
- **Deleted:** d
- **Deleted:** d
- **Deleted:** grain size estimates
- **Deleted:** grain-size
- **Deleted:** ,
- **Deleted:** ¶

order of 10%. At smaller amplitudes (P_{max}=90) and larger pulse broadening values (σ=1, 2 ns), the width of the recovered distribution increases with decreasing grain size, with the 5th and 95th percentiles of the distributions spanning around a factor of 5 for r_{eff}=50 µm and σ=2 ns. For input grain sizes up to about 75 µm (a factor of three times the minimum grain size tested), the waveform that best fit the simulated waveform was often the one with no scattering for the low-amplitude and broadened waveforms (A=90, σ=2 ns.) In these cases, the bottom of the distribution is not constrained on a log scale.

Our numerical experiments show that for synthetic data, the ratio between the amplitude of the pulse and the RMS of the noise added to the synthetic waveform plays a large role in the accuracy of the recovered grain size, with larger signal-to-noise ratios corresponding to higher precision. For measured field data, the total gain of the system was set in advance using a neutral density filter to avoid detector saturation over snow surfaces, while the noise values were nearly constant, likely determined by the digitizer and receiver electronics. This should result in data with maximum signal-to-noise ratios over flat fine-grained snow surfaces, and lower signal-to-noise ratios over rough, sloping, and/or coarse-grained surfaces. Fortunately, the model results suggest that we should be able to recover grain sizes with small fractional errors when the grain sizes are large, even when the signal-to-noise ratios are relatively large.

3.4 Predicting biases in ICESat-2 measurements.

We predict expected biases in ICESat-2 data based on measured ATM waveform shapes by using our model to interpret the measured ATM waveforms, using the effective grain size as a proxy for the degree of subsurface scattering, then using the model again to estimate the range delay that would result from an ICESat-2 measurement over the same surface. To explain why this is necessary, we present a general statement of the magnitude of the bias (B) in an altimetry measurement estimated from a waveform $W_s(t)$, due to subsurface scattering:

$$B\big(M, W_s(t)\big) = M\big(W_s(t)\big) - M\big(W(t)\big) \qquad 6$$

Here $W_s(t)$ is the waveform including the effects of scattering, $W(t)$ is the waveform excluding the effects of scattering, and $M()$ is a metric used to derive height measurements from waveforms (referred to here as a *retracker*). The ICESat-2 ATL06 algorithm (Smith et al., 2019b) provides a standard land-ice height parameter, h_{li}, that is based on the median photon elevation within a small (typically ±1.5 m) window around the surface. Ideally, to evaluate the expected biases in this parameter, we would use measured ATM waveforms to approximate $W_s(t)$, and use the ATM IRF to approximate $W(t)$, which would let us directly use Eq. 6 to calculate expected biases with the windowed waveform median as $M()$. This is not practical, however, because most ATM waveforms include digitizer output that is less than zero (see Fig. 2). ICESat-

Formatted: Superscript

Formatted: Superscript

Deleted: To evaluate the resolution and accuracy of this fitting procedure, we generated a set of test waveforms based on $I_{est}(t)$, for a range of grain sizes, pulse amplitudes, and broadening values. To each sample in each of these simulated waveforms, we added a random value drawn from a Gaussian distribution with a standard deviation of two digitizer counts. To demonstrate that the algorithm can converge when the input does not match the search data exactly, we generated the input test data for grain-size values that were offset from the set of searched values by half the spacing between the searched values. Figure 3 shows the relationship between the specified and recovered grain size for small amplitudes and large broadening values (A = 90, σ=2 ns), and for large amplitudes and small broadening values (A=225, σ=0 ns). For surfaces with little broadening (σ=0 ns), the fitting procedure consistently recovers grain sizes as small as 20 µm, converging to either the next larger or the next smaller grain size value among the searched values (separated by 10%) with a moderate preference for the next smaller value, giving recovered values whose distribution width (5th to 95th percentile) is on the order of 10%. For smaller amplitudes and larger roughness, the width of the recovered distribution increases with decreasing grain size, with distributions spanning a factor of 5 for r_{eff}=50 µm. For the smallest input grain sizes with the small-amplitude rough input, the waveform that best fit the simulated waveform was often the one with no scattering, so the bottom of the distribution is not constrained on a log scale.

Deleted: returned

Deleted: noise value

Commented [BS64]: *R1 - 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.*

Commented [BS65R64]: We revised this to specify "the RMS of the noise added to the synthetic waveform."

Commented [BS66]: *R1 - Lines 270-278: Much of this paragraph is dedicated to describing the effects ATM amplitude had (past-tense) on the uncertainty in the estimated grainsiz* ... [33]

Deleted: amplitudes

Deleted: however, there is no consistent relation between the reflectance of the surface and the amplitude of the received ... [32]

Deleted: AT

Commented [BS67]: R2: L280: Is the satellite not named "ICESat-2"? ... [34]

Commented [BS68R67]: Fixed.

Deleted: AT

Deleted: AT

Deleted: (Smith et al., 2019)

Deleted: .

Deleted: (6)

Deleted: bias

Deleted: figure

2 uses a photon-counting lidar, so the median elevation can be calculated directly from the distribution of photon heights within the window. For a waveform lidar, the waveform median can be approximated under the assumption that waveform's digitzer counts (W(t)) are proportional to the flux of photons into the detector:

$$T_{med}(W(t)) = t \,|\, \frac{\int_{t_0}^{t} W(t')dt'}{\int_{t_0}^{t_1} W(t')dt'} = 0.5 \qquad (7)$$

but if the relationship between the two is more complex (i.e. if I(t) in Eq. 3 is significantly different from a delta function), the waveform median may not be equal to the median time for the energy incident on the detector. This appears to be the case for ATM, where the recorded waveforms include negative values, implying a more complicated relationship between the photon flux and the recorded values.

Since we cannot apply the median retracker directly to the ATM waveforms, we model the effects of subsurface scattering on ATL06 biases by using Eq. 3 to generate synthetic scattering-affected waveforms for a range of grain sizes, based on an estimate of the ICESat-2 system IRF derived from pre-launch calibration measurements (Smith et al., 2018). We then use Eq. 6 to predict the bias in the ATL06 measurements from each modelled waveform. Figure 4 shows the expected range bias for three retrackers as a function of grain size: the median retracker applied to the ICESat-2 IRF (the ATL06 *h_li* parameter), for a windowed mean on the same IRF (the ATL06 *h_mean*), and for a 15%-threshold centroid retracker (the metric used to track ATM waveforms), using the ATM IRF. The biases are smallest for the median retracker for the ICESat-2 waveform, increasing from sub-centimeter levels for r_{eff} < 10 µm to around 35 cm for r_{eff} > 10000 µm. The mean-based ICESat-2 bias is around twice as large as the median-based ICESat-2 bias, and the ATM bias is a few percent larger than the ICESat-2 median. This plot illustrates one difficulty in measuring ICESat-2 subsurface-scattering biases using laser-altimetry data as a reference: Over coarse-grained surfaces, ATM measurements are expected to have approximately the same biases as ICESat-2 measurements.

Deleted: , and a waveform-based median is not defined for a waveform that is not uniformly non-negative.

Deleted: Instead

Formatted: Indent: First line: 0.5"

Commented [BS69]: *R1 - Line 290: Please consider elaborating further on why negative numbers affect the waveformmedian 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.*

Commented [BS70]: *R1 - 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.*

Deleted: as a function of grain siz

Deleted: e.

Deleted: in

Deleted: h_{li} as a function of grain size using the ICESat-2 IRF

Deleted: ,

Deleted: for the ATL06 windowed median,

Formatted: Font: Italic

Deleted: equivalent to

Deleted: parameter on ATL06

Deleted: : the centroid of the part of the waveform that has digitized values greater than 15% of the waveform's maximum

Deleted: AT

Deleted: S

Deleted: S

Deleted: falls in between the two, and increasing sharply for the larger grain sizes as the portion of the waveform tail that exceeds the 15% threshold increases with grain size

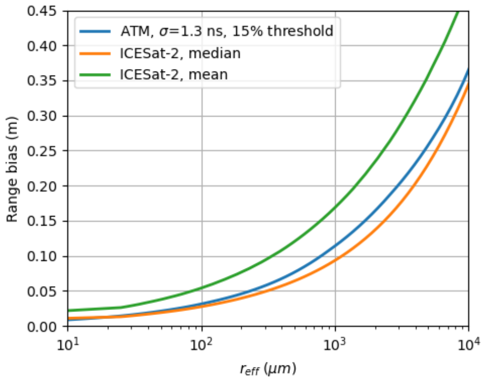

Figure 4. Predicted range bias for ATM and ICESat-2 waveforms. ATM biases are calculated using a mean-based retracker with a 15% amplitude threshold. ICESat-2 biases are calculated using a windowed median and a windowed mean retracker.

3.5 Subsurface-scattering-bias correction based on ATM and OLCI

To systematically correct ICESat-2 measurements, we need a spatially and temporally contiguous map of estimated subsurface scattering biases. In principle we could do this in two stages, using maps of grain size based on optical reflectance measurements (i.e. from OLCI), to interpolate a grain-size value for each ICESat-2 elevation measurement, and then calculating a range bias based on the relationship between grain size and bias using Eqn. 6. The accuracy of such a correction depends on the accuracy of the interpolated grain sizes and on the accuracy of the range bias predicted for each grain size. In particular the accuracy of the predicted range biases depends on whether the same scattering processes that influence the range bias determine the surface reflectance, which may not be true in all cases. For example, OLCI and AVIRIS-NG grain-size estimates are based in part the reflectance of infrared light (Nolin and Dozier, 2000; Vandecrux et al., 2022b), which does not penetrate as far below the snow surface as green light does (Smith et al., 2018), so the reflectance-based measurements may be more sensitive to near-surface layers than ICESat-2 would be. An ICESat-2 bias predicted based on surface reflectance measurements using our nominal 400 kg m^{-3} will also be imprecise by up to a factor of two for snow and ice surfaces with smaller or larger densities.

In contrast to reflectance-based grain-size estimates, ATM-waveform-based grain-size estimates involve the same physical processes involved in ICESat-2 subsurface scattering biases. This implies that if we use the same model to interpret ATM waveform shapes that we use to predict ICESat-2 biases, the

Deleted: were

Deleted: were

Deleted: We then use this relation between the IS2 median bias and grain size to estimate the temporal and spatial variation of ICESat-2 height biases based on ATM data. Although the estimated grain sizes may not be correct because of errors in the assumptions that we have made about the density of the ice and snow in the near subsurface, the recovered grain size should be an accurate way of describing how subsurface scattering affected the measured waveform, so the predicted h_{li} bias for a given recovered grain size should be consistent with the conditions that produced the ATM waveform, despite errors in the grain-size estimates related to surface-density variations. We note that this plot implies that ATM elevation products are likely to include decimeter-scale biases over surfaces with substantial subsurface scattering.

Formatted: Heading 2

Formatted: Indent: First line: 0.5"

Formatted: Superscript

Deleted: ,

Deleted: (

predicted ICESat-2 bias for a given recovered grain size should be consistent with the conditions that produced the ATM waveform, despite errors in the grain-size estimates related to surface-density variations. For this reason, we believe that we can use predicted biases based on ATM grain-size estimates to evaluate bias corrections based on OCLI grain-size estimates, even if the OLCI and ATM grain-size estimates do not agree on a point-for-point basis.

The simplest way to calculate an OLCI-based correction to the ATL06 *h_li* parameter is:

$$B(x,y,t) = B_{med}(r_{OLCI}(x,y,t)) \qquad 8$$

Here B(x, y, t) is the estimated bias at position (x, y) and time t, $r_{OLCI}(x,y,t)$ is the grain size estimated from the OLCI data at the same location and time, and B_{med} is the median ATL06 bias predicted using Eq. 6 and the ICESat-2 IRF. Based on our assumption that subsurface scattering affects ATM waveforms in the same way it affects ICESat-2 photon distributions, we treat biases based on ATM grain-size estimates as representative of the biases that would affect ICESat-2 if it measured the surface at the place and time where ATM made its measurements. This lets us evaluate B(x, y, t) by comparing it against $B_{med}(r_{ATM}(x,y,t))$, the ATL06 bias estimated for the grain size measured by ATM at the same location. Thus, the statistics of $B_{med}(r_{ATM}(x,y,t)) - B_{med}(r_{OLCI}(x,y,t))$ should allow us to estimate the statistics of ICESat-2 ATL06 data corrected using based on OLCI grain-size estimates.

As we will see in section 4.5, the OLCI measurements appear to become less sensitive to grain-size variations when the surface grain size is small. This leads us to also evaluate a threshold-based adjustment to the OLCI correction:

$$B_{thr}(x,y,t) = \begin{cases} B_0 & : r_{OLCI}(x,y,t) < r_{thr} \\ B_{med}(r_{OLCI}(x,y,t)) & : r_{OLCI}(x,y,t) > r_{thr} \end{cases} \qquad 9$$

Here r_{OLCI} is the OLCI-estimated grain size, $B_{med}(r_{OLCI})$ is the model predicted bias, r_{thr} is the threshold grain size above which the model produces reliable bias estimates, and B_0 is a constant bias value used for OLCI grain sizes smaller than r_{thr}. We can use the distribution of recovered grain size values to find values of B_0 and r_{thr} minimize the mean and spread of $B_{med}(r_{ATM}(x,y,t)) - B_{thr}(r_{OLCI}(x,y,t); B_0, r_{thr})$.

3.6 Robust measure of spread

Throughout the results of this study, we will measure the width of distributions using the *robust spread*, which we define as half the difference between the 16th and 84th percentiles of a distribution. This is analogous to the standard deviation of a normal distribution, in which the central 68% of the distribution falls within one standard deviation of the mean. It allows us to characterize the spread of the central peaks of distributions

Deleted:)
Deleted: predicted
Formatted: Font: Italic
Deleted: ¶
Deleted: distributions of photons measured by ICESat-2
Formatted: Heading 3
Deleted: sections
Formatted: Superscript
Formatted: Superscript

that are not necessarily normally distributed, and for which the standard deviation might be dominated by large outlying values.

4. Results

4.1 Recovered snow grain sizes from ATM and AVIRIS-NG

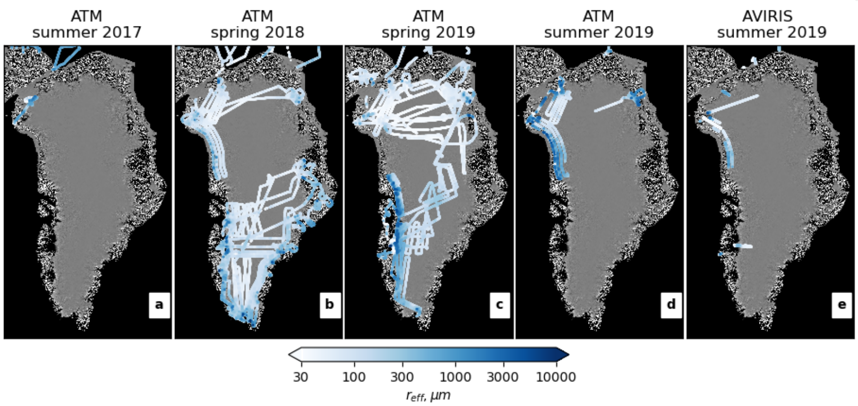

Figure 5. Recovered snow grain sizes from ATM and AVIRIS-NG. Colored points indicate recovered grain sizes for four ATM campaigns (a-d) and for AVIRIS-NG (e). Each color-coded points indicates a 1-km block median of recovered grain sizes, and the points have been plotted in order of grain size, so that coarser grain sizes overprint finer grain sizes. Background is the Mosaic of Greenland from 2012 (Scambos et al., 2012).

We obtained grain-size estimates from ATM for the summers of 2017 and 2019, and from the springs of 2018 and 2019. Figure 5 shows maps of recovered grain size from ATM and the valid AVIRIS-NG surveys for the late summer of 2019. These maps show a trend from large grain sizes at low elevation to small grain sizes at higher elevation, with notably larger grain sizes in the summer than in the spring where surveys overlap. The southern portion of the spring-2018 survey (Fig. 5b) was carried out earlier in the season than the corresponding portion of the spring-2019 survey (Fig. 5b), and encountered finer grain sizes, particularly along the coast, while grain sizes in the northern parts of both of these surveys were consistently fine. The summer surveys in 2017 (Fig. 5a) and 2019 (Fig. 5d) both encountered coarse grain sizes, particularly in the coast-parallel lines in 2019. The AVIRIS-NG survey from 2019 (Fig. 5e) has most of its overlap with the contemporaneous ATM survey along two coast-parallel lines, but a third coast-parallel line where ATM measured some of the coarsest grain sizes of the campaign was not covered.

Deleted: 3

Deleted: 3

Deleted: A-D

Deleted: E

Commented [BS71]: *R1- 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.*

Commented [BS72R71]: R2 — Fig. 5: AVIRIS is misspelled in Panel E

Commented [BS73R71]: Fixed

Deleted: for the summer of 2017, the spring seasons of 2018 and 2019, and the late summer of 2019,

Commented [BS74]: *R1 - 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*

Deleted: coarser

Deleted: by usable AVIRIS data.

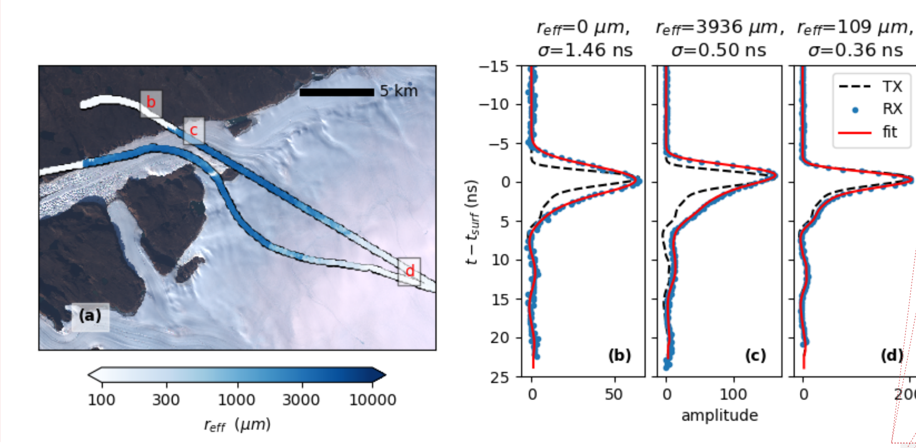

Figure 6. Grain size and waveforms. (a) True-color Landsat image of the Northeast Greenland ice sheet near Leidy Glacier from 6 August, 2019, with estimated effective grain size (r_{eff}) from ATM data collected 4 August 2019. For the ATM data, we plot the results of a 100-m blockmedian applied to r_{eff} and draw the outline of the swath in black. Panels (b), (c), and (d) show measured (RX) and best-fit modeled waveforms (fit), for three locations, as well as the input transmitted pulse (TX), scaled to match the amplitude of the received pulse. Bounding coordinates for panel (a) are presented in table S1.

To illustrate the spatial patterns of grain-size estimates recovered over a glacier during the melt season, Figure 6a shows a map of recovered grain size from Leidy glacier, northeast Greenland in the summer of 2019. We also show three waveforms, one measured over a rock/soil surface (Fig. 6b), one over low-elevation coarse-grained melting ice (Fig. 6c), and a third from finer-grained snow (Fig. 6d), as well as the corresponding best-fitting waveforms. The rock/soil waveform shows some broadening relative to the transmitted waveform, likely due to surface roughness, that is symmetric in time, with equal distortion of the upper and lower slopes of the waveform. The best fitting model waveform has an r_{eff} value of 0 μm, and a σ value of 1.46 ns. The coarse-grained waveform (Fig. 6c) is also broader than the transmitted waveform, but has different amounts of distortion for the leading (upper) and trailing (lower) edges of the waveform: It has a sharply sloping upper edge, but a more gradual slope on the lower edge, which is consistent with the predicted effects of subsurface scattering. The best-fitting model waveform has an r_{eff} value of 2896 μm, and a σ value of 0.26 ns. The higher-elevation waveform (Fig. 6d) has much less distortion than the low-elevation waveform, with a shape much more similar to the transmitted pulse, which is reflected in the best-fitting model parameters of r_{eff} =109 μm, σ = 0.26 ns. Elevations measured by ATM show that the outlet section of the glacier (near (c)) is at 400-500 m, and elevation increases to around 1200 m near (d). The mapped distribution of grain sizes (Fig. 6a)

Commented [BS75]: *R1 - Figure 6a: The Blues colormap used to represent estimates grainsize 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 grainsize estimates based on the ranges presented in Lines 250-251.*

Commented [BS76R75]: I (Ben) like the Blues colormap for this figure, in part because the blue-white ramp is analogous to the blue appearance of coarse grained snow and the white appearance of fine grained snow. To help distinguish the part of the track that extends to the right of the figure, we revised the figure so that the track is outlined in black.

Deleted: Grain size

Deleted: A.

Deleted: grain size

Deleted: reff

Deleted:)

Deleted: reff

Deleted: B

Deleted: C

Deleted: D

Commented [BS77]: *R1 - Figure 5 and Figure 6a: Please include coordinates for all maps.*

Commented [BS78R77]: Please see my initial comment about coordinates on maps. The bounding coordinates for 6A are now specified in supplemental table 1.

Deleted: grain size

Deleted: ,

Deleted: and t

Deleted: measured in Greenland

Deleted: panel

Deleted: B

Deleted: pane

Deleted: l

Deleted: C

Deleted: panel

Deleted: D

Deleted: grain size

Deleted: panel

Deleted: A

shows little or no subsurface scattering on rock and soil ($r_{eff} \approx 0$), strong subsurface scattering for low-elevation ice ($r_{eff} > 1000\ \mu m$), and weaker subsurface scattering at higher elevations ($r_{eff} < 200\ \mu m$). We suggest that the lower-elevation part of the glacier on the left-hand part of Fig. 6a has experienced stronger surface melt than the higher-elevation part to the right, which is roughly consistent with the gradient from bluer to whiter tones in the background Landsat image collected two days later.

4.2 Comparisons of recovered snow grain sizes between two independent ATM instruments

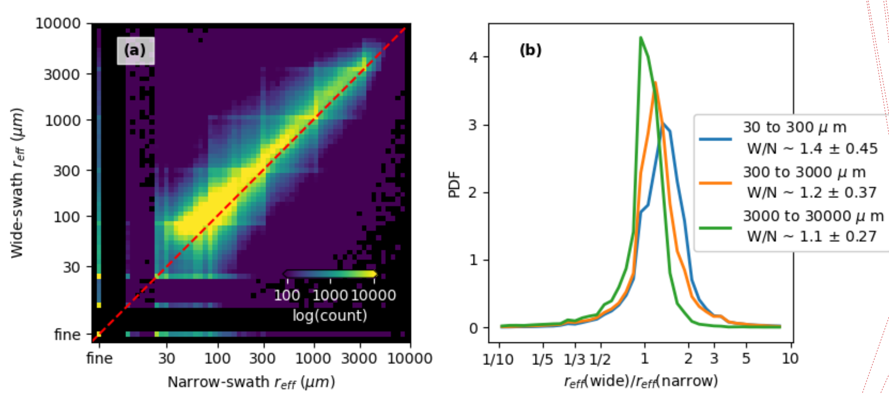

Figure 7. Recovered snow grain sizes from two ATM systems from the Summer 2019 campaign. Panel a shows the density of measurements as a function of recovered r_{eff} values from the narrow and wide-scan ATM systems (lighter colors represent a higher density of measurements). Points for which one of the systems found a best match with a scattering-free model waveform are reported along the rows/columns marked 'fine'. Panel b shows the distribution of wide-to-narrow r_{eff} ratios for different ranges of narrow-swath r_{eff}. The legend for panel b gives the median and robust spread of the ratios for each range.

Because the wide- and narrow-swath ATM instruments were installed on the same aircraft, there are abundant opportunities to compare measurements of the same surface at essentially the same time between the two, as a check on the self consistency of the measurements and as a check on whether the recovered grain size depends strongly on the incidence angle of the laser beam. Figure 7a shows a two-dimensional histogram of grain-size estimates from the wide and the narrow ATM sensors from the summer 2019 campaign. The estimates are clustered close to the 1:1 line, with slightly larger grain-size estimates from the wide-swath

Commented [BS79]: *R1 - 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 grainsize lower portion and the finer grainsize upper portions is not much. Has extensive melting been so highly concentrated in only the lower section*

Commented [BS80R79]: There is a strong elevation gradient across the image, which likely leads to the range of grainsizes. We now specify that our statement about the grain sizes is an interpretation rather than a certainty:
Elevations measured by ATM show that the outlet section of the glacier (near C) is at 400-500 m, and elevation increases to around 1200 m near D. The mapped distribution of grain sizes (panel A) shows little or no subsurface scattering on rock and soil (, strong subsurface scattering for low-elevation ice (, and weaker subsurface scattering at higher elevations (. We suggest that the lower-elevation part of the glacier on the left-hand part of panel 6A has experienced stronger surface melt than the higher-elevation part to the right of panel 6A, which is roughly consistent with the gradient from bluer to whiter tones in the background Landsat image collected two days later.

Deleted: that has experienced extensive melt

Deleted: .

Deleted: 3

Deleted: grain size

Commented [BS81]: *R1 - Figure 7a: Please include a colorbar as is done for Figures 9 and 10.*

Deleted: grain size

Deleted: A

Deleted: B

Deleted: B

Deleted: standard deviation

Commented [BS82]: *R1 - 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 leg* ... [35]

Deleted: self consistency

Deleted: -

Deleted: A

Deleted: grain-size estimates

Deleted: two

Commented [BS83]: L361: It would be useful to name ... [36]

Commented [BS84R83]: Named.

Commented [BS85]: R1 - Line 363: To me "around" d ... [37]

Commented [BS86R85]: Revised to 'close to'

Deleted: around

Deleted: grain-size estimates

instrument. The histogram shows horizontal and vertical streaks that correspond to grain-size values that the fitting algorithm selects preferentially as part of the effort to reuse previously computed model waveforms. These likely reflect small reductions in the accuracy of the recovered grain-size estimates, although not obviously to any large extent. For grain sizes smaller than around 25 µm, the fitting process for both datasets often selects a model waveform with no scattering model applied as best fitting the measurements. This results in a reduced number of recovered values at r_{eff}<25 µm, and spikes in the histogram for values where one or both estimates selected the scattering-free waveform. For display purposes, we have mapped these to the left of and below the range of possible fit values (labeled 'fine' in Fig. 7a). The two sets of measurements appear to be consistent for grain sizes as small as 30 µm, and the two datasets report effective-zero grain sizes (< 10 µm) for most of the same points: for 85% of points for which the wide swath grain size effectively zero, the narrow swath was also, and for 70% of points for which the narrow-swath grain size was effectively zero, the wide-swath grain size was also.

The distribution of ratios between the recovered grain sizes for the two systems is similar to a lognormal distribution, with a central parameter close to unity. Figure 7b shows histograms of ratios between wide-swath and narrow-swath estimates, for three ranges of grain sizes (as determined from the narrow-swath values). For large grain sizes (> 3000 µm) the median ratio is 1.1, with a robust spread of 0.27; the bias and spread increase with decreasing grain size, and for small grain sizes (30 to 300 µm) the median ratio is 1.2, with a spread of 0.45. One possible reason for the larger grain-sizes estimates from the wide-swath instrument is that the wide-swath beam had a larger incidence angle to the surface, so the return waveforms had somewhat larger Gaussian broadening. Our experiments with simulated data (section 3.3) suggest that 1 ns of pulse broadening can result in a small positive bias in recovered grain size for the 30-100 µm range of input sizes.

- **Deleted:** grain-size values
- **Deleted:** grainsize estimates
- **Deleted:** grain size
- **Deleted:** F
- **Commented [BS87]:** BS: missing "these"
- **Commented [BS88R87]:** Fixed
- **Deleted:**
- **Deleted:** 6
- **Deleted:** A
- **Deleted:** grain size
- **Deleted:** grain size
- **Deleted:** grain size
- **Deleted:** grain size
- **Deleted:** grain size
- **Deleted:** B
- **Deleted:** grain size
- **Deleted:** grain size
- **Deleted:** (equal to the half-width of the central 68% of the distribution, approximately the one-sigma range)
- **Deleted:** grain size
- **Deleted:** grain size
- **Deleted:** , indicating close agreement between the two systems.
- **Deleted:**

4.3 Comparisons between snow grain sizes derived from ATM and AVIRIS-NG

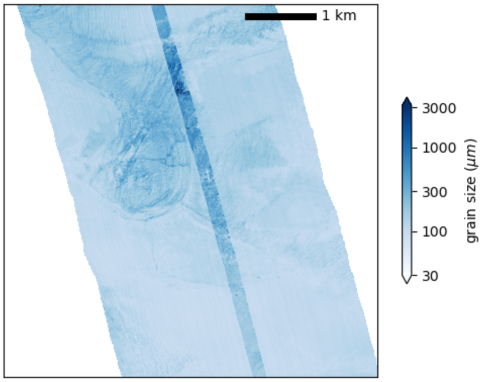

Figure 8. Sample of AVIRIS-NG- and ATM-derived snow grain-size estimates for a coastal location in Greenland. The grain size based on the complete 4-km AVIRIS-NG swath is shown, with a 10-m block median of the recovered grain size from the 250-m wide-scan ATM swath superimposed on top. The scene center is approximately 75.314° N, 33.464°E, and contains data from the AVIRIS-NG granule ang20190906t144855 and the ATM granule ILATMW1B_20190906_133000.atm6T6.h5.

Grain-size estimates from ATM and from AVIRIS-NG show consistent spatial variations, which are most easily identified in areas where the grain size varies on short spatial scales. Figure 8 shows maps of grain-size estimates from the wide-swath ATM scanner and from AVIRIS-NG for a short segment of a flight path in northwest Greenland. Both datasets show a range of surface grain sizes, with variations that that appear to correspond to spatial variations in surface weathering, likely over a drained supraglacial lake basin. Despite the agreement between the small-scale variations, the ATM data show consistently larger gain sizes than AVIRIS-NG, particularly in the upper part of the scene in the roughest part of the lake basin.

Deleted: 3

Deleted: grain size

Deleted: grain-size estimates

Deleted: grain size

Deleted: grainsize

Commented [BS89]: R1 - 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?

Deleted: grain size estimates

Deleted: grain size

Deleted: The patterns of variation are very similar between the sensors, although

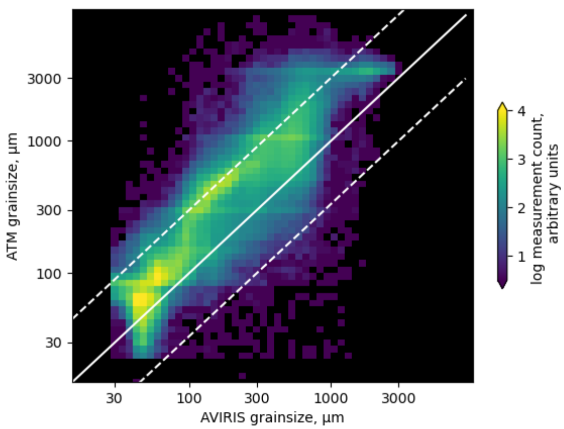

Figure 9. Two-dimensional histogram comparing AVIRIS-NG-derived grain size with narrow-swath ATM-derived grain size, with cells colored by the number of points observed. The solid white line shows the 1:1 relationship between the two datasets. To help illustrate the magnitude of the difference between the datasets, we plot two dashed lines that show the ATM : 3 x AVIRIS-NG (upper) and ATM : 1/3 x AVIRIS-NG (lower) relationships.

The general agreement between AVIRIS-NG and ATM grain-size estimates is illustrated by a comparison between 10-m blockmedians of narrow-swath ATM grain-size estimates and AVIRIS-NG grain-size maps interpolated at the locations of the ATM data (Fig. 9). This plot was generated based on all narrow-swath ATM waveform data available for the ice sheet, but excludes a single AVIRIS-NG transect measured on sea ice, as discussed in section 2.2. For grain sizes greater than about 50 µm, the two show a generally similar trend, although ATM grain sizes are typically around 2-3 times larger than the corresponding AVIRIS-NG grain sizes. This relationship does not hold towards the small-grain size side of the plot, where the AVIRIS-NG grain sizes are clustered in a near-vertical feature centered around 50 µm. We believe that this is because the AVIRIS-NG algorithm loses some of its sensitivity to grain size variations around 40-50 µm while, based on our synthetic-data experiments, we expect the ATM retrievals to be sensitive to grain sizes as small as 25 µm. The points where the ATM fit selected zero scattering are not shown in this plot; they amount to a small fraction (0.4 %) of observations.

Deleted: grain size

Deleted: grain size

Deleted: , the

Deleted: 2

Deleted: 0.5

Commented [BS90]: R2: Fig. 9: What is the justification for the position of the dashed lines? It would be useful to clarify that here.

Commented [BS91R90]: We now describe the choice of the position of the lines in the caption.

Deleted: ure

Deleted: shows a comparison between grain-size estimates from AVIRIS and those from ATM.

Commented [BS92]: R1 - 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?

Commented [BS93R92]: R2 — L398: Please could you clarify if this is **all** coincident grain sizes from Summer 2019 or just a sample.

Deleted: T

Deleted: grain size

Deleted: grain size

Deleted: is less pronounced

Deleted: grainsize

Deleted: we see more values where ATM grainsize approximately equal to, or smaller than, AVIRIS grainsize

Deleted: comes about

Commented [BS94]: R1 - 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.

Commented [BS95R94]: L401: "comes about" is clumsy, consider revising.

Commented [BS96R94]: We have revised this section to include the explanation for the vertical feature.

Commented [BS97R94]: "Comes about" replaced by "... [38]

Deleted: dataset

Deleted: includes few points smaller than about 40 µm, so points where the ATM fit estimated grain sizes smaller than this ar ... [39]

4.4 Comparison between OLCI, AVIRIS-NG, and ATM snow grain sizes

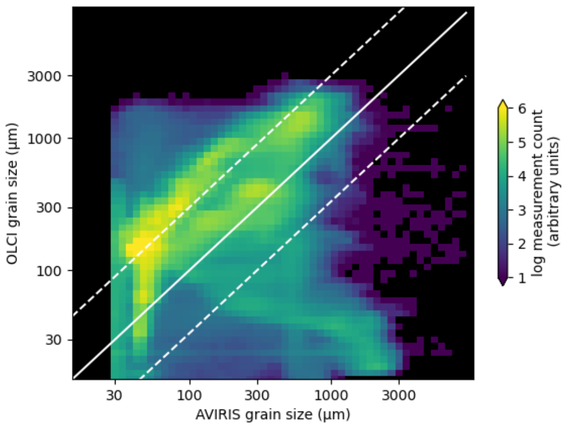

Figure 10. Comparison between AVIRIS-NG-derived grain sizes and OLCI-derived grain sizes. The solid white line shows the 1:1 relationship between the two datasets, the two dashed lines show the OLCI: 3 times AVIRIS-NG (upper) and OLCI : 1/3 AVIRIS-NG (lower) relationships. All OLCI measurements were collected within 1 day of the AVIRIS-NG measurement.

Direct comparisons between the AVIRIS-NG and OLCI grain-sizes help illustrate the reliability of each dataset on its own and in comparison with ATM. Figure 10 shows a 2-D histogram of AVIRIS-NG-derived grain sizes from the summer-2019 survey and OLCI-derived grain sizes collected within one day of the AVIRIS-NG measurements. The largest concentration of OLCI grain sizes is between three and four times larger than the corresponding AVIRIS-NG sizes. As in the comparison between ATM and AVIRIS-NG, there is a vertical feature in the distribution at AVIRIS-NG grain size = 40-50 µm, which likely corresponds to the fine-grained limit of the AVIRIS-NG data. The distribution of measurements for which the OLCI grain-size estimates are substantially finer than the AVIRIS-NG estimates may reflect contamination with undetected clouds in the OLCI imagery, which would tend to bias the OLCI estimates in the fine-grained direction.

Commented [BS98]: *R1 - 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?*

Deleted: 3

Deleted: satellite

Deleted: grain size

Deleted: grain size

Deleted: satellite

Deleted: grain size

Deleted: satellite

Deleted: satellite

Deleted: satellite

Deleted: ½

Deleted: ,

Deleted: shows a comparison between

Commented [BS99]: R2: L411: Which years?

Commented [BS100R99]: We now specify 2019

Deleted: grain size

Deleted: satellite

Deleted: grain size

Deleted: satellite

Deleted: grain size

Deleted: grain size

Deleted: satellite

Deleted: grain-size estimates

Deleted: satellite

Deleted: satellite

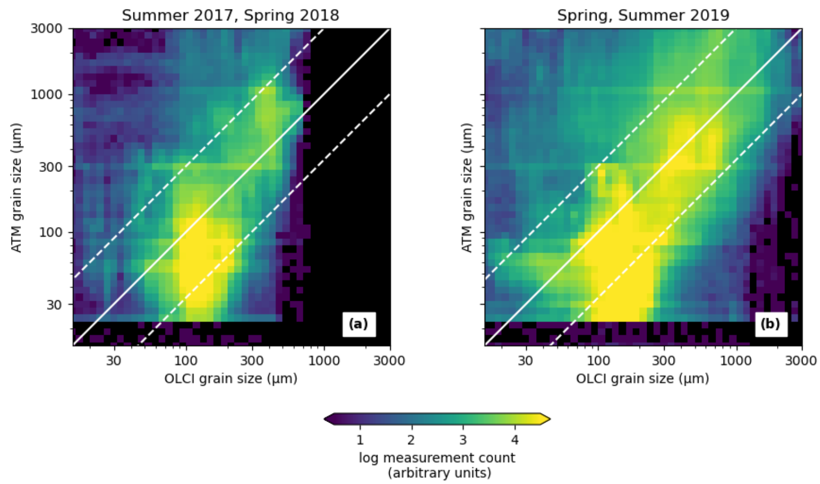

Figure 11. Comparison between narrow-swath-ATM-derived and OLCI-derived grain-size estimates. Panel A shows the distribution for the summer (July) of 2017 and the spring (March-May) of 2018, Panel B shows the distribution from the spring (April-May) and summer (September) of 2019. In both plots, the ATM grain sizes are derived from a 10-meter blockmedian of the data, and only those points for which the time difference between the OLCI measurement and the ATM measurement was less than 3 days are included. The solid lines indicates the 1:1 relationship between the datasets, the dashed lines indicate the 1:3 and the 3:1 relations.

Similarly, Figure 11 shows a comparison between OLCI-derived grain sizes and those from the narrow-swath ATM instrument, based on a combination of data from the summer of 2017 and the spring of 2018 (Fig. 11a) and from the spring and summer of 2019 (Fig. 11b). In each case, the distributions of both types of grain-size measurements roughly follow the 1:1 line, although for both years, the ATM measurements show a range of measurements smaller than 100 µm for which the OLCI measurements are clustered around 100 µm. This may indicate that there are conditions under which the OLCI measurements cluster around a moderately small grain size while ATM maintains sensitivity at smaller grain sizes. The 2017-2018 panel (fig. 11a) contains far fewer points with large grain sizes because the dataset for the Summer of 2017 has very limited spatial coverage compared to the summer of 2019, and the Spring-2019 dataset covered more melting surfaces than did the Spring-2018 dataset.

4.5 Comparing subsurface-scattering range bias estimates between OLCI and ATM data

Deleted: satellite

Deleted: grain-size estimates

Deleted: grain size

Deleted: satellite

Commented [BS101]: Fig. 11: I think the satellite grain size should be on the y-axis to be consistent with Fig. 10.

Deleted: satellite

Deleted: grain size

Deleted: 0

Deleted: A

Deleted: 0

Deleted: B

Deleted: grain size measurements

Commented [BS102]: R1 - 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?

Commented [BS103R102]: The correlation is not especially strong, and there is a substantial background of low-quality OLCI measurements evident in this plot. This general statement of the relationship between the two datasets is about as strong a statement as the data allow.

Deleted: satellite

Deleted: satellite

Deleted: grain size

Deleted: grain size

Deleted: A

Deleted: grain size

Deleted: 3

Deleted: satellite

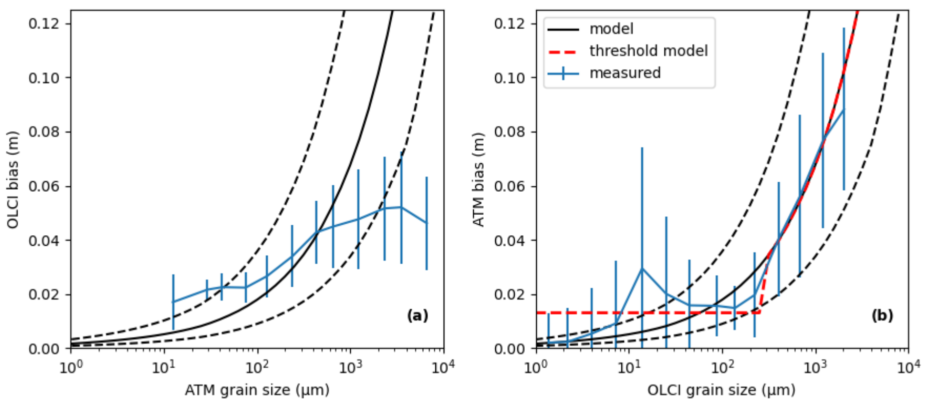

Figure 12. Range biases as a function of snow grain-size estimates for the complete 2017-19 dataset. (a) shows range biases predicted from OLCI grain-size estimates as a function of ATM grain size, (b) shows range biases estimated from ATM grain sizes as a function of OLCI grain-size estimates. For each panel, the vertical bars show the standard deviation of the range bias estimates for each grain size value, the black solid curve shows the modeled range bias as a function of grain size, and the dashed lines show the factor-of-two uncertainties in the model related to surface density.

Comparing grain sizes estimated from the different sensors (Figs. 9-11) demonstrates the consistency (or lack thereof) between the datasets, but to address the usefulness of OLCI data in correcting biases in ICESat-2 data, we need to compare biases predicted for ICESat-2 based on OLCI with biases estimated based on ATM waveforms. In these comparisons, the accuracy of the sensor is most important for large grain sizes because ICESat-2 biases predicted by our model (Fig. 3) are approximately zero for small grain sizes, so any correction we calculate will be small, with larger corrections expected for larger grain sizes.

To illustrate the ability of the OLCI data to predict the ICESat-2 bias at locations where ATM grain size estimates are available, we plot the ICESat-2 bias predicted based on OLCI measurements as a function of ATM-derived grain size (Fig. 12a). In this plot, we collect groups of ATM measurements in logarithmic bins with a spacing of $10^{0.25}$ and calculate the median and robust spread of biases of the biases calculated for the corresponding OCLI grain sizes. If we assume that biases calculated based on ATM waveforms are approximately correct, this suggests that the OLCI bias estimates underestimate the sensitivity of ICESat-2 biases to grain size. When we reverse the way in which we calculate the statistics and calculate the distribution of ICESat-2 biases predicted from ATM measurements as a function of OCLI-estimated grain size (Fig. 12b), we see a closer match between the ICESat-2 biases predicted based on the ATM data and those predicted

Deleted: grain size estimates

Deleted: A

Deleted: satellite

Deleted: grain-size estimates

Deleted: grain size

Deleted: B

Deleted: grain size

Deleted: satellite

Deleted: grain-size estimates

Deleted: grain-size

Deleted: grain size

Commented [BS104]: R1 - 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.

Commented [BS105R104]: R2 — Fig. 12: It took me a minute to figure out what this figure was showing and I think it is because the word "satellite" is being used to refer to both Sentinel-3 grain size and ICESat-2 range bias. Please modify the axes labels to clarify. Also if the black lines show the modeled range bias as a function of grain size then the x-axis label should just be "grain size" and the legend should provide information about where the grain sizes came from.

Deleted: f

Deleted: f

Deleted: f

Deleted: f

based on the OLCI measurements, at least for OLCI-estimated grain sizes larger than around 250 µm. At smaller grain sizes, the ATM-derived ICESat-2 bias estimates deviate from the OLCI biases, with a roughly uniform value close to 0.02 m for OLCI-derived grain sizes between 20 and 100 µm, a small peak for OLCI biases close to 15 µm, and approximately zero bias for finer grain sizes.

The two plots in Fig. 12 cover different ranges of grain sizes because of the different ways that the two sensors sample the ice sheet. Fig. 12a includes large values of grain size from ATM (up to around 4000 µm) because single ATM measurements occasionally sample features on the surface with large grain sizes, and includes no ATM measurements with grain sizes smaller than 30 µm because for smaller grain sizes, ATM often reports zero scattering. In Fig. 12b, grain sizes larger than 1000 µm do not appear, because the 1-km OLCI pixels rarely measure the small features where coarse grain sizes are observed. For the smallest OLCI-derived grain sizes, it appears that ATM often returned no-scattering estimates, so the estimated bias is effectively zero for both datasets.

4.6 Calculating a best-feasible correction.

Based on Fig. 12b, it appears reasonable to believe that OLCI grain-size estimates provide useful information about subsurface delays for coarse-grained snow, but not for fine-grained snow. To better account for this lack of sensitivity in OLCI at fine grain sizes, we used the ATM and OLCI grain sizes from 2017 - 2019 to find optimal parameter values for the threshold bias model (Eq. 9): For a range of B_0 and r_{thr}, we calculated the median and the robust spread of the distribution of ATM biases corrected using on the OLCI grain sizes, $B_{med}(r_{ATM}) - B_{thr}(r_{OLCI})$. To help match the resolution between the ATM and the OLCI grain-size estimates, we carried out these calculations on a 250-m blockmedian of the ATM measurements. Figure 13 (a,b) show how the median and the robust spread depend on the parameter values. For threshold values greater than about 150 µm, there is a fine-grain-bias (B_0) value that gives a median residual of zero, and for each fine-grain bias, there is a threshold value that gives the minimum robust spread; these curves intersect at B_0=0.012 m, r_{thr}=270 µm. Figure 13c shows the distributions of ATM-derived biases, ATM-derived biases corrected based on $B_{med}(r_{OLCI})$, and of ATM-derived biases corrected based on the optimized $B_{thr}(r_{OLCI})$ model. The uncorrected distribution of ATM-derived biases has a peak at around 0.01 m a median of 0.013 m, with a substantial tail of values extending in the positive direction, representing coarse-grained parts of the ice sheet where we would predict that ICESat-2 would measure elevations several cm too low. Applying the unmodified correction results in a more compact distribution of residuals, with a median of -0.007 m and a spread of 0.006 m, both of which are an improvement on the raw distribution but the bias is now in the opposite direction. The optimized threshold model yields a distribution of residuals with a zero median and a robust spread of 0.004 m.

Commented [BS106]: R2 — L485-489: I think the first half of this paragraph should be removed (or placed later in the discussion) because, as it is written, it seems like the main takeaway is the consistency of the different snow grain size estimates. However, there are substantial biases between the estimates (Figs. 9-11) which the authors are up front about later in the discussion and should be the focus. ... [40]

Deleted: f

Formatted: Line spacing: 1.5 lines

Deleted: f

Deleted: Figure 12 shows ICESat-2 range biases predicted ... [41]

Deleted: 3.4

Deleted: Based on figure 12, panel B, it appears reasonabl ... [42]

Deleted: the

Deleted: satellite

Deleted: grain size

Deleted: for this model

Deleted: *thr*

Commented [BS111]: R1 - Line 459: Could the autho ... [43]

Commented [BS112R111]: We realize that this is not ... [44]

Deleted: corrected

Deleted: *sat*

Deleted: For the sake of computational efficiency, and t

Deleted: satellite

Deleted: grain-size estimates

Deleted: panels

Deleted: A, B

Deleted: ed

Deleted: threshold

Deleted: 3

Deleted: 8

Deleted: Figure 13c shows the

Deleted: *sat*

Deleted: *sat*

Deleted: using the optimized parameters

Deleted: 1

Deleted: c

Deleted: 6

Deleted: 8

Deleted: 8

Deleted: biased

Deleted: 6

The preceding analysis used robust statistics (i.e. the median and robust spread), which show how the correction works for typical locations on the ice sheet (i.e. ignoring the most extreme scattering conditions), which we would expect to fall in the middle of our distribution of residuals. However, many users of altimetry data will explicitly or implicitly perform their analysis using non-robust statistics (i.e. by calculating mean elevation differences, or calculating the standard deviation of elevation differences). To show how the corrections work with statistics that are more sensitive to outlying values, we repeated the analysis using the mean and the standard deviation of the corrected datasets. This yields similar optimum B_0 and r_{thr} values (0.014 m and 260 µm, respectively) for the zero-mean-residual model with the smallest standard deviation, but finds that for this model, the standard deviation is approximately the same as that for the non-optimized correction (0.011 m vs. 0.012 m). This shows that with the right parameters, the correction can produce a near-zero corrected mean, but cannot necessarily make a substantial improvement in the standard deviation of the corrected data.

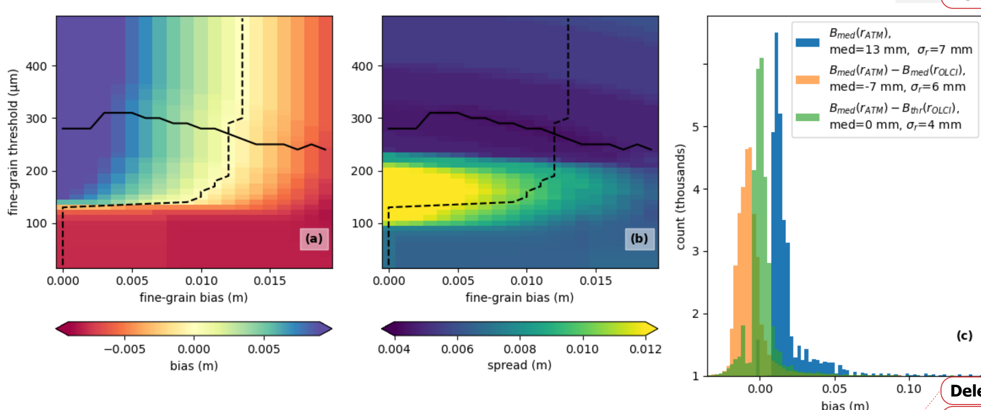

Figure 13. Tuning the threshold correction for ATM-based ICESat-2 bias estimates. Panels (a) and (b) show the median and robust spread of the distribution of ATM-derived ICESat-2 bias estimates corrected with the threshold model (Eq. 9) for different values of the fine-grain bias (B_0) and fine-grain threshold (r_{thr}). The dashed curves show the fine-grain bias corresponding to the minimum absolute value of the median for each value of the threshold, and the solid lines show the fine-grain threshold corresponding to the minimum value of the spread for each value of the fine-grain threshold. Panel (c) shows histograms of uncorrected bias estimates, bias estimates corrected based on $B_{med}(r_{OLCI})$ $(Eq. 8)$, and bias estimates corrected based on $B_{thr}(r_{OLCI})$ $(Eq. 9)$ for the optimum parameters, B_0=0.012 m, r_{thr}=270 µm. The median and robust spread of each distribution is given in the legend.

Deleted: We present this analysis using the median and robust spread because they are less affected by outlying data points.

Deleted:

Deleted: R

Deleted: ing

Deleted: yields similar

Deleted: 5

Deleted: of the optimal model

Deleted: 4

Deleted: in either case

Deleted: A

Deleted: B

Deleted: (equal to the half width of the central 68% of the distribution) …

Deleted: equation

Deleted: 7

Deleted: C

Commented [BS114]: R1 - 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

Deleted: 3

Deleted: 8

5 Discussion:

The comparison of measurements between the narrow and wide-swath instruments (Fig. 7) shows that ATM-based estimates of snow grain size are consistent to within a factor of two or better between two independent instruments, and are not strongly influenced by measurement geometry except at small grain size, where the larger angle between the wide-swath beam and the surface produces blurring of the returned waveform. Based on our modelling results (Fig. 3) and the expected relationship between incidence angle and return pulse width (Section 3.2), we expect this to result in larger scatter and bias in the wide-swath grain-size estimates. As estimates of grain size, the two sets of measurements have biases and uncertainties due to our assumptions about the density of the snow, but as measurements of photon delays due to subsurface scattering, they are reasonably consistent and should be useful in predicting biases in ICESat-2 data.

The comparisons between AVIRIS-NG and ATM grain size (Fig.9), and those between AVIRIS-NG and OLCI-derived grain size (Fig. 10) both show the AVIRIS-NG estimates as biased by a factor of 2-3 towards fine grain sizes relative to the other dataset; further, both the ATM and the OLCI estimates appear to produce usable estimates of grain size that are smaller than 30 µm, while the AVIRIS-NG measurements seem to have a fine-grained limit of resolution around 40 µm. These differences between the AVIRIS-NG measurements and ATM-based measurements are consistent with comparisons between this AVIRIS-NG survey and observations of apparent elevation differences between green and near-infrared altimetry measurements that also implied that the AVIRIS-NG data had underestimated grain sizes (Fair et al., 2024). Despite these limitations, the comparisons between ATM, OLCI, and AVIRIS-NG measurements show broad agreement between the three sets of data, with larger grain sizes in each dataset corresponding to larger grain sizes in the others. However, this relationship is not as consistent as we might have hoped, and for a substantial fraction of the points there is no clear relationship between the grain sizes from the different sensors. Part of this scatter may result from differences in resolution between the datasets. ATM resolves grain size on a sub-meter-sized footprint, which we then degrade to 10 m using our blockmedian filter, the AVIRIS-NG data have a 5-meter pixel size, and the OLCI-based measurements are posted at 1 km. Many of the measurements showing the coarsest grain sizes from ATM are from small features such as crevasses and stream channels, which are likely not resolved by the larger pixel size of the OLCI measurements. Similarly, the smallest, coarsest-grained features in the AVIRIS-NG dataset are not expected to be resolved in the OLCI data.

There may also be differences between the retrieved grain sizes related to the measurement techniques. The ATM scattering measurements rely on subsurface multiple scattering that may sample hundreds or thousands of scattering events, and in which photons may penetrate hundreds of times the grain diameter below the surface. By contrast, the AVIRIS-NG and satellite measurements both use portions of the reflectance spectrum extending into the near infrared, where the attenuation length of ice is as small as a few cm (Warren, 1982).

Commented [BS115]: R1 - 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.

Commented [BS116R115]: We will add more references to this section

Deleted: 4

Deleted: f

Deleted: grain size…now grain size are consistent to within a factor of two or better between two independent instruments, and are not strongly influenced by measurement geometry except at small grain size…rain size, where blurring ... [45]

Deleted: f

Commented [BS117]: R1 - 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.

Commented [BS118R117]: We now provide a more detailed explanation of this, and refer to our more thorough treatment of return pulse shape in section 3.2.

Deleted: by the angle between the beam

Deleted: and the surface likely results in a large…o result in larger scatter and bias in the wide-swath estimates…rain-size estimates..… As estimates of grain size…rain size, the two sets of measurements have biases and uncertainties due to our assumptions about the density of the snow, but as measurements of photon delays due to subsurface scattering, they are both ... [46]

Deleted: grain size…nd ATM grain size…rain size (Fig.9), and those between AVIRIS-NG grain size…and satellite…LCI-derived grain size…rain size (Fig. 10) both show the AVIRIS-NG estimates as biased by a factor of 2-3 towards fine grain size…rain sizes relative to the other dataset; further, both the ATM and the satellite…LCI …estimates appear to produce usable estimates of grain size…rain size that are smaller than 30 µm, while the ... [47]

Deleted: 3

Deleted: (Fair et al., n.d.)…. Despite these limitations, the comparisons between ATM, satellite…LCI, and AVIRIS-NG measurements show broad agreement a consistent trend …e ... [48]

Commented [BS119]: R1 - 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 autho ... [49]

Commented [BS120R119]: We have weakened our first statement about the agreement between the datasets (now "broad agreement"). The rest of this seems like it says what we wa ... [50]

Deleted: grain size…rain size on a sub-meter-sized footprint, which we then degrade to 10 m using our blockmedian filter, the AVIRIS-NG data have a 5-meter pixel size, and the satellite ... [51]

Deleted: grain size

This means that the ATM measurements are sensitive to grain size over a much larger range of depths than are the reflectance-based measurements. Particularly under melting surface conditions, we expect to see a layer of finer-grained ice on top of coarse-grained or water-saturated deeper layers (Cooper et al., 2018), which would lead us to expect that the reflectance-derived grain sizes would be finer than those derived from ATM. This effect is not expected to be as important under colder conditions, especially where fresh snow is present at the surface, because returns from a snow layer a few centimeters thick will contain only a very small minority of photons that have experienced long path delays (Smith et al., 2018).

We believe that it is also likely that there are disagreements between reflectance-derived measurements of grain size and ATM-based measurements because of the simplified relationship we have used between grain size and scattering properties. Our model of subsurface scattering assumes that the scattering is from independent spheres of ice suspended in air, and that the density of the medium is 400 kg m^{-3}. In fact, surface densities in the accumulation zone are often lower than that assumed by our model (Medley et al., 2022) while ablation-zone densities can approach that of compact glacier ice (800 kg m^{-3} and higher), and the presence of liquid water in the snow can result in reduced scattering efficiency per grain compared to that expected for spheres in ice. Over fresh, low-density snow, we expect our ATM-based measurements to overestimate grain size because our model does not fully account for the path length between scattering events and assumes that the extra path delay comes about because of time spent traveling through ice grains. Over compact ice surfaces the situation is more complex, because the surface density is likely larger than our reference density, leading to an underestimate of grain sizes, but close packing of grains and the presence of water should each lead to less efficient scattering from each grain, leading to an overestimate of grain size. Under most circumstances, we expect the latter effects to be more significant, because the effect of density alone is unlikely to be larger than a factor of two (see Fig. 1).

The comparison between predicted ICESat-2 biases derived from ATM and those from the OLCI measurements (Fig.12) suggests that while OLCI measurements cannot accurately predict the measurement bias for each laser-based measurement, the mean bias at the kilometer scale is more likely to be reliable. The difference between the two ways of plotting the biases as seen in Fig. 12 likely relates to the spatial resolution of the two sensors. ATM, with sub-meter resolution, captures small-scale features on the ice sheet, including crevasses, water channels, and ponds that all have large grain sizes. These features do not appear in the OLCI maps, which reflect the average grain size over 1-km pixels, which results in underestimates of bias for the ATM measurements with coarse grain sizes. Conversely, the average over OLCI measurements shows good agreement with the predicted grain size-vs-bias curve, likely because the median biases for large, spatially distributed collections of ATM measurements are only weakly affected by the minority of ATM measurements collected over large-grain-size features. Further, the discrepancies between ATM and OLCI-derived grain sizes in the fine-grained regime (Fig. 11) should have relatively little impact on the accuracy of a OLCI-based

Deleted: grain size

Deleted:

Deleted: (Cooper et al., 2018)

Deleted: grain size

Deleted: (Smith et al., 2018)

Deleted: grain size

Deleted: grain size

Deleted: (Medley et al., 2022)

Deleted: grain size

Deleted: it

Deleted: grain size

Deleted: grain size

Deleted: f

Deleted: ure

Deleted: ¶

Deleted: the satellite

Commented [BS121]: R2— L532: Again, which satellite measurements?

Deleted: satellite

Deleted: in

Deleted: f

Deleted: grain size

Deleted: satellite

Deleted: grain size

Deleted: grain size

Deleted: satellit

Deleted: e

Deleted: grain-size

Deleted: the

Deleted: grain-size

Deleted: satellite

Deleted: grain size

Deleted: small-grain-sizegrain size

Deleted: f

Deleted: ure

Deleted: satellite

prediction of biases in ICESat-2 data, because whatever their disagreements about grain sizes, the two datasets agree that the bias correction should be small. We hypothesize that the peak in the ATM-bias-vs-OLCI-grain size plot around 20 μm in Fig. 11b reflects undetected clouds in the OLCI data set; for these measurements, the ATM bias can have a large range of values, while the OLCI reports a grain size appropriate for polar clouds, resulting in an apparent positive shift in the ATM biases. Errors such as these might be ameliorated in part by combining reflectance-based grain-size estimates with a model of firn evolution, which might help identify unlikely values of grain size, but this kind of analysis is beyond the scope of this study.

Our experiments with a correction for ICESat-2 biases based on the OLCI-derived grain-size estimates (Fig. 13) show that for the full dataset, the mismatch between OLCI and ATM resolution and the imprecisions of the two datasets for small grain sizes result in a net overcorrection of the biases (shown in Fig 13c, where the median of the corrected range biases is less than zero) but a reduction in the spread of the corrected biases. Implementing a threshold-based simplification of the bias model that assigns a constant value to the corrections for small grain size removes this bias and further reduces the spread of the residuals. However, the optimum parameters of this threshold model are likely determined in large part by the characteristics of the input data, including the distribution of grain sizes included in the surveys and the accuracy of the OLCI grain-size estimates on the particular days during which each survey was conducted. Researchers interested in applying the same correction to a different set of satellite-based grain-size estimates would need to perform a similar analysis to calibrate the threshold values. To calibrate a new dataset of independent grain-size estimates against the ATM-based biases, researchers would need to repeat the analysis that is summarized in Fig. 13:

1. Generate grain-size estimates for each ATM data point ($r_{est,sat}$)
2. Generate bias estimates for each grain size estimate ($B_{est,sat}$)
3. For a range of threshold values, calculate the median and spread of $B_{med}(r_{ATM}) - B_{thr}(r_{sat})$ (Eq. 9)
4. Select the threshold value that gives the minimum spread for a zero median

In our case, the threshold values that gave a zero median residual included those that gave a nearly optimal spread, but this would not necessarily be the case for other datasets, which would require more careful consideration of the trade-off between bias and spread in the correction. This kind of analysis is only feasible for satellite data that have temporal overlap with the existing ATM survey.

6 Conclusions

In this study, we have demonstrated a technique for the retrieval of ice-sheet surface grain size using the shape of pulses returned by a green-light laser. We showed that the shapes of the measured waveforms agree with the results of a simplified theoretical model of how subsurface scattering should affect the shape of green laser

Deleted: The…peak in the ATM-bias-vs-satellite…LCI-g ... [52]

Commented [BS122]: L542-543: This statement seems at odds with L584-585 which states that elevation biases could be decimeter scale. Consider revising.

Commented [BS123R122]: No? We specify here that there is no problem in the small-grain size regime, while the large grain sizes to which we refer in 584-585 are in areas with small grain sizes.

Deleted: (

Deleted: ure…11figure 11, panel… B ... [53]

Deleted:)

Deleted: likely …eflects undetected clouds in the satellite …LCI data set; for these measurements, the ATM bias can have a large range of values, while the satellite …LCI reports a grain siz ... [54]

Formatted: Font: Italic

Deleted:

Deleted: grain-size estimates…rain-size estimates with a model of firn evolution, which might help identify unlikely values of grain size ... [55]

Deleted: satellite…LCI-derived grain-size estimates…rain-size estimates (Fig.figure…13) show that for the full dataset, the mismatch between OLCIsatellite…and ATM resolution and the imprecisions of the two datasets for small grain size…rain sizes result in a net overcorrection of the biases (shown in Fig.figure …3, panel …C… where the median of the corrected range biases is less than zero) but a reduction in the spread of the corrected biases. Implementing a threshold-based simplification of the bias model that assigns a constant value to the corrections for small grain size…rain size removes this bias and further reduces the spread of the residuals. However, the optimum parameters of this threshold model are likely determined in large part by the characteristics of the input data, including the distribution of grain size…rain sizes included in the surveys and the accuracy of the satellite …LCI grain-size estimates…rain-size estimates on the particular days during which each survey was conducted. Researchers interested in applying the same correction to a different set of satellit…atellitee…based grain-size estimates…rain-size estimates would need to perform a similar analysis to calibrate the threshold values. To calibrate a new dataset of independent grain-size estimates ... [56]

Deleted: f

Deleted: ure

Deleted: grain-size estimates

Deleted: grain-size

Deleted: equation …q. 97 ... [57]

Deleted: 5… Conclusions ... [58]

Formatted: Heading 2

Deleted: grain size

pulses, and experiments with synthetic data suggest that matching waveforms with the model results should allow accurate estimates of grain size over a wide range of conditions. We showed that measurements are consistent between two independent versions of the same instrument flown on the same aircraft at the same time with different look angles, showing that the grain size recovery is repeatable, and is not strongly sensitive to the geometry of the measurements, except at small grain sizes for which the larger incidence angles associated with the wide-swath scanner begin to degrade the sensitivity of the system. Comparisons with reflectance-based estimates of grain size show agreement between the trends in the data, but not especially close point-for-point agreement between the ATM measurements and the reflectance-based measurements. However, comparisons between different reflectance-based measurements also do not show point-for-point agreement, and we are unsure whether we should claim to have validated the novel ATM-based measurements with the better-established reflectance-based techniques or whether we should claim that our ATM-based measurements provide relatively precise ground truth for the reflectance-based measurements.

Returning to the original goal of this study, which was to predict biases in ICESat-2 data, we feel that the close agreement between ATM waveforms and the shapes predicted by our model validates our use of the model to predict ICESat-2 biases due to subsurface scattering. The widespread large grain sizes we estimate in the low-elevation parts of Greenland suggest that there are large areas of the ice sheet for which we can expect decimeter-scale biases in ICESat-2 data. To date, our efforts to identify subsurface scattering bias in ICESat-2 data have been stymied by the need to collect data from tens or hundreds of pulses to resolve the shape of the return waveform, which is difficult over the rough surfaces typical of low-elevation Greenland in the summer. This suggests to us that routine correction of ICESat-2 data based on ICESat-2 return-pulse characteristics will not be feasible, except perhaps for limited areas with unusually flat topography. However, the synthesis of the ATM and OLCI-based predictions of scattering delays (Figures 12b, 13) suggests that a correction based on satellite-derived estimates of grain size is feasible for the large grain sizes where biases are largest, and that an empirical adjustment of the relation between grain-size estimates and predicted biases can be used to find a correction that yields an unbiased estimate with smaller variance than either the raw predicted biases or the unmodified correction model. Improvements in satellite-derived and model-derived estimates (Mei et al., 2021; Painter et al., 2009) of snow grain size are a potential way to improve the precision of a correction of this kind. One avenue for improvement might be to derive grain-size estimates from satellites with resolution finer than the half-kilometer OCLI data used here. A similar correction using LANDSAT and/or Sentinel-2 data could provide data at 30-meter resolution, although with coarser time resolution and with a less optimal selection of spectral bands. Another possible data source for corrections of this type would be grain size predictions driven by a grain size-evolution model driven by meteorological data or model output, which would have the advantage over purely satellite-driven grain-size estimates of providing

Deleted: grain size

Deleted: and slightly different

Deleted: ,

Deleted: grain-size

Commented [BS124]: R1 - 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

Commented [BS125R124]: We now specify that the two versions of the same instrument have different look angles, and qualify that the sensitivity of the wide-swath system is degraded at larger look angles.

Deleted: grain size

Deleted: based on ATM measurements

Deleted: grain size

Deleted: that clearly show the effects of subsurface scattering

Deleted: satellite

Deleted: f

Deleted: B

Deleted: grain size

Deleted: grain size

Deleted: grain-size estimates

Deleted: Improvements in satellite-derived and model-derived estimates

Deleted:

Deleted: (Mei et al., 2021; Painter et al., 2009)(Mei et al., 2021; Painter et al., 2009)

Deleted: of grain size are a potential way to improve the precision of a correction of this kind; the satellite images on which the corrections we used are based are available at half-kilometer resolution, which could help the per-point accuracy of the correction; a similar correction using LANDSAT and/or Sentinel-2 data could provide data at 30-meter resolution, although with less frequent collection of imagery and with a less optimal selection of spectral bands.

Deleted: grain-size

Deleted: grain-size

Deleted: grain-size estimates

estimates that would not be limited by the availability of cloud-free observations. Any such comparison would require careful consideration of the relationship between physical grain size (calculated in the grain size model) and the effective grain sizes considered in our scattering model, which might best be handled by calibrating model output overlapping the Greenland ATM surveys against ATM data.

Deleted: grain size

Deleted: grain-size

Deleted: grain size

Deleted: A final possibility would be to apply algorithms such as those tested here to ICESat-2 photon-return-time distributions. At present, attempts at such retrievals over coastal Greenland have been stymied by the need to aggregate photons from large numbers of ICESat-2 pulses over complex topography,

Formatted: English (US)

Data availability:

ATM waveform data are available from the National Snow and Ice Data Center (Studinger, 2018a, b). Ground calibration data used to derive the ATM instrument response is available at: https://zenodo.org/record/7225937. OLCI-based grain-size estimates are available through GEUS dataverse (Vandecrux et al., 2022a). AVIRIS-NG grain-size estimates are available by FTP from https://popo.jpl.nasa.gov/avng/y19/, and ATM-based grain-size estimates are available from the National Snow and Ice Data Center (NSIDC): https://doi.org/10.5067/1207YUVC7KOO.

Deleted: (Studinger, 2018a, b)

Deleted: Satellite

Deleted: grain-size estimates

Deleted: (Vandecrux et al., 2022a)

Deleted: grain-size estimates

Deleted: grain-size estimates

Deleted: are in the process of submission to the

Deleted: , and will be accessible by the time this manuscript is published

Competing interests:

At least one of the (co-)authors is a member of the editorial board of The Cryosphere.

References

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Harding, D., Kwok, R., Lefsky, M., Markus, T., Marshak, A., Neumann, T., Palm, S., Schutz, B., Smith, B., Spinhirne, J., and Webb, C.: The ICESat-2 Laser Altimetry Mission, Proc. IEEE, 98, 735–751, https://doi.org/10.1109/jproc.2009.2034765, 2010.

Allgaier, M. and Smith, B. J.: Diffuse optics for glaciology, Opt. Express, 29, 18845, https://doi.org/10.1364/oe.425630, 2021.

Allgaier, M., Cooper, M. G., Carlson, A. E., Cooley, S. W., Ryan, J. C., and Smith, B. J.: Direct measurement of optical properties of glacier ice using a photon-counting diffuse LiDAR, J. Glaciol., 68, 1210–1220, https://doi.org/10.1017/jog.2022.34, 2022.

Brunt, K. M., Smith, B. E., Sutterley, T. C., Kurtz, N. T., and Neumann, T. A.: Comparisons of Satellite and Airborne Altimetry With Ground-Based Data From the

Interior of the Antarctic Ice Sheet, Geophys. Res. Lett., 48, https://doi.org/10.1029/2020gl090572, 2021.

Cooper, M. G., Smith, L. C., Rennermalm, A. K., Miège, C., Pitcher, L. H., Ryan, J. C., Yang, K., and Cooley, S. W.: Meltwater storage in low-density near-surface bare ice in the Greenland ice sheet ablation zone, Cryosphere, 12, 955–970, https://doi.org/10.5194/tc-12-955-2018, 2018.

Fair, Z., Flanner, M., Neumann, T., Vuyovich, C., Smith, B., and Schneider, A.: Quantifying volumetric scattering bias in ICESat-2 and Operation IceBridge altimetry over snow-covered surfaces, Earth and Space Science, 2023.

Fausto, R. S., Box, J. E., Vandecrux, B., As, D. van, Steffen, K., MacFerrin, M. J., Machguth, H., Colgan, W., Koenig, L. S., McGrath, D., Charalampidis, C., and Braithwaite, R. J.: A Snow Density Dataset for Improving Surface Boundary Conditions in Greenland Ice Sheet Firn Modeling, Front. Earth Sci., 6, 51, https://doi.org/10.3389/feart.2018.00051, 2018.

Flock, S. T., Patterson, M. S., Wilson, B. C., and Wyman, D. R.: Monte Carlo modeling of light propagation in highly scattering tissues. I. Model predictions and comparison with diffusion theory, IEEE Trans. Biomed. Eng., 36, 1162–1168, https://doi.org/10.1109/tbme.1989.1173624, 1989.

Gallet, J. C., Domine, F., Zender, C. S., and Picard, G.: Measurement of the specific surface area of snow using infrared reflectance in an integrating sphere at 1310 and 1550 nm, Cryosphere, 3, 167–182, https://doi.org/10.5194/tc-3-167-2009, 2009.

Gardner, A. and Sharp, M.: A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization, J Geophys Res Earth Surf, 115, https://doi.org/10.1029/2009jf001444, 2010.

Harding, D., Dabney, P., Valett, S., Yu, A., Vasilyev, A., and Kelly, A.: AIRBORNE POLARIMETRIC, TWO-COLOR LASER ALTIMETER MEASUREMENTS OF LAKE ICE COVER: A PATHFINDER FOR NASA'S ICESAT-2 SPACEFLIGHT MISSION, 2011 IEEE Int. Geosci. Remote Sens. Symp., 1, 3598–3601, https://doi.org/10.1109/igarss.2011.6050002, 2011.

Hofton, M. A., Blair, J. B., Luthcke, S. B., and Rabine, D. L.: Assessing the performance of 20–25 m footprint waveform lidar data collected in ICESat data corridors in Greenland, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008gl035774, 2008.

Hu, Y., Lu, X., Zeng, X., Stamnes, S. A., Neuman, T. A., Kurtz, N. T., Zhai, P., Gao, M., Sun, W., Xu, K., Liu, Z., Omar, A. H., Baize, R. R., Rogers, L. J., Mitchell, B. O., Stamnes, K., Huang, Y., Chen, N., Weimer, C., Lee, J., and Fair, Z.: Deriving Snow Depth From ICESat-2 Lidar Multiple Scattering Measurements, Front. Remote Sens., 3, 855159, https://doi.org/10.3389/frsen.2022.855159, 2022.

Kokhanovsky, A., Lamare, M., Danne, O., Brockmann, C., Dumont, M., Picard, G., Arnaud, L., Favier, V., Jourdain, B., Meur, E. L., Mauro, B. D., Aoki, T., Niwano, M., Rozanov, V., Korkin, S., Kipfstuhl, S., Freitag, J., Hoerhold, M., Zuhr, A., Vladimirova, D., Faber, A. K., Steen-Larsen, H. C., Wahl, S., Andersen, J. K., Vandecrux, B., As, D. van, Mankoff, K. D., Kern, M., Zege, E., and Box, J. E.: Retrieval of Snow Properties from the Sentinel-3 Ocean and Land Colour Instrument, Remote Sens-basel, 11, 2280, https://doi.org/10.3390/rs11192280, 2019.

Krabill, W. B., Abdalati, W., Frederick, E. B., Manizade, S. S., Martin, C. F., Sonntag, J. G., Swift, R. N., Thomas, R. H., and Yungel, J. G.: Aircraft laser altimetry measurement of elevation changes of the greenland ice sheet: technique and accuracy assessment, J. Geodyn., 34, 357–376, https://doi.org/10.1016/s0264-3707(02)00040-6, 2002.

Lu, X., Hu, Y., Zeng, X., Stamnes, S. A., Neuman, T. A., Kurtz, N. T., Yang, Y., Zhai, P.-W., Gao, M., Sun, W., Xu, K., Liu, Z., Omar, A. H., Baize, R. R., Rogers, L. J., Mitchell, B. O., Stamnes, K., Huang, Y., Chen, N., Weimer, C., Lee, J., and Fair, Z.: Deriving Snow Depth From ICESat-2 Lidar Multiple Scattering Measurements: Uncertainty Analyses, Front. Remote Sens., 3, 891481, https://doi.org/10.3389/frsen.2022.891481, 2022.

MacGregor, J. A., Boisvert, L. N., Medley, B., Petty, A. A., Harbeck, J. P., Bell, R. E., Blair, J. B., Blanchard-Wrigglesworth, E., Buckley, E. M., Christoffersen, M. S., Cochran, J. R., Csathó, B. M., Marco, E. L., Dominguez, R. T., Fahnestock, M. A., Farrell, S. L., Gogineni, S. P., Greenbaum, J. S., Hansen, C. M., Hofton, M. A., Holt, J. W., Jezek, K. C., Koenig, L. S., Kurtz, N. T., Kwok, R., Larsen, C. F., Leuschen, C. J., Locke, C. D., Manizade, S. S., Martin, S., Neumann, T. A., Nowicki, S. M. J., Paden, J. D., Richter-Menge, J. A., Rignot, E. J., Rodríguez-Morales, F., Siegfried, M. R., Smith, B. E., Sonntag, J. G., Studinger, M., Tinto, K. J., Truffer, M., Wagner, T. P., Woods, J. E., Young, D. A., and Yungel, J. K.: The Scientific Legacy of NASA's Operation IceBridge, Rev. Geophys., 59, https://doi.org/10.1029/2020rg000712, 2021.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S., Shum, C. K., Schutz, B. E., Smith, B., Yang, Y. K., and Zwally, J.: The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation, Remote Sens. Environ., 190, 260–273, https://doi.org/10.1016/j.rse.2016.12.029, 2017.

Medley, B., Neumann, T. A., Zwally, H. J., Smith, B. E., and Stevens, C. M.: Simulations of firn processes over the Greenland and Antarctic ice sheets: 1980-2021, Cryosphere, 16, 3971–4011, https://doi.org/10.5194/tc-16-3971-2022, 2022.

Mei, L., Rozanov, V., Pohl, C., Vountas, M., and Burrows, J. P.: The retrieval of snow properties from SLSTR Sentinel-3 – Part 1: Method description and sensitivity study, Cryosphere, 15, 2757–2780, https://doi.org/10.5194/tc-15-2757-2021, 2021.

Nolin, A. W. and Dozier, J.: A hyperspectral method for remotely sensing the grain size of snow, Remote Sens Environ, 74, 207–216, https://doi.org/10.1016/s0034-4257(00)00111-5, 2000.

Nolin, A. W. and Payne, M. C.: Classification of glacier zones in western Greenland using albedo and surface roughness from the Multi-angle Imaging SpectroRadiometer (MISR), Remote Sens. Environ., 107, 264–275, https://doi.org/10.1016/j.rse.2006.11.004, 2007.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J.: Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, Remote Sens. Environ., 113, 868–879, https://doi.org/10.1016/j.rse.2009.01.001, 2009.

Petty, A. A., Keeney, N., Cabaj, A., Kushner, P., and Bagnardi, M.: Winter Arctic sea ice thickness from ICESat-2: upgrades to freeboard and snow loading estimates and an assessment of the first three winters of data collection, Cryosphere, 17, 127–156, https://doi.org/10.5194/tc-17-127-2023, 2022.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical recipes : the art of scientific computing, 2007.

Smith, B., Fricker, H. A., Holschuh, N., Gardner, A. S., Adusumilli, S., Brunt, K. M., Csatho, B., Harbeck, K., Huth, A., Neumann, T., Nilsson, J., and Siegfried, M. R.: Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter, Remote Sensing of Environment, 233, 111352, https://doi.org/10.1016/j.rse.2019.111352, 2019a.

Smith, B., Fricker, H. A., Holschuh, N., Gardner, A. S., Adusumilli, S., Brunt, K. M., Csatho, B., Harbeck, K., Huth, A., Neumann, T., Nilsson, J., and Siegfried, M. R.: Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter, Remote Sens. Environ., 233, 111352, https://doi.org/10.1016/j.rse.2019.111352, 2019b.

Smith, B. E., Gardner, A., Schneider, A., and Flanner, M.: Modeling biases in laser-altimetry measurements caused by scattering of green light in snow, Remote Sens Environ, 215, 398–410, https://doi.org/10.1016/j.rse.2018.06.012, 2018.

Smith, B. E., Fricker, H. A., Gardner, A. S., Medley, B., Nilsson, J., Paolo, F. S., Holschuh, N., Adusumilli, S., Brunt, K., Csatho, B., Harbeck, K., Markus, T., Neumann, T., Siegdried, M. R., and Zwally, H. J.: Pervasive ice sheet mass loss driven by competing,ocean and,atmosphere processes, Science, 368, 1239–1242, https://doi.org/10.1126/science.aaz5845, 2020.

Studinger: IceBridge ATM L1B Elevation and Return Strength with Waveforms, Version 1, NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/ezq5u3r3xwbs, 2018a.

Studinger, M.: IceBridge Narrow Swath ATM L1B Elevation and Return Strength with Waveforms, Version 1 [Data Set]., Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center., https://doi.org/10.5067/v25x7lhdpmzy, 2018b.

Studinger, M., Manizade, S. S., Linkswiler, M. A., and Yungel, J. K.: High-resolution imaging of supraglacial hydrological features on the Greenland Ice Sheet with NASA's Airborne Topographic Mapper (ATM) instrument suite, Cryosphere, 16, 3649–3668, https://doi.org/10.5194/tc-16-3649-2022, 2022a.

Studinger, M., Linkswiler, M. A., Manizade, S. S., and Yungel, J. K.: NASA's Airborne Topographic Mapper (ATM) ground calibration data for waveform data products, 2022b.

Studinger, M., Smith, B. E., Kurtz, N., Petty, A., Sutterley, T., and Tilling, R.: Estimating differential penetration of green (532 nm) laser light over sea ice with NASA's Airborne Topographic Mapper: observations and models, Cryosphere Discuss., 2023, 1–35, https://doi.org/10.5194/tc-2023-126, 2023.

Sutterley, T. C., Markus, T., Neumann, T. A., Broeke, M. van den, Wessem, J. M. van, and Ligtenberg, S. R. M.: Antarctic ice shelf thickness change from multimission lidar mapping, Cryosphere, 13, 1801–1817, https://doi.org/10.5194/tc-13-1801-2019, 2019.

Thompson, D. R., Boardman, J. W., Eastwood, M. L., Green, R. O., Haag, J. M., Mouroulis, P., and Gorp, B. V.: Imaging spectrometer stray spectral response: In-flight characterization, correction, and validation, Remote Sens. Environ., 204, 850–860, https://doi.org/10.1016/j.rse.2017.09.015, 2018.

Vandecrux, B., Box, J., Mankoff, K., and Wehrlé, A.: Snow broadband albedo, specific surface area and optical grain diameter from Sentinel-3's OLCI, daily 1 km mosaics, Greenland, GEUS dataverse, https://doi.org/10.22008/fk2/oiajvo, 2022a.

Vandecrux, B., Box, J. E., Wehrle, A., Kokhanovsky, A. A., Picard, G., Niwano, M., Hoerhold, M., Faber, A. K., and Steen-Larsen, H. C.: The Determination of the Snow Optical Grain Diameter and Snowmelt Area on the Greenland Ice Sheet Using Spaceborne Optical Observations, Remote Sens-basel, 14, 932, https://doi.org/10.3390/rs14040932, 2022b.

Warren, S. G.: Optical properties of snow, Rev. Geophys., 20, 67–89, https://doi.org/10.1029/rg020i001p00067, 1982.

Warren, S. G. and Brandt, R. E.: Optical constants of ice from the ultraviolet to the microwave: A revised compilation, J Geophys Res Atmospheres, 113, https://doi.org/10.1029/2007jd009744, 2008.

Wiscombe, W. J. and Warren, S. G.: A Model for the Spectral Albedo of Snow. I: Pure Snow, J. Atmos. Sci., 37, 2712–2733, https://doi.org/10.1175/1520-0469(1980)037<;2712:amftsa>2.0.co;2, 1980.

Yi, D., Zwally, H. J., and Sun, X.: ICESat measurement of Greenland ice sheet surface slope and roughness, Ann. Glaciol., 42, 83–89, https://doi.org/10.3189/172756405781812691, 2005.

Yu, A. W., Harding, D. J., and Dabney, P. W.: Laser transmitter design and performance for the slope imaging multi-polarization photon-counting lidar (SIMPL) instrument, Solid State Lasers XXV: Technol. Devices, 97260J-97260J–8, https://doi.org/10.1117/12.2213005, 2016.

Deleted:

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Harding, D., Kwok, R., Lefsky, M., Markus, T., Marshak, A., Neumann, T., Palm, S., Schutz, B., Smith, B., Spinhirne, J., and Webb, C.: The ICESat-2 Laser Altimetry Mission, Proc. IEEE, 98, 735–751, https://doi.org/10.1109/jproc.2009.2034765, 2010.

Allgaier, M. and Smith, B. J.: Diffuse optics for glaciology, Opt. Express, 29, 18845, https://doi.org/10.1364/oe.425630, 2021.

Allgaier, M., Cooper, M. G., Carlson, A. E., Cooley, S. W., Ryan, J. C., and Smith, B. J.: Direct measurement of optical properties of glacier ice using a photon-counting diffuse LiDAR, J. Glaciol., 68, 1210–1220, https://doi.org/10.1017/jog.2022.34, 2022.

Cooper, M. G., Smith, L. C., Rennermalm, A. K., Miège, C., Pitcher, L. H., Ryan, J. C., Yang, K., and Cooley, S. W.: Meltwater storage in low-density near-surface bare ice in the Greenland ice sheet ablation zone, Cryosphere, 12, 955–970, https://doi.org/10.5194/tc-12-955-2018, 2018.

Fair, Z., Flanner, M., Neumann, T., Vuyovich, C., Smith, B., and Schneider, A.: Quantifying volumetric scattering bias in ICESat-2 and Operation IceBridge altimetry over snow-covered surfaces, Earth and Space Science, n.d.

Gallet, J. C., Domine, F., Zender, C. S., and Picard, G.: Measurement of the specific surface area of snow using infrared reflectance in an integrating sphere at 1310 and 1550 nm, Cryosphere, 3, 167–182, https://doi.org/10.5194/tc-3-167-2009, 2009.

Gardner, A. and Sharp, M.: A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization, J Geophys Res Earth Surf, 115, https://doi.org/10.1029/2009jf001444, 2010.

Harding, D., Dabney, P., Valett, S., Yu, A., Vasilyev, A., and Kelly, A.: AIRBORNE POLARIMETRIC, TWO-COLOR LASER ALTIMETER MEASUREMENTS OF LAKE ICE COVER: A PATHFINDER FOR NASA'S ICESAT-2 SPACEFLIGHT MISSION, 2011 IEEE Int. Geosci. Remote Sens. Symp., 1, 3598–3601, https://doi.org/10.1109/igarss.2011.6050002, 2011.

Hofton, M. A., Blair, J. B., Luthcke, S. B., and Rabine, D. L.: Assessing the performance of 20–25 m footprint waveform lidar data collected in ICESat data corridors in Greenland, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008gl035774, 2008.

Hu, Y., Lu, X., Zeng, X., Stamnes, S. A., Neuman, T. A., Kurtz, N. T., Zhai, P., Gao, M., Sun, W., Xu, K., Liu, Z., Omar, A. H., Baize, R. R., Rogers, L. J., Mitchell, B. O., Stamnes, K., Huang, Y., Chen, N., Weimer, C., Lee, J., and Fair, Z.: Deriving Snow Depth From ICESat-2 Lidar Multiple Scattering Measurements, Front. Remote Sens., 3, 855159, https://doi.org/10.3389/frsen.2022.855159, 2022.

Kokhanovsky, A., Lamare, M., Danne, O., Brockmann, C., Dumont, M., Picard, G., Arnaud, L., Favier, V., Jourdain, B., Meur, E. L., Mauro, B. D., Aoki, T., Niwano, M., Rozanov, V., Korkin, S., Kipfstuhl, S., Freitag, J., Hoerhold, M., Zuhr, A., Vladimirova, D., Faber, A. K., Steen-Larsen, H. C., Wahl, S., Andersen, J. K., Vandecrux, B., As, D. van, Mankoff, K. D., Kern, M., Zege, E., and Box, J. E.: Retrieval of Snow Properties from the Sentinel-3 Ocean and Land Colour Instrument, Remote Sens-basel, 11, 2280, https://doi.org/10.3390/rs11192280, 2019.

Lu, X., Hu, Y., Zeng, X., Stamnes, S. A., Neuman, T. A., Kurtz, N. T., Yang, Y., Zhai, P.-W., Gao, M., Sun, W., Xu, K., Liu, Z., Omar, A. H., Baize, R. R., Rogers, L. J., Mitchell, B. O., Stamnes, K., Huang, Y., Chen, N., Weimer, C., Lee, J., and Fair, Z.: Deriving Snow Depth From ICESat-2 Lidar Multiple Scattering Measurements: Uncertainty Analyses, Front. Remote Sens., 3, 891481, https://doi.org/10.3389/frsen.2022.891481, 2022.

... [59]

Page 2: [1] Deleted **Ben Smith** **4/22/24 10:55:00 AM**

Page 2: [2] Commented [BS8R6] **Ben Smith** **2/14/24 5:03:00 PM**

Wordy sentence: Now split into two pieces.

L36: Added ice sheets.

Page 2: [3] Commented [BS9] **Ben Smith** **1/2/24 2:47:00 PM**

R2 — L32-25: Long, wordy sentence, consider splitting.

Page 2: [4] Commented [BS10R9] **Ben Smith** **1/2/24 2:48:00 PM**

Also:

R2: L36: Just glaciers? Or ice sheets as well? It seems like these two terms are being used interchangeably which is at odds with the first couple of sentences.

Page 2: [5] Commented [BS11R9] **Ben Smith** **2/14/24 5:07:00 PM**

Added "ice sheets"

Page 2: [6] Commented [BS12] **Ben Smith** **1/2/24 2:49:00 PM**

R2: L41-49: Even though this is a well-known phenomenon, it might be useful to add some references here which describe this in more detail.

Page 2: [7] Commented [BS13R12] **Ben Smith** **2/14/24 5:08:00 PM**

Added references to two previous studies

Page 2: [8] Deleted **Ben Smith** **4/22/24 4:06:00 PM**

Page 3: [9] Deleted **Ben Smith** **2/14/24 5:10:00 PM**

Page 3: [9] Deleted **Ben Smith** **2/14/24 5:10:00 PM**

Page 3: [10] Commented [BS17] **Ben Smith** **12/21/23 4:38:00 PM**

R1- 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.

Page 3: [11] Commented [BS18R17]	Ben Smith	5/9/24 3:52:00 PM

We now say: (i.e. the measured waveform in an analog lidar, or the distribution of photon timing in a photon-counting lidar)

Page 3: [12] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [12] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [13] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [13] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [13] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [13] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [14] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [14] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [14] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [15] Commented [BS19]	Ben Smith	1/2/24 2:49:00 PM

L56-59: I would need access to these manuscripts to judge overlap and novelty of this paper.

Page 3: [16] Deleted	Ben Smith	4/30/24 5:02:00 PM

Page 3: [16] Deleted	Ben Smith	4/30/24 5:02:00 PM

Page 3: [16] Deleted	Ben Smith	4/30/24 5:02:00 PM

Page 3: [17] Deleted	Ben Smith	4/30/24 5:03:00 PM

Page 3: [17] Deleted	Ben Smith	4/30/24 5:03:00 PM

Page 3: [17] Deleted	Ben Smith	4/30/24 5:03:00 PM

Page 3: [18] Deleted	Ben Smith	2/15/24 4:25:00 PM

Page 3: [18] Deleted	Ben Smith	2/15/24 4:25:00 PM

Page 3: [18] Deleted	Ben Smith	2/15/24 4:25:00 PM

Page 3: [18] Deleted	Ben Smith	2/15/24 4:25:00 PM

Page 3: [18] Deleted	Ben Smith	2/15/24 4:25:00 PM

Page 3: [18] Deleted	Ben Smith	2/15/24 4:25:00 PM

Page 3: [19] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [19] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [19] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [19] Deleted	Ben Smith	1/10/24 4:41:00 PM

Page 3: [19] Deleted **Ben Smith** **1/10/24 4:41:00 PM**

Page 3: [19] Deleted **Ben Smith** **1/10/24 4:41:00 PM**

Page 3: [19] Deleted **Ben Smith** **1/10/24 4:41:00 PM**

Page 3: [19] Deleted **Ben Smith** **1/10/24 4:41:00 PM**

Page 4: [20] Deleted **Ben Smith** **5/1/24 2:45:00 PM**

Page 4: [21] Commented [BS21] **Ben Smith** **12/21/23 4:40:00 PM**

R1 - 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.

Page 4: [22] Commented [BS22R21] **Ben Smith** **2/20/24 3:35:00 PM**

We now point the reader to section 3 for measurement locations.

Page 4: [23] Commented [BS23] **Ben Smith** **1/2/24 2:50:00 PM**

R2- L84: Define acronym on first use of term "ATM" (L75) rather than here.

Page 6: [24] Commented [BS37] **Ben Smith** **12/21/23 4:43:00 PM**

R1 - 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.

Page 6: [25] Commented [BS38] **Ben Smith** **1/2/24 2:52:00 PM**

R2- L114: "Verify" seems a bit strong. Validate or evaluate might be better.

Page 6: [26] Deleted **Ben Smith** **4/30/24 5:05:00 PM**

Page 6: [27] Deleted **Ben Smith** **4/18/24 3:26:00 PM**

Page 6: [28] Deleted **Ben Smith** **1/10/24 4:41:00 PM**

Page 6: [29] Deleted **Ben Smith** **4/18/24 3:29:00 PM**

Page 7: [30] Deleted **Ben Smith** **1/22/24 4:48:00 PM**

Page 10: [31] Deleted **Ben Smith** **1/30/24 4:05:00 PM**

Page 14: [32] Deleted **Ben Smith** **7/2/24 2:56:00 PM**

Page 14: [33] Commented [BS66] **Ben Smith** **12/21/23 4:56:00 PM**

R1 - Lines 270-278: Much of this paragraph is dedicated to describing the effects ATM amplitude had (past-tense) on the uncertainty in the estimated grainsize. The issue I find however, is that the ATM data results have not been covered yet. What do the authors expect the reader to takeaway from this paragraph when they have not seen the grainsize 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 grainsize estimates produced from the analysis of the ATM data that they mention here?

Page 14: [34] Commented [BS67] **Ben Smith** **1/2/24 2:56:00 PM**

R2: L280: Is the satellite not named "ICES**at**-2"?

Page 20: [35] Commented [BS82] **Ben Smith** **12/21/23 5:08:00 PM**

R1 - 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.

Page 20: [36] Commented [BS83] **Ben Smith** **1/2/24 2:56:00 PM**

L361: It would be useful to name the two ATM sensors since this is the first sentence of the paragraph

Page 20: [37] Commented [BS85] **Ben Smith** **1/2/24 1:19:00 PM**

R1 - 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"?

Page 23: [38] Commented [BS97R94] **Ben Smith** **4/22/24 10:29:00 AM**

"Comes about" replaced by "is"

Page 23: [39] Deleted **Ben Smith** **2/8/24 3:55:00 PM**

Page 27: [40] Commented [BS106] **Ben Smith** **1/2/24 3:06:00 PM**

R2 — L485-489: I think the first half of this paragraph should be removed (or placed later in the discussion) because, as it is written, it seems like the main takeaway is the consistency of the different snow grain size estimates. However, there are substantial biases between the estimates (Figs. 9-11) which the authors are up front about later in the discussion and should be the focus.

Page 27: [41] Deleted **Ben Smith** **2/23/24 4:40:00 PM**

Page 27: [42] Deleted **Ben Smith** **7/2/24 4:16:00 PM**

Page 27: [43] Commented [BS111] **Ben Smith** **1/2/24 1:28:00 PM**

R1 - 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.

Page 27: [44] Commented [BS112R111] **Ben Smith** **3/6/24 11:21:00 AM**

We realize that this is not a standard metric, and now describe it in our 'methods' section.

Page 29: [45] Deleted **Ben Smith** **4/24/24 1:44:00 PM**

Page 29: [45] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [45] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [46] Deleted	**Ben Smith**	**3/6/24 11:26:00 AM**

Page 29: [46] Deleted	**Ben Smith**	**3/6/24 11:26:00 AM**

Page 29: [46] Deleted	**Ben Smith**	**3/6/24 11:26:00 AM**

Page 29: [46] Deleted	**Ben Smith**	**3/6/24 11:26:00 AM**

Page 29: [46] Deleted	**Ben Smith**	**3/6/24 11:26:00 AM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	**Ben Smith**	**4/24/24 1:44:00 PM**

Page 29: [47] Deleted	Ben Smith	4/24/24 1:44:00 PM

Page 29: [47] Deleted	Ben Smith	4/24/24 1:44:00 PM

Page 29: [47] Deleted	Ben Smith	4/24/24 1:44:00 PM

Page 29: [47] Deleted	Ben Smith	4/24/24 1:44:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [48] Deleted	Ben Smith	1/10/24 4:42:00 PM

Page 29: [49] Commented [BS119]	Ben Smith	1/2/24 1:32:00 PM

R1 - 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?

Page 29: [50] Commented [BS120R119] **Ben Smith** **3/6/24 3:10:00 PM**

We have weakened our first statement about the agreement between the datasets (now "broad agreement"). The rest of this seems like it says what we want to say— the agreement between the datasets is imperfect, unlike the point-for-point agreement that we might have hoped for. The rest of the paragraph explains how this disagreement came about. I hope it is less jarring without the repetition of the word "consistent."

Page 29: [51] Deleted **Ben Smith** **4/24/24 1:44:00 PM**

Page 29: [51] Deleted **Ben Smith** **4/24/24 1:44:00 PM**

Page 29: [51] Deleted **Ben Smith** **4/24/24 1:44:00 PM**

Page 29: [51] Deleted **Ben Smith** **4/24/24 1:44:00 PM**

Page 29: [51] Deleted **Ben Smith** **4/24/24 1:44:00 PM**

Page 31: [52] Deleted **Ben Smith** **6/26/24 5:29:00 PM**

Page 31: [52] Deleted **Ben Smith** **6/26/24 5:29:00 PM**

Page 31: [52] Deleted **Ben Smith** **6/26/24 5:29:00 PM**

Page 31: [53] Deleted **Ben Smith** **7/16/24 4:18:00 PM**

Page 31: [53] Deleted **Ben Smith** **7/16/24 4:18:00 PM**

Page 31: [53] Deleted **Ben Smith** **7/16/24 4:18:00 PM**

Page 31: [54] Deleted **Ben Smith** **6/26/24 5:29:00 PM**

Page 31: [54] Deleted **Ben Smith** **6/26/24 5:29:00 PM**

Page 31: [54] Deleted	Ben Smith	6/26/24 5:29:00 PM

Page 31: [54] Deleted	Ben Smith	6/26/24 5:29:00 PM

Page 31: [55] Deleted	Ben Smith	4/24/24 1:46:00 PM

Page 31: [55] Deleted	Ben Smith	4/24/24 1:46:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [56] Deleted	Ben Smith	3/6/24 4:29:00 PM

Page 31: [57] Deleted	Ben Smith	1/3/24 2:51:00 PM

1.

Page 31: [57] Deleted	Ben Smith	1/3/24 2:51:00 PM

2.

Page 31: [58] Deleted	Ben Smith	1/3/24 2:44:00 PM

Page 31: [58] Deleted	Ben Smith	1/3/24 2:44:00 PM

Page 38: [59] Deleted	Ben Smith	1/10/24 4:42:00 PM