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]

5   [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.

10   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 NASA's ICESat-2 (Ice Cloud, and land Elevation Satellite-2) mission. In this study, we investigate these biases using a model of subsurface scattering, altimetry

15   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 argue 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

20   subsurface scattering, and demonstrate that on the basis of large-scale averages, the corrections calculated based on the optical imagery match the biases in the data. This work demonstrates that waveform-based altimetry data has the potential to measure the optical properties of granular surfaces, and that corrections based on optical grain-size estimates have the potential to correct for subsurface-scattering biases in ICESat-2 data.

1 Introduction.

25   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 atmospheric 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

30 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 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 efficient detectors to measure the arrival time of individual return photons, increasing its overall precision and efficiency relative to that of ICESat, but it leads to potential biases in its altimetry measurements because ice absorbs green light weakly, allowing

35 photons to scatter over relatively long distances within the snow before returning to the surface and, potentially, the satellite. These biases are relevant to interpretations of ICESat-2 altimetry measurements over glaciers, because ICESat-2 was designed to make precise measurements of glacier elevation change, and time varying biases in ICESat-2 measurements over glaciers and ice shelves can produce spurious signals that might be interpreted as ice-sheet mass changes, and spatially varying biases in measurements over sea ice might be interpreted as variability in freeboard and thus ice thickness (Harding et al., 2011;

40 Smith et al., 2018).

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, and light leaving the surface includes a distribution of photon delays,

45 which both delays the mean time of the returning photons and changes the shape of the returning pulse. The mean delay of the photons and the shape of the returning pulse depend on the scattering properties of the material, with lower densities and coarser grain sizes corresponding to weaker scattering, broader returns, and longer delay times. Light absorption within the scattering medium can also influence time distribution of returning photons, with stronger absorption producing narrower distributions and smaller net delays.

50 The dependence of return photon timing distribution on ice optical properties has been explored in previous studies (Smith et al., 2018; Allgaier and Smith, 2021; Hu et al., 2022), and 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 cm (Allgaier et al., 2022). Although other studies studies have noted the potential for remote-sensing studies to apply these theories to laser measurements, only a few studies have attempted to infer

55 snow and firn properties based on remotely sensed lidar scattering measurements (Hu et al., 2022; Lu et al., 2022; Harding et al., 2011). Two studies submitted contemporaneously with ours (Unpublished manuscripts provided as attachments) explore, respectively, the estimation of snow grain size based on apparent elevation differences between green and near-infrared altimetry measurements over northern Greenland snow surfaces, and the biases associated with subsurface scattering of green laser light in altimetry measurements over sea ice. The second of these studies demonstrates that subsurface scattering of

60 green laser light can lead to elevation biases in sea ice that make it appear that the surface of floating ice is below the surface of the surrounding water, which emphasizes the importance of subsurface scattering for both land ice and sea ice.

In this study, we investigate the scattering properties of Greenland snow and ice surfaces with the goal of developing a correction for the biases that subsurface scattering can introduce into ICESat-2 data. Using waveform measurements from an airborne laser-altimetry system over the Greenland ice sheet, we test the model of subsurface scattering developed previously (Smith et al., 2018) based on a comparison between the shapes of the returned pulses and timing distributions expected based on the model, and demonstrate that by adjusting the scattering parameters in the model to match modelled waveforms to measured waveforms we can recover an estimate of the near-surface optical grain size. We then test whether the spatial variations in parameters found in this way match variations observed based on optical reflectance measurements calculated at the same time. We then demonstrate the use of grain-size estimates derived from satellite imagery 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-resolution altimetry-based measurements can lead to mismatch in calibrations based on optical data.

2. Data

This study is based on ATM waveform data, grain-size estimates from an airborne spectrometer, and satellite-based grain-size estimates.

2.1 Altimetric waveforms from the Airborne Topographic Mapping lidar systems.

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 return-pulse shape. Because the recorded pulse for a laser altimeter corresponds to the temporal convolution of the expected 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, the effects of subsurface scattering become easier to measure for narrower transmitted pulses, higher bandwidth recording systems, flatter surfaces, and smaller beam divergence values.

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 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. Similar measurements have been made using the Land, Vegetation, and Ice Sensor (LVIS) (Hofton et al., 2008), but that sensor's longer pulse duration and infrared wavelength provide only limited sensitivity of waveform shapes 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-
95   counting detection strategy that is not compatible with the processing software used in this study.

Waveform measurements in this study come from the ATM wide-swath and narrow-swath waveform products (Studinger,
2018a, b), which provide digitized transmitted and received waveforms associated with each transmitted pulse in each airborne
campaign. The waveforms have a temporal sampling of 0.25 ns, and are quantized at 8 bits, to produce digital values between
0 and 255. During flights, a variable neutral density filter in front of the receiver is adjusted to compensate for varying
100   conditions and avoid digitizer saturation. Therefore the amplitude of the recorded pulse does not have a consistent relation
with the intensity of the received signal. 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. The instruments record both the received and
transmitted waveforms. The recorded transmitted waveform is delayed through a multimode fiber-optic cable to eliminate
105   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. Because of this, we
disregard the measured transmitted pulse shapes, and instead assume that the system IRF is consistent with the most recent
110   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 verify that the ATM-derived waveforms were consistent with the returns we would expect from known surface
115   conditions, we used data collected using AVIRIS-NG (the Airborne Visible/Infrared Imaging Spectrometer, Next-Generation),
on a Basler aircraft that followed the aircraft carrying ATM on five subsequent days in the autumn of 2019. 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). These data provide
grain-size estimates on a swath that is several times wider than the ATM swath, with pixel sizes between five and six meters,
120   depending on the height of the aircraft above the surface. The quality of the grain-size retrievals depends strongly on the solar
illumination of the ice-sheet surface during the acquisition; we found that some of the data files contained grain-size estimates
with uniform values, or contained areas of small grain sizes that did not appear consistent with expected surface characteristics.
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.

125

2.3 Grain-size estimates from satellite measurements

To demonstrate potential corrections for ICESat-2 height biases, we use a set of satellite measurements (Vandecrux et al., 2022b) derived from the Ocean and Land Colour Instrument (OLCI) 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 against in-situ datasets found that the OCLI-based estimates were systematically larger than estimates derived from the infrared (1310 nm) reflectance (Gallet et al., 2009) of surface-snow samples collected at EastGRIP in northeast Greenland, but showed the expected decreases during snowfall events, and increases during melt events (Vandecrux et al., 2022b). 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. Values interpolated from this grain-size look-up table give an estimate of the grain size, under the assumption that the grain size did not change since the previous observation, and values interpolated from the measurement-age lookup table let us edit out data points whose observations have not recently been updated. This lets us derive independent grain-size estimates coincident with the ATM and AVIRIS-NG measurements, and, because each of these datasets was collected under mainly cloud-free conditions, we expect that this assumption was correct for most the data ATM and AVIRIS data considered in this study.

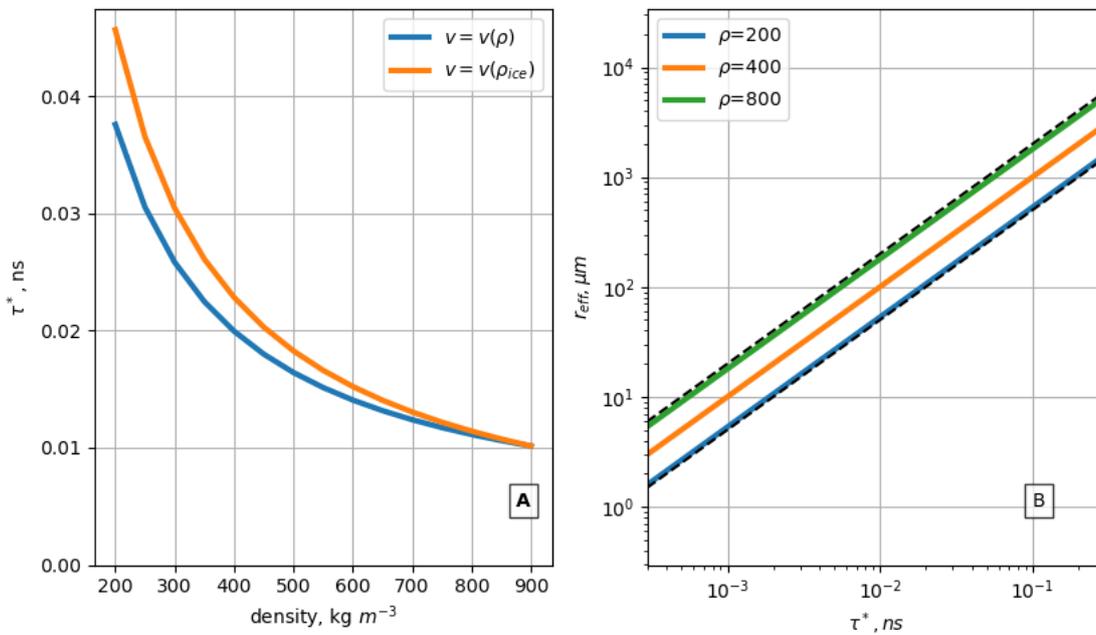
3. Methods

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 delays (Smith et al., 2018). This model is similar to that used in other studies (Allgaier and Smith, 2021), 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

160 medium) but less so for the backscatter geometry used here, because diffusion theory can produce unphysical results for very short time delays, and the horizonal displacement between the source and the detector means that these short delays are not observed. By directly modelling the time of flight for the incident beam and the first few scattering events, our Monte Carlo model avoids this problem.

165 **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 1000 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 $\rho = 400\ kg\ m^{-3}$.**

170 Returns from our model can be described as:

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

where

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

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,
175 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
180 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 $n_{air} = 1$.

185 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 eqn. 1 solely in terms of k_{abs} and r_{eff}. 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
190 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
195 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. Combining these gives:

200

$$W(t - t_{surf}) = I(t) \otimes SRF(t) \otimes \delta z \left(\frac{ct}{2} \right) \qquad 3a$$

$$W_{model}(t - t_{surf}, r_o) = I_{est}(t) \otimes SRF_m(t; r_o) \otimes G \left(\frac{ct}{2}, 0, \sigma \right) \qquad 3b$$

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. 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

205   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 $SRF_m(t; r_o)$ is calculated from Eq. 1, $I_{est}(t)$ is an estimate of the IRF, and $G(t, 0, \sigma)$ is a Gaussian function representing the combination of surface roughness and return broadening due to surface slope.

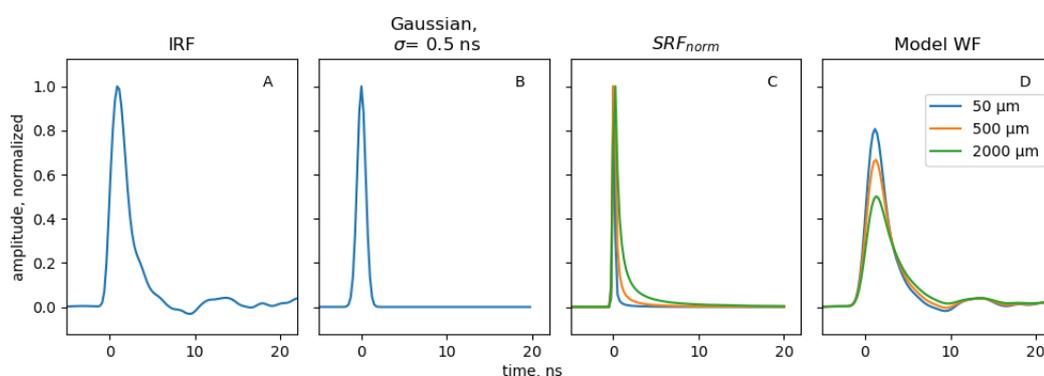



Figure 2. Components of the waveform model. The ATM IRF (A) is convolved with a Gaussian function representing surface roughness (B) and the surface response function (C) to produce the model waveform (D). Three SRFs and corresponding waveforms are shown in (C) and (D), for r_{eff}=50, 500, and 2000 μm.

Figure 2 shows the components of equation 3b, and resulting waveforms, based on the system IRF measured 9 March 2018, for a surface roughness equivalent to 0.5 ns (*i.e.* 7.5 cm), and for three grain sizes: 50, 500, and 2000 μm. The modeled

215   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 eqn. 3, but the consistency between modeled and recovered waveforms (see

220   the next section) 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 identified the first sample at which the waveform exceeded 50% of its maximum amplitude

225   and assumed that all samples more than 3 ns before this sample contained only a DC offset and noise, whose values we calculated as the mean and standard deviation (N_{est}) of the sample values in this region. We then corrected each waveform by subtracting this DC offset.

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

230

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

Here $P_m(t_i)$ is the waveform sampled at times t_i, corrected for the background rate, and W is the modelled waveform.

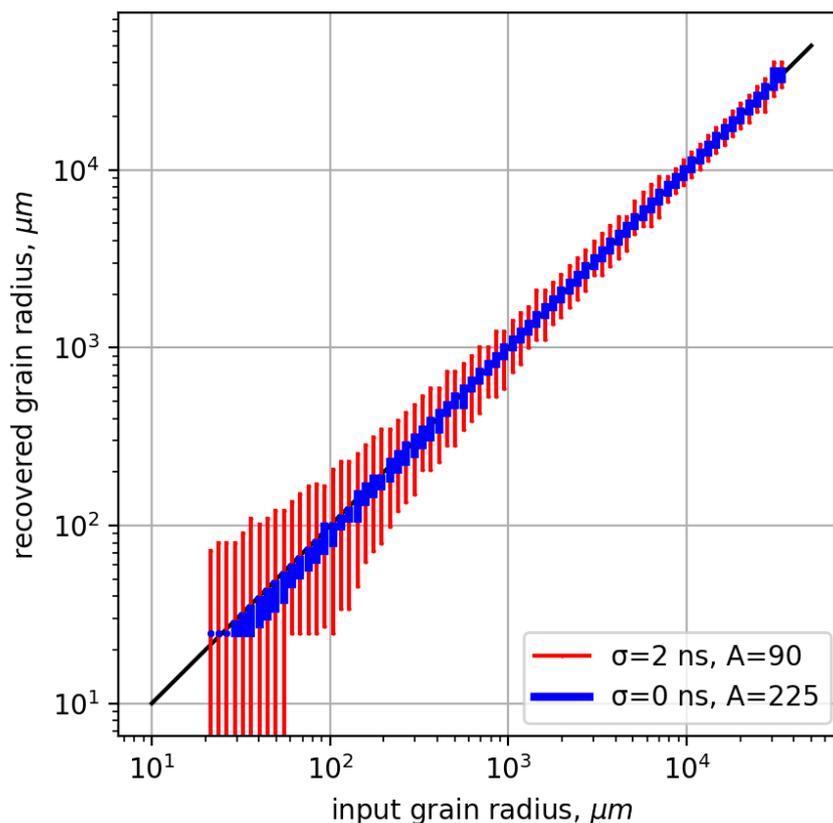



Figure 3. Fitting test data. Vertical bars show the range of recovered grain sizes for each input grain size value, for a low -amplitude, rough-surface case with A=90, σ=2 ns, and a high-amplitude, smooth-surface case, with A=255, σ=0 ns. Bars indicate the 5th and 95th percentiles of the recovered grain 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 waveform.

We find optimal values for our adjustable parameters using a three-stage golden-section search (Press et al., 2007) in σ, r_{eff},

240   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 Eqn. 3b are computationally costly, so we keep track of all

245   waveforms we had calculated, and, whenever possible, used 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-

250   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.

255   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

260   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

265   and larger roughness, the width of the recovered distribution increases with decreasing grain size, with distributions spanning around 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.

270   Our numerical experiments show that for synthetic data, the ratio between the amplitude of the returned pulse and the noise value plays a large role in the accuracy of the recovered grain size, with larger amplitudes corresponding to higher precision. For measured field data, however, there is no consistent relation between the reflectance of the surface and the amplitude of the received pulse, because the digitizer consistently produced noise values of a few counts, while the transmittance of the

ATM receiver optics was adjusted by hand during surveys to avoid saturating the detector. In areas where the surface was
275   consistently dark or had consistently large grain sizes, the receiver might have been tuned to capture the returns at high gain,
but in areas where bright snow was mixed with darker surfaces, or where the range to the surface was highly variable because
of rugged topography, the strongest returns were often captured with settings that produced large amplitudes, while the darker,
more distant surfaces had lower amplitude, and thus lower precision grain-size estimates.

3.4 Predicting biases in ICESat-2 measurements.

280   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:

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

285   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., 2019) 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, 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 (6) to calculate expected
290   biases with the windowed waveform bias as $M(\ )$. This is not practical, however, because most ATM waveforms include
digitizer output that is less than zero (see figure 2), and a waveform-based median is not defined for a waveform that is not
uniformly non-negative. Instead, 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. We then
use Eq. 6 to predict the bias in the ATL06 measurements as a function of grain size. Figure 4 shows the expected bias in h_{li} as
295   a function of grain size using the ICESat-2 IRF, for the ATL06 windowed median, for a windowed mean (equivalent to the
h_mean parameter on ATL06), and the metric used to track ATM waveforms, using the ATM IRF: the centroid of the part of
the waveform that has digitized values greater than 15% of the waveform's maximum. 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 IS2 bias is around twice as large as the median-based IS2 bias, and the ATM bias falls in between the
300   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.

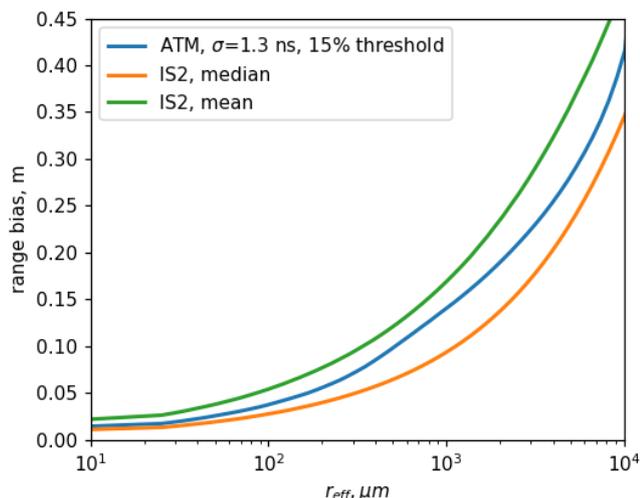



Figure 4. Predicted range bias for ATM and ICESat-2 waveforms. ATM biases were calculated using a mean-based retracker with a 15% amplitude threshold. ICESat-2 biases were calculated using a windowed median and a windowed
305   **mean retracker.**

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

310   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.

315

3. Results

3.1 Recovered grain sizes

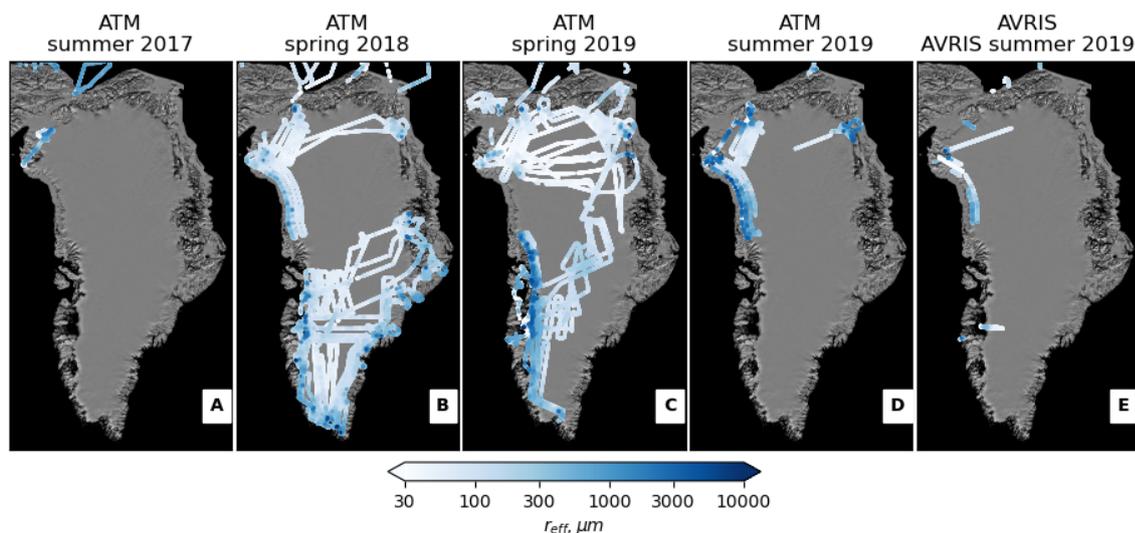



Figure 5. Recovered grain sizes from ATM and AVIRIS. Colored points indicate recovered grain sizes for four ATM campaigns
320 **(A-D) and for AVIRIS (E). Each color-coded points indicates a 1-km block median of recovered grain sizes, and the points have been plotted in order of grain size, so that coarser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

Figure 5 shows maps of recovered grain size from ATM for the summer of 2017, the spring seasons of 2018 and 2019, and the late summer of 2019, and the valid AVIRIS surveys for the late summer of 2019. These maps show a trend from large

325 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 was carried out earlier than the corresponding portion of the spring-2019 survey, and encountered coarser grain sizes, particularly along the coast, while grain sizes in the northern parts of both surveys were consistently fine. The summer surveys in 2017 and 2019 both encountered coarse grain sizes, particularly in the coast-parallel lines in 2019. The AVIRIS survey from 2019 has most of its overlap with the

330 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 by usable AVIRIS data.

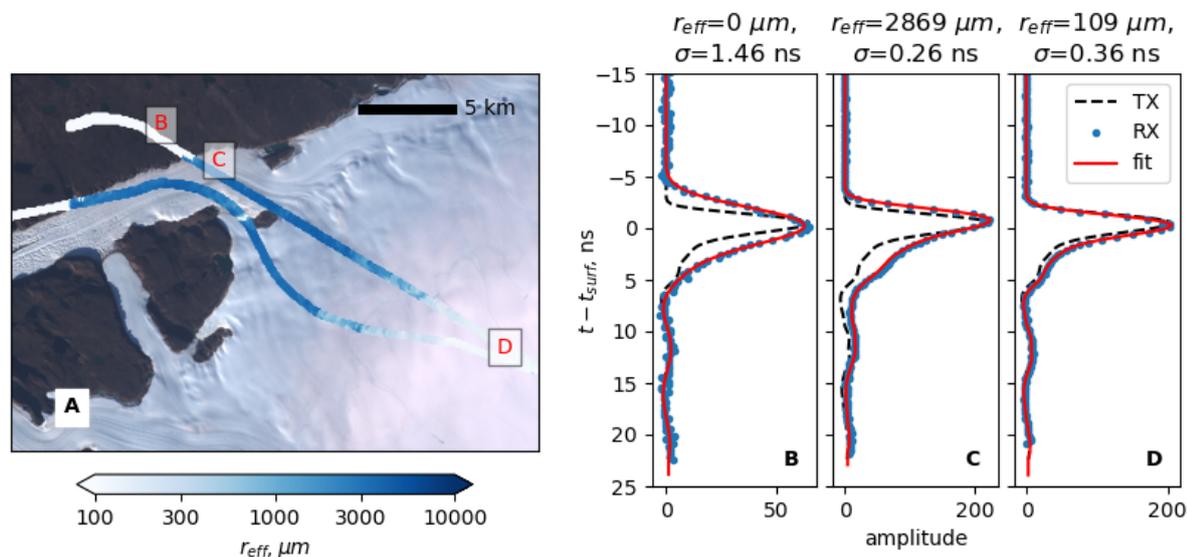



Figure 6. Grain size and waveforms. A. True-color Landsat image of the Northeast Greenland ice sheet near Leidy Glacier from 6
August, 2019, with estimated effective grain size (reff)) from ATM. For the ATM data, we plot the results of a 100-m blockmedian
applied to reff. Panels B, C, and D show measured (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.

Figure 6 shows a map of recovered grain size, and three waveforms measured in Greenland, one measured over a rock/soil
surface (panel B), one over low-elevation coarse-grained melting ice, and a third from finer-grained snow, 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 (panel C)
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 (panel D) 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. The mapped distribution of grain sizes (panel A) shows little or no subsurface scattering on rock and soil
($r_{eff} \approx 0$), strong subsurface scattering for low-elevation ice that has experienced extensive melt ($r_{eff} > 1000 \, \mu m$), and
weaker subsurface scattering at higher elevations ($r_{eff} < 200 \, \mu m$).

3.2 Comparisons of recovered grain sizes between two independent ATM instruments

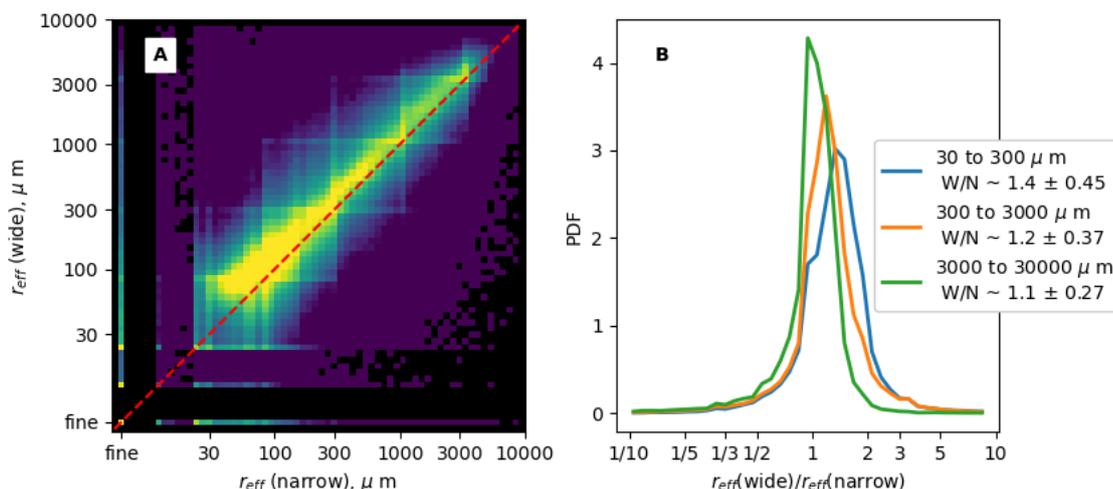



Figure 7. Recovered grain sizes from two ATM systems from the Summer 2019 campaign. Panel A shows the density of measurements as a function of recovered r_{eff} values from the narrow and wide-scan ATM systems (lighter colors represent a higher density of measurements). Points for which one of the systems 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 standard deviation of the ratios for each range.

Figure 7A shows a two-dimensional histogram of grain-size estimates from the two ATM sensors. The estimates are clustered around the 1:1 line, with slightly larger grain-size estimates from the wide-swath 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 grainsize 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 to the left of and below the range of possible fit values (labeled 'fine' in 6A). 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.

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 (equal to the half-width of the central 68% of the distribution, approximately the one-sigma range) 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, indicating close agreement between the two systems.

3.3 Comparisons between grain sizes derived from ATM and AVIRIS

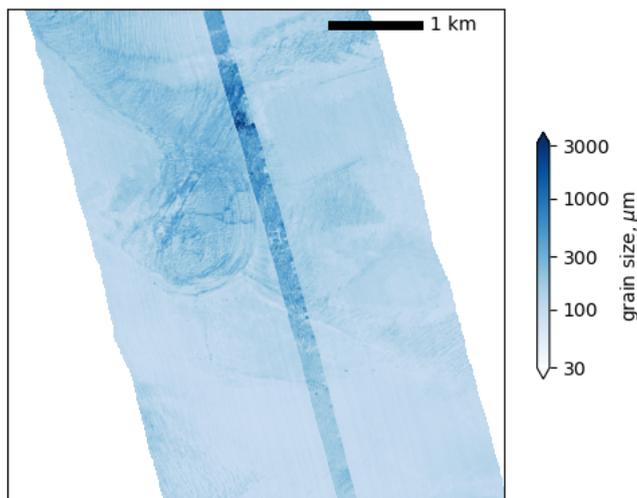



Figure 8. Sample of AVIRIS- and ATM-derived grain-size estimates for a coastal location in Greenland. The grain size based on the complete 4-km AVIRIS swath is shown, with a 10-m block median of the recovered grainsize from the 250-m wide-scan ATM swath superimposed on top. The scene center is approximately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

Figure 8 shows maps of grain size estimates from the wide-swath ATM scanner and from AVIRIS 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. The patterns of variation are very similar between the sensors, although 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.

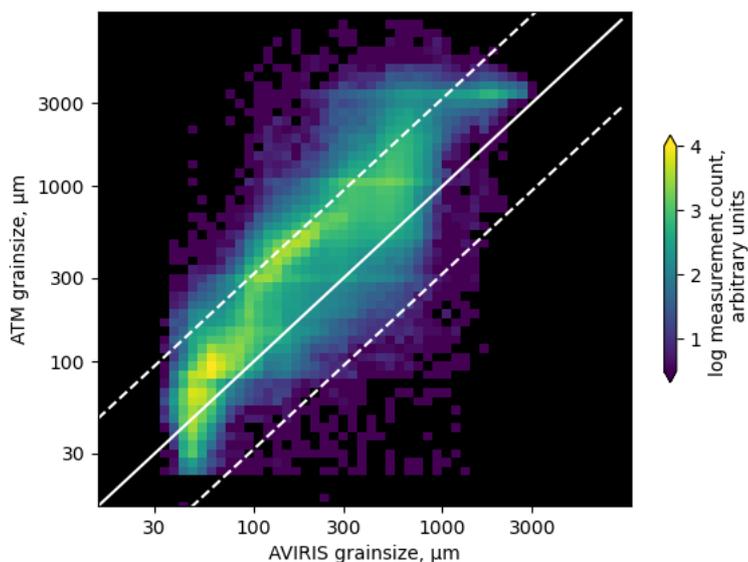



Figure 9. Two-dimensional histogram comparing AVIRIS-NG-derived grain size with ATM-derived grain size, with cells colored by the number of points observed. The solid white line shows the 1:1 relationship between the two datasets, the two dashed lines show the ATM : 2 x AVIRIS-NG (upper) and ATM : 0.5 x AVIRIS-NG (lower) relationships.

Figure 9 shows a comparison between grain-size estimates from AVIRIS and those from ATM. The two show a generally similar trend, although ATM grain sizes are typically around 2-3 times larger than the corresponding AVIRIS grain sizes. This relationship is less pronounced towards the small-grainsize side of the plot, where we see more values where ATM grainsize approximately equal to, or smaller than, AVIRIS grainsize. We believe that this comes about because the AVIRIS dataset includes few points smaller than about 40 μm, so points where the ATM fit estimated grain sizes smaller than this are mapped into a near-vertical feature at the left-hand side of the plot. 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.

3.4 Comparison between satellite, AVIRIS, and ATM grain sizes

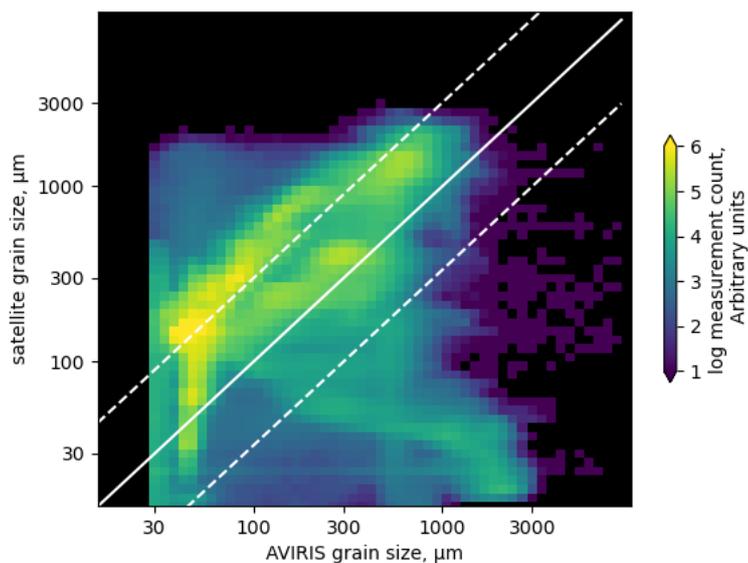



Figure 10. Comparison between AVIRIS-derived grain sizes and satellite-derived grain sizes. The solid white line shows the 1:1 relationship between the two datasets, the two dashed lines show the satellite: 3 times AVIRIS (upper) and satellite : 1/3 AVIRIS (lower) relationships. All satellite measurements were collected within ½ day of the AVIRIS measurement,

Figure 10 shows a comparison between AVIRIS-derived grain sizes and satellite-derived grain sizes. The largest concentration of satellite grain sizes is between three and four times larger than the corresponding AVIRIS sizes. As in the comparison between ATM and AVIRIS, there is a vertical feature in the distribution at AVIRIS grain size = 40-50 µm, which likely corresponds to the fine-grained limit of the AVIRIS data. The distribution of measurements for which the satellite grain-size estimates are substantially finer than the AVIRIS estimates may reflect contamination with undetected clouds in the satellite imagery, which would tend to bias the satellite estimates in the fine-grained direction.

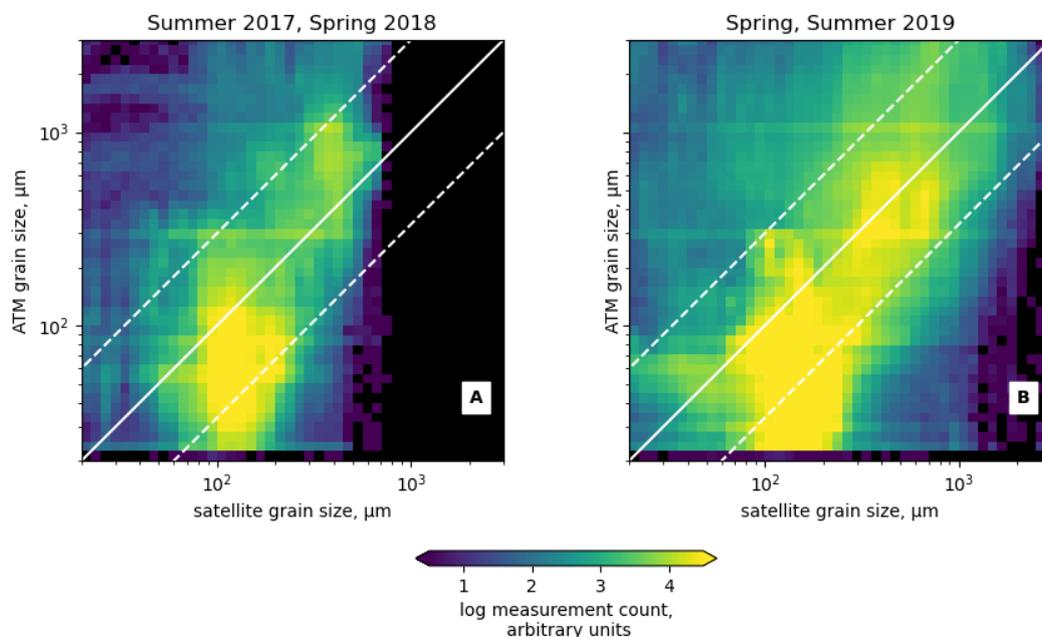



Figure 11. Comparison between ATM-derived and satellite-derived grain-size estimates. Panel A shows the distribution for the summer of 2017 and the spring of 2018, Panel B shows the distribution from the spring and summer of 2019. In both plots, the ATM grain sizes are derived from a 10-meter blockmedian of the data, and only those points for which the time difference between the satellite 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.

Figure 11 shows a comparison between satellite-derived grain sizes and those from ATM, based on a combination of data from

the summer of 2017 and the spring of 2018 (10A) and from the spring and summer of 2019 (10B). In each case, the

distributions of both types of grain size measurements roughly follow the 1:1 line, although for both years, the ATM

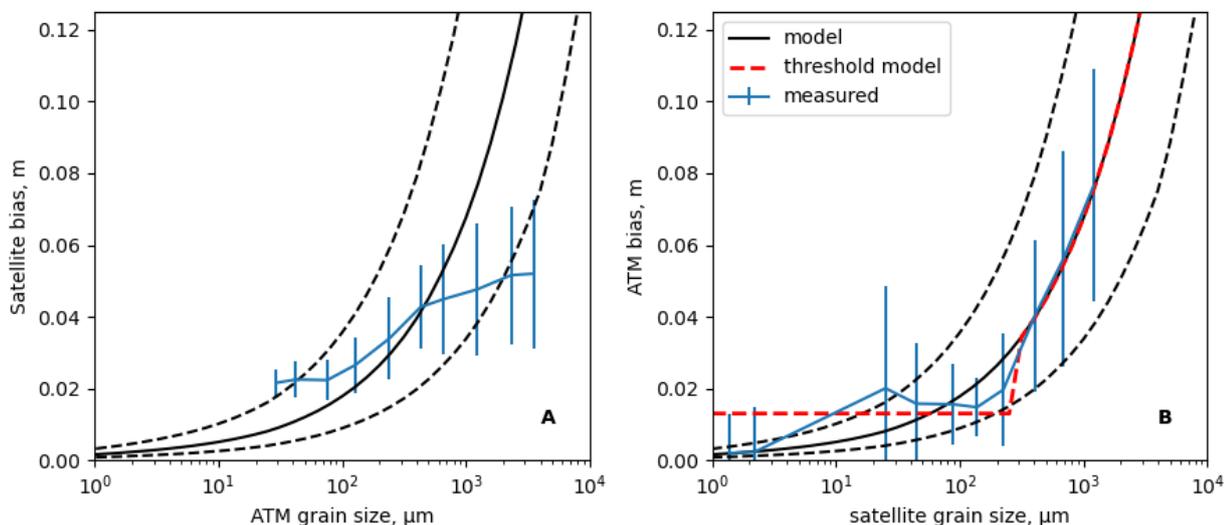
measurements show a range of measurements smaller than 100 μm for which the satellite measurements are clustered around

100 μm. This may indicate that there are conditions under which the satellite measurements cluster around a moderately small

grain size while ATM maintains sensitivity at smaller grain sizes. The 2017-2018 panel (A) 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.

3.5 Comparing subsurface-scattering bias estimates between satellite and ATM data

435 **Figure 12. Range biases as a function of grain size estimates for the complete 2017-19 dataset. (A) shows range biases predicted from satellite grain-size estimates as a function of ATM grain size, B shows range biases estimated from ATM grain sizes as a function of satellite 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.**

440 Figure 12 shows ICESat-2 range biases predicted based on ATM and based on satellite grain-size estimates. To estimate the accuracy of satellite-based grain-size estimates for each ATM measurement, we calculate the median of all satellite-derived range-bias predictions for ranges of ATM- measured grain size (in logarithmic bins with spacing $10^{0.25}$). The relationship between the two shows that the satellite-derived grain sizes generally increase with ATM grain size but badly underestimate the grain size for large values of the ATM grain size. Second, to estimate the accuracy of satellite-derived grain sizes as a

445 representation of ICESat-2 biases averaged over large numbers of measurements, we plot the median of ATM-derived bias estimates over ranges of satellite-derived grain sizes (also in logarithmic bins spaced by $10^{0.25}$) against the bias estimates calculated for the satellite-derived grain sizes. These values show a close match between the biases predicted based on the ATM data and those predicted based on the satellite measurements for satellite-estimated grain sizes larger than around 250 µm; at smaller grain sizes, the ATM biases deviate from the satellite biases, with a roughly uniform value close to 0.02 m.

450

3.4 Calculating a best-feasible correction.

Based on figure 12, panel B, it appears reasonable to believe that ATM waveforms and satellite grain size estimates provide information about subsurface delays for coarse-grained snow, but not for fine-grained snow. This suggests that we might correct for subsurface penetration with a function of the form:

$$B_{thr}(r_{sat}) = \begin{cases} B_0 & : r_{sat} < r_{thr} \\ B_{med}(r_{sat}) & : r_{sat} > r_{thr} \end{cases} \qquad 7$$

455

Here r_{sat} is the satellite-estimated grain size, $B_{med}(r_{sat})$ 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 threshold used for satellite grain sizes smaller than r_{thr}. We used the ATM and satellite grain sizes from 2017 - 2019 to find optimal parameter values for this model: For a range of B_{thr} and r_{thr}, we calculated the median and the robust spread of the distribution of corrected ATM biases, $B_{med}(r_{ATM}) -$

460 $B_{thr}(r_{sat})$. For the sake of computational efficiency, and to help match the resolution between the ATM and the satellite grain-size estimates, we carried out these calculations on a 250-m blockmedian of the ATM measurements. Figure 13 (panels A, B) show how the median and the robust spread depended on the parameter values. For threshold values greater than about 150 μm, there is a threshold (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=0.013 m, r_{thr}=280 μm. Figure 13c shows the

465 distributions of ATM-derived biases, ATM-derived biases corrected based on $B_{med}(r_{sat})$, and of ATM-derived biases corrected based on $B_{thr}(r_{sat})$ using the optimized parameters. The uncorrected distribution of ATM-derived biases has a peak at around 1 cm a median of 0.016 m, with a substantial tail of values extending in the positive direction. Applying the unmodified correction results in a more compact distribution of residuals, with a median of -0.08m and a spread of 0.008 m, both of which are an improvement on the raw distribution but biased in the opposite direction. The optimized threshold model

470 yields a distribution of residuals with a zero median and a robust spread of 0.006 m. We present this analysis using the median and robust spread because they are less affected by outlying data points. Repeating the analysis using the mean and the standard deviation of the corrected datasets yields similar optimum B_0 and r_{thr} values (0.014 m and 250 μm, respectively) for the zero-mean-residual model with the smallest standard deviation, but finds that for this model, the standard deviation of the optimal model is approximately the same as that for the non-optimized correction (0.014 m in either case).

475

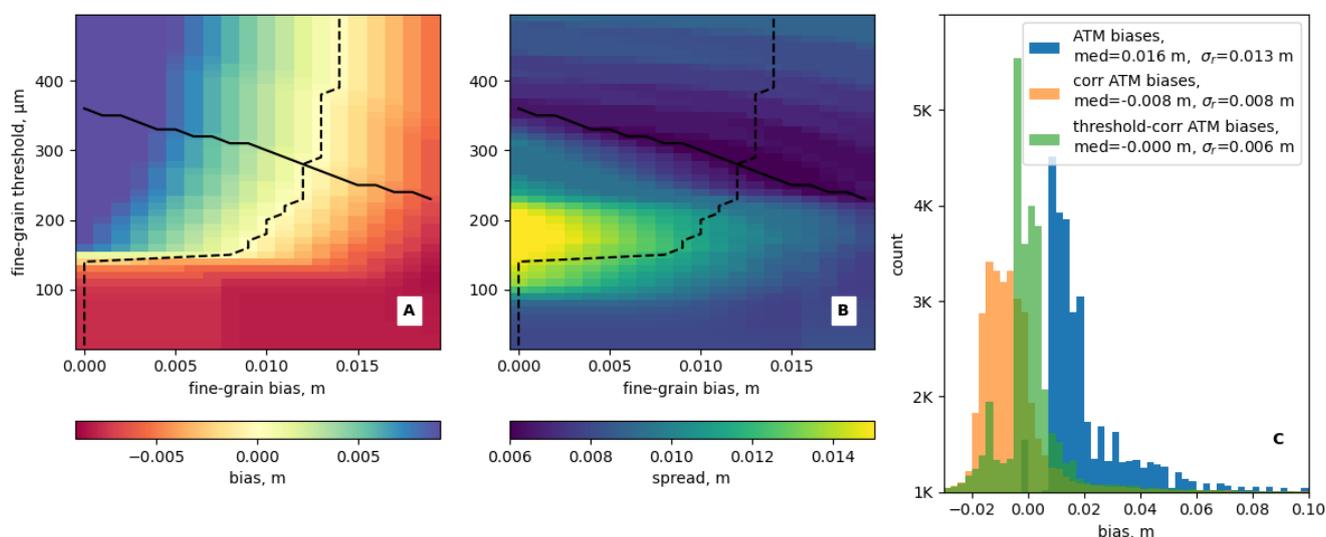



Figure 13. Tuning the threshold correction for ATM-based ICESat-2 bias estimates. Panels A and B show the median and robust spread (equal to the half width of the central 68% of the distribution) of the distribution of ATM-derived ICESat-2 bias estimates

corrected with the threshold model (equation 7) 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 $B_{med}(r_{sat})$, and bias estimates corrected based on $B_{thr}(r_{sat})$ for the optimum parameters, B_0=0.013 m, r_{thr}=280 μm. The median and robust spread of each distribution is given in the legend.

4. Discussion:

The comparison of measurements between the narrow and wide-swath instruments shows that ATM-based estimates of 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 blurring of the returned waveform by the angle between the beam and the surface likely results in a large bias in the wide-swath 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 both consistent and should be useful in predicting biases in ICESat-2 data. The comparisons between AVIRIS-NG grain size ATM grain size, and those between AVIRIS-NG grain size and satellite-derived grain size 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 satellite 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 these AVIRIS-NG measurements and ATM-based measurements are consistent with comparisons between this AVIRIS survey and observations of apparent elevation differences between green and near-infrared altimetry measurements that also implied that the AVIRIS data had underestimated grain sizes (Fair et al., n.d.). Despite these limitations, the comparisons between ATM, satellite, and AVIRIS measurements show a consistent trend between the three sets of data, with larger grain sizes in each dataset corresponding to larger grain sizes in the others. 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 satellite-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 satellite measurements. Similarly, the smallest, coarsest-grained features in the AVIRIS-NG dataset are not expected to be resolved in the satellite 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 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. 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

515 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

520 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 it

525 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

530 to be larger than a factor of two (see figure 1).

The comparison between predicted ICESat-2 biases derived from ATM and those from the satellite measurements suggests that while satellite 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 in seen in

535 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 satellite 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 satellite measurements shows good agreement with the predicted grain-size-vs-bias curve, likely because the median biases for the collections of ATM measurements are only

540 weakly affected by the minority of ATM measurements collected over large-grain-size features. Further, the discrepancies between ATM and satellite-derived grain sizes in the small-grain-size regime (figure 11) should have relatively little impact on the accuracy of a satellite-based prediction of biases in ICESat-2 data, because whatever their disagreements, the two datasets agree that the bias correction should be small. The peak in the ATM-bias-vs-satellite-grain-size (figure 11, panel B) likely reflects undetected clouds in the satellite data set; for these measurements, the ATM bias can have a large range of

545 values, while the satellite reports a grain size appropriate for polar clouds. 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 satellite-derived grain-size estimates (figure 13) show that for the full dataset, the mismatch between satellite and ATM resolution and the imprecisions of the two datasets for small grain

550 sizes result in a net overcorrection of the biases (shown in figure 13, 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 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 satellite

555 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 figure 13:

    1.   Generate grain-size estimates for each ATM data point ($r_{est,sat}$)

560   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 $B_{med}(r_{ATM}) - B_{thr}(r_{sat})$ (equation 7)

    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

565 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.

5. Conclusions

In this study we have demonstrated a technique for the retrieval of ice-sheet surface grain size using the shape of pulses returned

570 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 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 and slightly different versions of the same instrument flown on the same aircraft at the same time, showing that the grain-size recovery is repeatable, and is not strongly sensitive to

575 the geometry of the measurements. 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
580   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 based on ATM measurements, 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-
585   2 data. To date, our efforts to identify ICESat-2 data that clearly show the effects of subsurface scattering 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. However, the synthesis of the ATM and satellite-
590   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 grain size are a potential way to improve the precision of a correction of
595   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. 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
600   grain-size estimates of providing 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. 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
605   have been stymied by the need to aggregate photons from large numbers of ICESat-2 pulses over complex topography, but
such techniques might be possible for limited areas with very flat and smooth surface topography.

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. Satellite-based grain-size
610   estimates are available through GEUS dataverse (Vandecrux et al., 2022a) AVIRIS grain-size estimates are available by FTP

from https://popo.jpl.nasa.gov/avng/y19/, and ATM-based grain-size estimates are in the process of submission to the National Snow and Ice Data Center (NSIDC), 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.

615 **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,
620 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,
625 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
630 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.

635 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
640 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.

645 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,
650 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.

655 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.

660 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.,
665 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.

Medley, B., Neumann, T. A., Zwally, H. J., Smith, B. E., and Stevens, C. M.: Simulations of firn
670 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.

675 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.

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.

680 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.

685 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, 2019.

Smith, B. E., Gardner, A., Schneider, A., and Flanner, M.: Modeling biases in laser-altimetry
690 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,
695 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,
700 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.

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.

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. 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.

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.