

## Response to Comments from Reviewer 1

*AUTHORS: We thank the Reviewer for their comments; edits based on their input have improved this manuscript.*

The authors present new elevation data collected using kinematic GPS in Antarctica and discuss their relevance toward validation of satellite laser altimetry data from the ICESat-2 mission. The paper was concise and well-written, and I am happy with publication almost as-is.

I have three major comments:

(1) While the authors discussed elevation changes associated with sastrugi migration, there was no discussion of other surface processes, primarily firn compaction in the context of

(a) Temporary (perhaps elastic) compaction of snow/firn from the weight of the PistenBully, which might not be captured from the track depth measurements. How heavy were the PistenBullys and is this effect negligible?

*AUTHORS: The reviewer is correct: Our track-depth measurements almost certainly would not capture the elastic effect you are describing. PistenBullys weigh about 10,000 kg, spread over a 10 m length and 2 m width scale. While the elastic effect of that weight pushing temporarily on the firn might not be negligible, our suspicion is that the impact is below our leading error term, which is the track depth uncertainty and non-uniformity ( $\sim 6 \text{ cm} \pm 1.5 \text{ cm}$ ). We also note that if the elastic effect is less than the precision of our GPS data ( $\sim 4 \text{ cm}$ ) then we may not be able to observe this without creative survey methods. If extant, such an elastic effect would present as a negative bias of the vehicles relative to a non-invasive data set, such as airborne or spaceborne lidar. We intend to experiment with some creative survey methods, as time allows, at the start of this field season. But currently, this is beyond the scope of this paper. Further, we have other strategic plans to reduce the track-depth uncertainty; this includes running the GPS from the sled (which seemed to float at a consistent depth in the snow), as opposed to the PistenBullys (where the tracks seemed to dig into the snow at varying depths).*

(b) Climate-driven firn compaction over < seasonal to multi-year timescales showing up as elevation differences between GPS- and Operation IceBridge-derived estimates. I think it could be useful if the authors included a time series of modelled elevation change from firn processes (data available at <http://www.staff.science.uu.nl/~ligte104/DATA/>) at one or more locations along this transect.

*AUTHORS: The reviewer is correct that time variable climate-driven firn compaction could manifest as an elevation difference between our measurements and the airborne data presented here. The seasonal effect is less of a concern as the elevation response to climate-driven firn processes are effectively annual, and in this region should be driven by the annual cycle of temperature. As the UAF data were nearly coincident in time with the GPS data collection, the impact of seasonal firn processes on elevation should be minimal. The ATM data was collected with a larger offset with respect to time of year ( $\sim 2$  months) which may be impacting our comparisons. The secular trend in elevation owing to climate-driven firn processes is a thornier problem, as in this sector the climate drivers (accumulation rate, temperature) are poorly known. Both the secular and annual trends in elevation owing to firn processes will be a primary concern in our second paper on the topic, which will compare two seasons of ground-based data collection with airborne and spaceborne lidar data. We have added the following text to the document:*

*"We note that our analysis does not attempt to account for elevation changes due to the temperature- and accumulation-rate-driven effects of firn compaction (Li and Zwally, 2015). In this region, we expect variation in firn compaction rate to be driven by changes in firn temperature, which have a large seasonal amplitude and a much smaller secular trend. As the firn warms each austral spring, the surface elevation along our traverse should decrease. Since the UAF lidar data and ground-based GPS data were collected within a month, we expect firn compaction to have a negligible effect on our results. Conversely, the  $\sim 2$  month seasonal lag between the*

*ATM and GPS data collection means that we may be sensitive to the seasonality of firn compaction rate, as well as any secular trend over the 4 year interval between these data sets."*

(2) The authors mentioned that there were anomalous elevations at the edge of the UAF lidar swath. This raised a red flag for me: Is there a possible connection between scan angle and elevation accuracy for the airborne lidar systems (even in campaigns where elevations did not have an across-track tilt)? The authors could include elevation differences between GPS and airborne lidar as a function of scan angle, especially since their data will likely capture a fairly wide range of scan angles. This analysis will tell us if we should only use airborne lidar data from a particular range of scan angles when comparing with ICESat-2 altimetry.

*AUTHORS: We do not think that scan angle is the root cause of this artifact. We note that the trough is isolated to one side (in our case the southern side) of the transect. Our expectation is that if the artifact was associated with scan angle, some matching form of the trough would also appear on the other (northern edge).*

(3) It appears to me that the UAF lidar data from 3rd December 2017 were of lower quality than those from 30th November 2017 (Table 1, 2, and Figure 6 Panel C). However, biases appear to be within  $1\sigma$  uncertainties, so perhaps this difference isn't significant enough to require further discussion.

*AUTHORS: Our experience is that mature airborne lidar accuracies and precisions are generally under  $10\text{ cm} \pm 15\text{ cm}$ . Therefore, because data from both flights fell within this general rule of thumb, we put no stock in one having a lower bias; we do not think that this suggests that that flight is significantly better than the other. The reviewer makes a great point; and there is language (some new language) that touches on this in the text (e.g., first few lines of section 4.3, which we have augmented):*

*"Both altimeters had elevation biases less than 10 cm and surface measurement precisions less than 15 cm; we note that these values are similar to those in Brunt et al. (2017), which is a similar study in a similar geophysical setting."*

Other minor comments:

1) Page 1, Line 20: change "set to launch" to "launched". Yay!

*AUTHORS: Done! Yippee!!*

2) Section 3.1: What cut-off angle was used in the processing?

*AUTHORS: Good addition. We added the following text:*

*"We used a GPS satellite elevation mask, or a cut-off angle, of 7.5 degrees, to minimize the effects of multipath."*

3) It would be nice to have larger font sizes in Figures 6 and 7.

*AUTHORS: This was done in conjunction with other edits suggested by Reviewer 2.*

Discussion: I think a significant number of issues with using GPS data to validate airborne and satellite altimetry could potentially be mitigated in the future with the use of a terrestrial laser scanner (TLS) mounted near the GPS antenna on the PistenBullys.

*AUTHORS: We plan on attempting to make a change in the coming season that starts to get at the reviewer's point. We will try to integrate a downward looking laser, next to the GPS antenna, off of the side of the rear of the sled. This should give us an 'along-track' profile of surface roughness. While this does not provide the 3D sense of roughness around our GPS survey data (that the TLS would provide), it may help beat down the error term associated with track depth.*

Thank you again,  
Brunt, Neumann, Larsen

## Response to Comments from Reviewer 2

*AUTHORS: We thank the Reviewer for their attention to detail; their input has contributed significantly to this manuscript.*

The paper reports the acquisition and processing of ground GPS data along the 88S Traverse and compares the surface elevations to those derived from airborne laser altimetry. The ultimate goal of the 88 Traverse is to validate ICESat-2 surface elevation data. Ground GPS data were processed by using Precise Point Positioning (PPP) post-processing method. The paper provides a detailed analysis of surface elevation mapping results from ground-based GPS and airborne laser altimetry.

The manuscript would benefit from including a summary on how the ground and airborne elevation datasets will be used for ICESat-2 validation. While the authors make a strong argument for using the 88S traverse because of the abundance of ICESat-2 measurements, they don't elaborate on how the data sets collected at different times and characterized by different errors will help to validate the ICESat-2.

*AUTHORS: We've added a long paragraph to the end of the discussion that ties together the ground-based and airborne datasets and their relation to ICESat-2 validation. We mention the strength of this dataset and that the methods presented here provide a framework for data analysis in the future:*

*"Data collected from the 88S Traverse (and data collected on subsequent surveys of the same route) will provide 300 km of in situ data for direct comparison with ICESat-2 elevation data products. The GPS data collection strategies and post-processing methods presented here provide accurate and precise data for such an assessment. Further, the data analysis presented here provides guidance on how to make similar comparisons between ground-based and satellite elevations, given the satellite footprint size and associated rejection criteria. Approximately 3 to 4 ICESat-2 reference ground tracks will intersect this region daily to produce many statistical crossover points between the GPS and ICESat-2 datasets. While the crossover points represent only a small segment of along-track ICESat-2 data, the analysis will be based on data from many ICESat-2 reference ground tracks over the course of the entire satellite mission. Thus, the analysis of the derived ICESat-2 bias and surface measurement precision relative to these GPS data will provide an assessment of how ICESat-2 is performing through time, independent of errors associated with single orbits or single points in time. Results of Brunt et al. (2017) and results presented here also provide an assessment of the accuracy and surface measurement precision of 3 airborne lidars that NASA has routinely deployed over the ice sheets (ATM, LVIS, and the UAF lidar). With a statistical understanding of how these instruments perform on the relatively flat ice-sheet interiors, longer flight lines can be constructed over similar such flat ice-sheet surfaces to create better statistics associated with comparisons using long length scales of along-track ICESat-2 data. In summary, the strategic location of the ground-based 88S Traverse provides a validation of ICESat-2 that is independent of the errors that are correlated with respect to most satellite time scales, and these ground-based data provide s with a better understanding of airborne lidars that will survey longer length scales of data, for better satellite error statistics."*

In my opinion, the analysis presented in the paper does not provide sufficient support for the assumption that the ground GPS data depicts the real surface and thus provides ground truth for the airborne laser altimetry. While the PPP processing used for processing kinematic ground GPS data can provide vertical accuracy up to 3 cm, errors can be up to 10 cm (e.g., <https://www.novatel.com/an-introduction-to-gnss/chapter-5-resolving-errors/gnss-data-post-processing/>), or somewhat lower because of the filtering the authors applied. The vertical accuracy of the ATM laser altimetry points is similar to the accuracy of PPP processing. For example, Martin et al., 2012 (Table 7) quoted a vertical accuracy of 6.6 cm with a vertical precision of 3 cm for an operating altitude of 500-750 meter above the surface.

To demonstrate the accuracy of the ground GPS surveys, the authors compared the measurements made with vehicles A and B and obtained a small elevation difference of 1.14.1 cm. However, this comparison does not fully characterize the absolute accuracy of surveys. Some of the errors, in particular, the ionosphere and troposphere errors, could be quite large and correlated within a short time that elapsed between data acquired with the GPS systems on the two vehicles.

**AUTHORS:** These 2 comments are very similar and are addressed in this single response. The reviewer makes valid points about our results and the limitations of PPP post-processing. We note that we are limited with respect to options for processing to the best possible position solution; there are very long baselines to any ground-based permanent station, which removes the differential GPS (DGPS) option. Therefore, we tried to make an assessment of the quality of these data, with only the 2 GPS data streams to work with. The crossover comparison seemed like the best method for making this assessment, and follows similar approaches of Siegfried et al. (2011), Kohler et al. (2013) and Brunt et al. (2017).

We specifically state that we take our GPS measurements to “*represent truth, with zero errors. In actuality, these errors are not zero and are a function of: ... 2) ionosphere and troposphere errors; ....*” Thus, we feel that we made an effort to highlight the issue that the reviewer mentions.

But we fully agree with the reviewer’s assessment that the GPS A to GPS B statistics are probably better than would be if the vehicles surveyed the region on different days (which is not possible for operational safety reasons). Therefore, we put further caveats on the ‘truth’ statement above (in Section 3) and in the Discussion section, where we discuss the results of the GPS data analysis (Section 5):

Section 3:

*“represent truth, with zero errors. In actuality, these errors are not zero and are a function of: ... 2) ionosphere and troposphere errors; ... We note that given the short distance between the 2 survey vehicles, our results are somewhat blind to the full magnitude of the error terms that can be correlated on short time scales, such as those associated with the ionosphere and troposphere.”*

Section 5:

*“As stated above, we note that we are blind to errors introduced by the close spatial coincidence of the GPS receivers (~50 m) and to those introduced by the common processing of the GPS data. Errors in ionospheric or tropospheric modeling would impact both GPS-based data sets similarly and would introduce a bias between the GPS measurements and the actual ice sheet surface.”*

In light of this, and subsequent conversations, we plan on attempting to make a change in the coming season that starts to get at the reviewer’s point. We will try to survey the first 10 or so km of the traverse route a few days prior to the actual start of the full traverse. These data will help assess the magnitude of the troposphere errors.

We note that we also added the numbers of Martin et al. (2012) in the Discussion section.

Thus, the results presented in the paper do not adequately support the attribution of the elevation difference between ground GPS and airborne lidar surveys to a single source, i.e., to the random error (precision) of the airborne laser altimetry data sets. Instead, the elevation differences between the different airborne surveys (not included) as well as relative to the ground GPS survey (Figures 4 and 5) should be analyzed with the inclusion of any available calibration and/or validation information. For example, the correlated patterns in the residual elevation differences between the airborne and ground survey have characteristic spatial wavelengths of 15-50 km that corresponds to 2-7 hours with the ground speed of  $2 \text{ m s}^{-2}$ . This pattern might indicate modeling errors of the ionospheric and tropospheric corrections which typically decorrelate within a few hours. Only a comprehensive analysis could shed light on the source of the errors and attribute them to the different surveys.

AUTHORS: Using these 3 datasets, we do not feel that it is not possible to definitively prove that any particular data set measures the true surface elevation, owing in part to some of the reviewer's concerns (e.g., tropospheric effects on the ground-based dataset). However, by assuming that one data set represents the surface elevation (i.e., 'the truth') we can evaluate the corresponding data sets in reference to it. See comments above on how we have softened language associated with taking our ground-based data as the truth and then making assessments of the other datasets based on that.

Detailed comments:

General comments:

- The description of the horizontal and vertical reference frames is somewhat confusing in the paper. The reference of the geographic coordinates is not mentioned and the elevations are described to be referenced to the ITRF reference frame of the WGS-84 ellipsoid. The correct information for ATM laser altimetry is that elevations are given in the ITRF08 reference frame, and geographic coordinates are referenced to the WGS84 ellipsoid (e.g, <https://nsidc.org/data/ILATM1B>). Similar information should be included for the other data sets. AUTHORS: We've made these clarifications for all 3 references to ITRF (in the traverse data, and both lidar datasets).

Page 1, lines 19-25: of course, these sentences should be in past tense after the successful launch of ICESat-2  
AUTHORS: The other reviewer caught this too! We've made this change (although if we had used past tense in the submitted version of this paper, we surely would have jinxed ourselves!)

Line 24: include the spatial scale of the ICESat-2 mission requirement of 0.4 cm/yr

AUTHORS: While most ICESat-2 requirements have along-track length scales, this particular requirement is a 'whole ice sheet' requirement. But we did add text associated with the time requirement (on an annual basis).

Page 2, lines 13-18: according to the authors the ground tracks will be spaced randomly in time (line 15) and evenly (line 16). It is a confusing description and needs clarification. Also, how are the ground tracks distributed relative to the ICESat-2 subcycles?

AUTHORS: Great point. We think that this edit mitigates your primary concern:

*"Therefore, the 20% of the 1387 unique tracks that are intersected by this survey represent data collected from the whole 91-day orbital cycle, and not a specific 20% of the cycle. Further, since the ground tracks intersected by the 88S Traverse are spread throughout the 91-day orbital cycle, ..."*

ICESat-2 does have a ~30-day subcycle to meet the sea ice requirement of a monthly sea ice freeboard data product. The requirement specifically stipulates 35 km spacing at 70° latitude. But there is no operational impact of this subcycle requirement and it's not germane to the ice sheet analysis described here. We have therefore decided not to introduce this concept, as it may confuse readers.

Lines 24-30: these statements are too concise to understand without consulting the referenced studies. For example, what are the surveys to which the surface measurement precisions refer?

AUTHORS: We have broken this section into 2 distinct pieces (Brunt et al., 2018, and Kohler et al., 2013) and added text for clarity; the text addresses numbers of campaigns (for Brunt et al.) or length-scales of data (i.e., 'the entire traverse', for Kohler et al.):

*Using Precise Point Positioning (PPP) post-processing methods for 6 ground-based GPS surveys, elevation biases for the associated 6 ATM airborne surveys (conducted between 2009 and 2016) ranged from -10.8 to 0.8 cm, while surface measurement precisions were equal to or better than 8.7 cm. Using the same methods for 2*

*ground-based GPS surveys, elevation biases for 2 LVIS airborne surveys (conducted in 2007 and 2010) ranged from -2.7 to 8.2 cm, while surface measurement precisions were equal to or better than 6.1 cm. Their results suggest that for a flat, relatively smooth and homogeneous surface, these altimeters provide consistent results, which are required for an airborne component of an ICESat-2 validation strategy. Kohler et al. (2013) collected 5000 km of ground-based kinematic GPS data along the Norway-USA East Antarctic Traverse, in 2 different vehicles and over the course of 2 different field campaigns, for direct comparison with ICESat elevation data from all of the satellite campaigns typically used for data analysis (e.g., L2A through L2E). Using PPP post-processing methods, elevation biases for ICESat, based on ground-based data from the entire traverse, ranged from -12 to -2 cm, while surface measurement precisions were equal to or better than 15.8 cm. Their results were based on cross-over analysis between ground-based measurements and the last 2 years of ICESat data.”*

Lines 32-33: use the words spaceborne and airborne consistently, preferably without a Hyphen  
AUTHORS: Good edit. We went with ‘airborne’ and ‘spaceborne’ throughout.

Line 33: does the surface elevation change minimally or not? In my opinion, enough is known about the elevation changes of the East Antarctic Plateau to make a definitive statement and even quantify the changes  
AUTHORS: The authors would argue that very little is known about this region with respect to overall change and specifically to the terms listed (ice flow, snow accumulation, and surface melt). This area is generally within the satellite pole hole of most polar-orbiting satellites, and surface measurements are sparse in the region. Thus, we are comfortable with the language used here.

Page 3, section 2.1: more details about the kinematic GPS surveys should be included. What was the typical distance (space and/or time) between the two vehicles during data acquisition? Did they always travel in the same order or did they switch place? Was there any static GPS collected?

AUTHORS: All great additions. We have added text throughout section 2.1 that addresses all of these comments. Plus, the other reviewer asked about cut-off angles in processing. We have added these details.

Line 10-12: How many times was the antenna height measured? If only twice (line 10), the standard deviation mentioned in line 12 is not meaningful.

AUTHORS: Good point. We removed the standard deviations and instead listed the measurements, e.g.: “*The average antenna heights for the 2 vehicles were 281.3 cm (vehicle A, 280.7 cm and 281.9 cm) and ...*”

UAF lidar: the configuration of the UAF lidar – line scanner – should be mentioned.

AUTHORS: Good point. We added this text.

ATM: please refer to the information included in Martin et al., (2012) about the accuracy and precision of the ATM system, instead of the somewhat outdated Krabill et al., 2002

AUTHORS: We have added the Martin reference in this section of the paper; and the details of their vertical accuracy/precision results ( $6.6 \pm 3$  cm) was added to page 8, where it was more applicable.

Page 4, lines 11-14: What is the difference between the ATM T4 and T6 systems? What was the flying height of those ATM flights that are not included? Is there a study/personal communication that can be referenced about the presence and cause of the across-track tilt, i.e., uncorrected attitude error?

AUTHORS: ATM: Our group is not knowledgeable on the differences between the T4 and T6 versions of ATM, and given that we do not move forward with the 2016 data, this is a bit outside the scope of the paper. We have added the survey altitude (~450 m AGL) and a personal communication remark (Michael Studinger) associated with cross-track tilt; but as of writing, it is our understanding that the source was still unknown.

Page 5, lines 4-15: it would be useful to repeat the data acquisition times in this section.

**AUTHORS:** This is a good addition. We were able to add times for ATM, but the UAF lidar do not have associated times; but we added text that could point future users to the exact files.

Line 11: use “geographic coordinates” instead of latitudes and longitudes.

**AUTHORS:** Done.

Line 26: what is the spatial scale to compute the standard deviation? Why using 1- sigma instead of the more widely used 2-sigma or 3-sigma?

**AUTHORS:** All of the biases and standard deviations presented here are calculated using the full run of the dataset (~300 km). With respect to the use of 1-sigma, we are building off of results from other ICESat and ICESat-2 validation efforts (e.g., Fricker et al., 2005, Kohler et al., 2013, Borsa et al., 2014, and Brunt et al., 2016). Results presented here can therefore be directly compared with the results from the other papers. We note that all of these papers are very specific about exactly what they are presenting (e.g., but stating ‘SD’, or ‘1-sigma’).

Page 6, line 24: which flight provided higher elevations?

**AUTHORS:** Data from the 30 November 2017 flight were lower than data from the 3 December 2017 flight; we’ve added this text to the manuscript.

Page 7, lines 5-22: Were the biases removed from the differences before computing the variograms? Was there any additional preprocessing applied? Using the standard terminology for the description of the variograms (sill, range, etc.) would improve the section. The difference between elevations from GPS unit A and ATM (Figure 6, top panel) has a non-zero nugget effect, while the other differences do not exhibit any. What could cause this difference?

**AUTHORS:** The biases were not removed prior to the analysis, and no further pre-processing was applied. We have added the ‘sill’ and ‘range’ terms to a few places in the text.

With respect to GPS A and ATM (which is now the bottom panel of Fig 6 to be more uniform with other figures and the layout of the text), the reviewer makes a good point. We note that this is also the comparison with the greatest range, at ~50 m; the others tend to be <20 m. We assume that this is associated with the sastrugi migration, since it has to be either (or a combination of) measurement error or small scale variation (smaller than sampling scale). The sastrugi migration would be in the approximate size/distance scale. But this is a bit of speculation. The effect of sastrugi on the differences in the overall results is discussed more coherently in the ‘Discussion’ section.

Lines 30-32: contrary to the statement about “no discernable signal is evident,” in my opinion, all differences show a clear and correlated spatial variation, which requires an explanation.

**AUTHORS:** We respectfully disagree. We see three instances where there is a spatial correlation across panels 5b, 5c, and 5d (at 0 easting, 110 easting and 170 easting). These could be the result of many things, but track depth changes would be an obvious candidate. And we have already commented that we feel that this is the leading error term in our survey and analysis (Section 5).

We note that this text was ultimately removed, as it was not relevant to the GPS-to-GPS paragraph topic.

Figures:

Figure 1: adding the easting/northing would allow cross-referencing this figure with figures 4-5.

**AUTHORS:** Good addition. This has been added.

Figure 3: explain the significance and meaning of the different shades of red

**AUTHORS:** The red distinguishes the 2 GPS units. We have added text for clarification:

*... "and the 88S Traverse ground-based GPS data are in shades of red (GPS A is in light red, while GPS B is in dark red)."*

Figure 4-5: X and Y axis titles should use larger fonts. The vertical axis should refer to the computed quantity, eg., "Residual elevation difference (m)" rather than the interpretation "Precision." Adding the data sources, e.g., GPS A minus UAF Lidar Dec 3, 2017, as a title to each panel would make it possible to view the figure without reference to the figure caption

**AUTHORS:** All great changes; these have been made.

Figure 6-7: axis titles and labels are tiny, should be larger. The figure caption is too descriptive, use the standard terminology of variograms (sill, range, nugget) instead.

**AUTHORS:** We have made these changes to the figure. We removed some of the figure-caption text and added sill, range, and nugget to the main text.

Figure 8: describe the artifact in the title, e.g., narrow ridge of elevations parallel with flight direction

**AUTHORS:** We added language to the figure caption: *"...UAF Lidar elevations are anomalously lower within the artefact and manifest as a narrow trough, parallel to the direction of flight."*

References:

Martin, C. F., Krabill, W. B., Manizade, S. S. & Russell, R. L. Airborne Topographic Mapper Calibration Procedures and Accuracy Assessment. NASA Technical Memorandum 2012-215891, (2012).

**AUTHORS:** We have added this reference.

Thank you again,  
Brunt, Neumann, Larsen

# Assessment of altimetry using ground-based GPS data from the 88S Traverse, Antarctica, in support of ICESat-2

Kelly M. Brunt<sup>1,2</sup>, Thomas A. Neumann<sup>2</sup>, and Christopher F. Larsen<sup>3</sup>

<sup>1</sup>Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, USA

<sup>2</sup>NASA Goddard Space Flight Center, Greenbelt, MD, USA

<sup>3</sup>Geophysical Institute, University of Alaska, Fairbanks, Fairbanks, AK, USA

Correspondence to: Kelly M. Brunt (kelly.m.brunt@nasa.gov)

**Abstract.** We conducted a 750 km kinematic GPS survey, referred to as the 88S Traverse, based out of South Pole Station, Antarctica between December 2017 and January 2018. This ground-based survey was designed to validate spaceborne altimetry and airborne altimetry developed at NASA. The 88S Traverse intersects 20% of the ICESat-2 satellite orbits on a route that has been flown by 2 different Operation IceBridge airborne laser altimeters: the Airborne Topographic Mapper (ATM; 26 October 2014) and the University of Alaska, Fairbanks (UAF) Lidar (30 November and 3 December, 2017). Here we present an overview of the ground-based GPS data quality and a quantitative assessment of the airborne laser altimetry over a flat section of the ice-sheet interior. Results indicate that the GPS data are internally consistent ( $1.1 \pm 4.1$  cm). Relative to the ground-based 88S Traverse data, the elevation biases for ATM and the UAF Lidar range from -9.5 to 3.6 cm, while surface measurement precisions are equal to or better than 14.1 cm. These results suggest that the ground-based GPS data and airborne altimetry data are appropriate for the validation of ICESat-2 surface elevation data.

## 1 Introduction

The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) is a next-generation laser altimeter developed by the National Aeronautics and Space Administration (NASA) and ~~set to launch~~launched in 15 September 2018 (Markus et al., 2017). ICESat-2 will carry a single instrument, the Advanced Topographic Laser Altimeter System (ATLAS), a 6-beam photon-counting system using  $<2$  ns, 532 nm wavelength pulses with a 10 kHz repetition rate. ICESat-2 will continue NASA's multidecade effort to measure changes in the polar regions (Markus et al., 2017; Webb et al., 2012; Zwally et al., 2011), with mission requirements that include the determination of the annual ice-sheet elevation change rates to an accuracy of less than or equal to  $0.4 \text{ cm a}^{-1}$  (Markus et al., 2017).

Plans for the post-launch validation of ICESat-2 elevation data products include utilizing both ground-based and airborne elevation datasets. The relatively short ground-based datasets, such as presented here, will provide error assessments for airborne surveys, such that longer airborne surveys can then be designed with sufficient length scales to provide the data volume required for meaningful statistics of satellite data validation. The ground-based activities include the kinematic GPS

validation efforts at Summit Station, Greenland (Brunt et al., 2017) and airborne activities, such as those associated with NASA's Operation IceBridge (OIB; Koenig et al., 2010), which includes a lidar as part of the instrument suite.

In support of the ground-based component of ICESat-2 data validation, we conducted a 750 km traverse based out of South Pole Station (Fig. 1), referred to as the 88S Traverse (28 December 2017 – 10 January 2018). Kinematic GPS data collected along this traverse was used to validate airborne data and will ultimately be used for validation of ICESat-2's spaceborne datasets.

ICESat-2 will have 1387 unique orbits over a 91-day orbital cycle (i.e., all 1387 unique tracks are sampled every 91 days, or 4 times per year). The orbit has an inclination of 92°, allowing for data collection between 88° north and south. Since ICESat-2 is a 6 beam instrument, we refer to the imaginary centerline of the beam pattern as the reference ground track for each of the 1387 tracks. The 88S Traverse was designed specifically to include 300 km of data along the 88° S line of latitude, which is the latitude limit of ICESat-2 and where the ICESat-2 reference ground tracks will converge. This 300 km traverse along 88° S represents 20% of the total length of this line of latitude; the traverse route will therefore intersect 20% (277) of the 1387 ICESat-2 reference ground tracks. Because of Earth rotation, time-sequential ground tracks are not geographically sequentially spaced along the 88° S line of latitude. Therefore, the 20% of the 1387 ICESat-2 unique tracks intersected by this survey ~~are spaced randomly in time over the~~ represent data collected from the whole 91-day orbital cycle, and not a specific 20% of the cycle. Further, since the ground tracks intersected by the 88S Traverse are spread evenly throughout the 91-day orbital cycle, data from the 88S Traverse mitigates weather limitations (i.e., cloud cover) that have had an impact on other validation campaigns, which utilize only a few tracks within a small area of interest (e.g., Fricker et al., 2005).

The design of the 88S Traverse was based on validation studies associated with ICESat and OIB research. Brunt et al. (2017) used data from ~~an the~~ 11 km ground-based kinematic GPS Summit Station Traverse (~~Brunt et al., 2017; Siegfried et al., 2011) traverse near Summit Station~~, in the center of the Greenland Ice Sheet, to assess the elevation bias and surface measurement precision of OIB laser altimeters, including the Airborne Topographic Mapper (ATM) and Land, Vegetation, and Ice Sensor (LVIS). Using Precise Point Positioning (PPP) post-processing methods, for 6 ground-based GPS surveys, elevation biases for ~~the associated these altimeters~~ 6 ATM airborne surveys (conducted between 2009 and 2016) ranged from -10.8 to 8.20.8 cm, while surface measurement precisions were equal to or better than 8.7 cm. Using the same methods for 2 ground-based GPS surveys, elevation biases for 2 LVIS airborne surveys (conducted in 2007 and 2010) ranged from -2.7 to 8.2 cm, while surface measurement precisions were equal to or better than 6.1 cm. Their results suggest that for a flat, relatively smooth and homogeneous surface, these altimeters provide consistent results, which are required for an airborne component of an ICESat-2 validation strategy. Kohler et al. (2013) collected 5000 km of ground-based kinematic GPS data along the Norway-USA East Antarctic Traverse across East Antarctica, in 2 different vehicles and over the course of 2 different field campaigns, for direct comparison with ICESat elevation data from all of the satellite campaigns typically used for data analysis (e.g., L2A through L2E). Using PPP post-processing methods, elevation biases for ICESat, based on ground-based data from the entire traverse, ranged from -12 to -2 cm, while surface measurement precisions were equal to or

better than 15.8 cm. Their results were based on cross-over analysis between ground-based measurements and the last 2 years of ICESat data.

~~These include studies from the and the Norway USA East Antarctic Traverse (Kohler et al., 2013).~~

Here we present results from the first 88S Traverse and show that 1) this part of Antarctica is ideal for this type of airborne and spaceborne data validation and 2) the surface elevation is probably changing minimally, with respect to ice flow, snow accumulation, and surface melt, making it an ideal absolute elevation validation surface, but that there is some level of snow redistribution (sastrugi migration) necessitating near coincident airborne surveys in space and in time to improve estimates of surface measurement precision.

## 2 Data

### 2.1 88S Traverse GPS data

We conducted a 750 km kinematic GPS survey near Amundsen-Scott South Pole Station, Antarctica using 2 tracked vehicles (PistenBullys) provided by the US Antarctic Program. The 88S Traverse departed from South Pole Station on 28 December 2017 and traveled for 4 days to the 88° S line of latitude. The traverse route then followed this line of latitude for ~300 km, before returning to South Pole Station on 10 January 2018 (Fig. 1). The kinematic GPS survey used dual-frequency Trimble NetR9 receivers recording at 1 and 2 Hz with Trimble Zephyr 2 Geodetic GNSS (TRM57971) antennas, mounted to the roof of each PistenBully. The GPS units collected data during the day; they were powered down in the evenings based on operational constraints (these included charging the batteries and the fact that the satellite phones that we used in the evenings interfered with the GPS receivers, a problem that will be rectified in subsequent surveys). Some opportunistic static GPS data were collected during routine breaks throughout the day.

The height of each roof-mounted GPS antenna was measured twice along the 88S Traverse; specifically, the measurement made was the distance between the antenna base plane and the bottom of the indentation of the tracks of the PistenBully into the snow (Fig. 2). The average antenna heights for the 2 vehicles were 281.3 cm (vehicle A, ~~1 $\sigma$  standard deviation 0.9 cm~~ 280.7 cm and 281.9 cm) and 282.3 cm (vehicle B, 282.6 cm and 281.9 cm ~~1 $\sigma$  standard deviation 0.4 cm~~). The depths of the tracks of each of the vehicles into the snow surface were measured 30 times along the traverse. The average track depths for the 2 vehicles were 6.2 cm (vehicle A, 1 $\sigma$  standard deviation 1.6 cm) and 5.8 cm (vehicle B, 1 $\sigma$  standard deviation 1.2 cm). The antenna-height and track-depth measurements are ultimately required to calculate the distance from each of the GPS antenna phase centers to the snow surface (Fig. 2).

Surveys were conducted at  $\sim 2 \text{ m s}^{-1}$ ; at a 2 Hz sampling rate, this generated data points with non-uniform footprint spacing of  $\sim 1 \text{ m}$ . The leading PistenBully set the traverse route and the trailing ( $\sim 50 \text{ m}$ ) PistenBully routinely crossed the tracks of the leading vehicle to create statistical crossover points within the data (Fig. 3). GPS unit A was always on the trailing PistenBully, behind GPS unit B.

## 2.2 UAF Lidar

The University of Alaska, Fairbanks (UAF) Lidar is a [line-scanner](#) laser altimeter that has typically been deployed during Alaska-based OIB campaigns (Johnson et al., 2013). The UAF Lidar surveyed the 88S Traverse on 2 separate flights (30 November and 3 December, 2017) while integrated in a commercial (Airtec) BT-67 (Basler). The UAF system is a commercial RIEGL LMS-Q240i scanning laser altimeter transmitting in the 905 nm wavelength part of the spectrum. The system has a full scanning angle of 60°. The 2 surveys over the 88S Traverse were conducted at an aircraft speed of  $\sim 85 \text{ m s}^{-1}$ , at an altitude of  $\sim 450 \text{ m AGL}$  (above ground level). At this speed and altitude, and with an effective repetition rate of 10 kHz, the UAF Lidar generates a  $\sim 1.3 \text{ m}$  diameter footprint with a total across-track swath-width of  $\sim 500 \text{ m}$ . Within a 10 m by 10 m area, the UAF Lidar produces  $\sim 20$  to 25 returns, with nearly uniform footprint spacing of  $\sim 2 \text{ m}$  (Fig. 3).

## 2.3 Airborne Topographic Mapper (ATM)

The Airborne Topographic Mapper (ATM; Krabill et al., 2002; [Martin et al., 2012](#)) is a laser altimetry system used by many OIB campaigns in both the Arctic and Antarctic. ATM collected data along the 88S Traverse on 26 October 2014, while integrated on the NASA DC-8. For that deployment, ATM (version T4) consisted of a dual instrument configuration, with both wide-scan and narrow-scan lidar systems integrated simultaneously. The wide-scan lidar system is more appropriate for ice sheet surveys and has a full scanning angle of 30°. The ATM lidars are full-waveform conically-scanning system, transmitting 532 nm wavelength 6 ns pulses. Surveys were conducted at an aircraft speed of  $\sim 100 \text{ m s}^{-1}$ , at an altitude of  $\sim 450 \text{ m AGL}$ . At this speed and altitude, and with a 3 or 5 kHz repetition rate, the wide-scan (30°) ATM lidar generates a  $\sim 1 \text{ m}$  diameter footprint with a scanning swath width of  $\sim 250 \text{ m}$ . Within a 10 m by 10 m area, the wide-scan ATM produces  $\sim 6$  to 8 returns, with non-uniform footprint spacing of  $\sim 5 \text{ m}$ ; data are most dense along the edge of the swath (Fig. 3).

For completeness, we note that ATM also conducted a mission that included the 88S Traverse on 26 October and 15 November 2016 (also integrated on the NASA DC-8 [and flying at  \$\sim 450 \text{ m AGL}\$](#) ) using the T6 version of ATM. However, analysis of these data and other flights during this campaign suggest that there is an across-track tilt within these data, that represented a 10 to 15 cm spurious elevation variation across the wide-scan ATM swath ([personal communication, Michael Studinger, NASA](#)). We therefore exclude the 2016 ATM flights from further discussion.

## 3 Methods

### 3.1 88S Traverse GPS data

Following the data processing methods of Brunt et al. (2017), we post-processed 88S Traverse GPS data using PPP methods. PPP solutions use precise GPS satellite orbit and clock information to determine the kinematic GPS antenna position. Position solutions for each vehicle were determined using NovAtel's Inertial Explorer (v.8.6); processing for each vehicle was done on nearly continuous stretches of GPS data, which typically represented 1 full day of driving, or approximately 50

km. Position solutions were solved to the L1 phase center of each antenna; ~~and the elevations are given in reference to the ITRF08 reference frame of and the geographic coordinates are referenced to the~~ WGS84 ellipsoid. ~~We used a GPS satellite elevation mask, or a cut-off angle, of 7.5 degrees, to minimize the effects of multipath.~~ Inertial Explorer provides an estimate of a given point-position vertical accuracy; this value was used to filter suspect elevation data that had a vertical sigma of more than 8 cm.

The elevation of the snow surface, relative to the position solutions of the L1 phase center of each antenna (Fig. 2), was then determined using data from the field and the appropriate National Geodetic Survey (NGS) antenna model phase-center offset. The height of the snow surface ( $h$ ) for each vehicle was determined based on the position solutions of the GPS antenna phase centers ( $GPS_{PC}$ ) based on the following equation:

$$h = GPS_{PC} - h_{AntHeight} - h_{NGSmodel} + h_{TrackDepth} , \quad (1)$$

where  $h_{AntHeight}$  is the mean distance between the antenna base plane and the indentation of the tracks in the snow (281.3 or 282.3 cm, depending on the vehicle),  $h_{NGSmodel}$  is the distance between the antenna phase center and the base plane based on the NGS model for the Trimble Zephyr 2 Geodetic antenna (4.1 cm), and  $h_{TrackDepth}$  is the mean depth of the PistenBully track indentations into the snow surface (6.2 or 5.8 cm, depending on the vehicle).

### 3.2 Airborne lidar data

We obtained the UAF Lidar Scanner L1B Geolocated Surface Elevation Triplets, Version 1 data (Larsen, 2010) through the National Snow and Ice Data Center (NSIDC) OIB Data Portal (<http://nsidc.org/icebridge/portal/>) for the 30 November 2017 and 3 December 2017 flights over the 88S Traverse area (~~files available at NSIDC are associated with whole Julian days, or days 334 and 337~~). The data files consist of latitudes, longitudes, and elevations that were derived from an integrated on-board GPS (Trimble) and inertial system (OxTS Inertial+2). GPS post-processing used PPP methods using Novatel's GrafNav (v.8.4). Processing of the lidar data, including the incorporation of the GPS and inertial data used a commercial software package (RiProcess) developed by RIEGL. These data are distributed with the elevations given in the ITRF08 reference frame and the geographic coordinates referenced to the WGS84 ellipsoid~~in the ITRF08 reference frame of the WGS84 ellipsoid.~~

We obtained the ATM IceBridge ATM L1B Elevation and Return Strength with Waveforms, Version 1 data (Studinger, 2018) also through the NSIDC for the 26 October 2014 flight over the 88S Traverse area (17:04 to 19:45 UTC). The data files include ~~latitudes, longitudes, geographic coordinates,~~ and elevations derived from an integrated on-board GPS (Javad) and inertial system (Applanix POS AV). Differential GPS (DGPS) post-processing methods, which use a base station installed at the departure airport for this deployment. DGPS was accomplished using a software package developed by the ATM team called GITAR (GPS Inferred Trajectories for Aircraft and Rockets; Martin, 1991). These data are distributed with the elevations given in the ITRF08 reference frame and the geographic coordinates referenced to the WGS84 ellipsoid~~in the ITRF08 reference frame of the WGS84 ellipsoid.~~

### 3.3 Comparison strategy

We based our comparison strategy on Brunt et al. (2017). We compared the post-processed snow surface elevations from the 88S Traverse with the airborne surface elevation data, using a ‘nearest-neighbor’ approach. In this method, we compared the closest lidar data point to every single ground-based GPS data point. We limited our statistical analysis based on a distance criterion, making elevation comparisons only where the lidar footprints and the GPS measurements were within a distance 1 m of one another. We then assessed the difference between the filtered GPS and ATM and UAF lidar surface elevation datasets.

Once the lidar elevation data ( $Lidar_{elevation}$ ) were associated with the GPS elevation data ( $GPS_{elevation}$ ), the mean elevation difference ( $GPS_{elevation} - Lidar_{elevation}$ ) is the lidar elevation bias ( $B$ ). We note that we take the GPS elevation data to be the ground truth.

The  $1\sigma$  standard deviation of this airborne lidar elevation bias ( $B$ ) is the spread of the lidar data, or the precision, about the mean. This is also the vertical dispersion of the lidar measurements about the mean surface. The vertical dispersion, or the surface measurement precision, includes both instrument precision and geophysical properties of the surface that will affect the measurement. Instrument precision is related to factors such as instrument timing errors, geolocation knowledge, and footprint size. Geophysical properties that will affect the measurement include atmospheric effects, surface roughness, and surface slope, although we note that our analysis is limited to a region of low (less than 1 degree) surface slope. These instrument and geophysical effects cannot be uniquely distinguished within the surface measurement precision. Ultimately, we report elevation accuracies and surface measurement precisions as a residual, following the convention of mean bias  $\pm 1\sigma$  standard deviation, or  $0.0 \pm 0.0$  cm.

The lidar biases and precisions reported here are determined relative to the GPS data, which we take represent truth, with zero errors. In actuality, these errors are not zero and are a function of: 1) formal GPS errors, which include factors such as ephemeris and clock errors; 2) ionosphere and troposphere errors; 3) multipath errors; and 4) errors due to geophysical effects, such as variable snow surface strength causing variable vehicle sinking or antenna motion due to short-scale surface undulations (sastrugi). We note that given the short distance between the 2 survey vehicles, our results are somewhat blind to the full magnitude of the error terms that can be correlated on short time scales, such as those associated with the ionosphere and troposphere.

## 4 Results

### 4.1 Ground-based GPS data evaluation

We compared the GPS position solutions of each vehicle to assess consistency of the ground-based data. After the 88S Traverse GPS data for each vehicle were post-processed, the data were then filtered based on the 8 cm vertical sigma; this reduced each GPS data set by about a third (GPS unit A: 316,948 data points were reduced to 203,603; GPS unit B: 321,689

data points were reduced to 209,253). The mean vertical sigma values for the data used in further analysis were 7.16 and 7.19 cm for ground-based GPS units A and B, respectively. We then used a nearest-neighbor approach, limited based on a 0.5 m distance criterion, and calculated the mean elevation residual between the elevation measured by the two vehicles. This residual was  $1.1 \pm 4.1$  cm ( $n=26,442$ ).

5 PPP GPS post-processing methods are often used in regions where long-term base-station data are not available for DGPS methods, such as the center of ice sheets. Brunt et al (2017) showed that PPP position solutions for their traverse outside of Summit Station, in the center of the Greenland Ice Sheet, were comparable to GPS position solutions using differential methods. Therefore, while we are limited with respect to the availability of permanent GPS base stations for post processing, we feel confident that our methods provide consistent and accurate results and are appropriate for this data analysis.

#### 10 4.2 Airborne lidar evaluation

To assess the internal consistency of the UAF Lidar, we compared the processed elevation data from the 30 November 2017 flight to the 3 December 2017 flight, using a nearest-neighbor approach, limited based on a 1 m distance criteria, and calculated the mean elevation residual. This residual was  $8.1 \pm 10.5$  cm ( $n>1.5$  million data points); data from the 30 November 2017 flight were lower than data from the 3 December 2017 flight. A similar assessment of internal consistency of the ATM data could not be made since our analysis was limited to a single flight, after rejecting the 2016 ATM data due to an observed across-track tilt.

#### 4.3 GPS to airborne lidar results

Table 1 lists the results for the nearest-neighbor analysis of the ground-based GPS and lidar elevation comparisons for both ATM and the UAF Lidar. Both altimeters had elevation biases less than 10 cm and surface measurement precisions less than 15 cm; we note that these values are similar to those in Brunt et al. (2017), which is a similar study in a similar geophysical setting. Figure 4, panel A, shows the elevations of ground-based GPS unit A. Panel B shows the difference between GPS A elevations and the 30 November UAF Lidar elevations, minus the mean difference. Panel C is similar to panel B but using the 3 December UAF Lidar data; and panel D compares the GPS data to the 2014 ATM data. Figure 5 is the same as Figure 4, but the results are relative to ground-based GPS unit B.

25 We examined the spatial correlation of the elevation differences calculated between the ground-based GPS data and the airborne lidar data (Motyka et al., 2010; Rolstad et al., 2009). When measurements are made within close spatial proximity of one another, they are generally similar, and measurement errors tend to be correlated; over increasing distances, measurement errors become uncorrelated. Similar to Rolstad et al. (2009), who proved which is a detailed summary of semivariograms, we created semivariograms of the elevation differences, This analysis is intended to which provide an assessment of the length scales where measurement errors become independent of one another, or uncorrelated. Figures 6 and 7 provide the semivariograms for GPS unit A and B, respectively, relative to the ATM flight (top panels) and the UAF Lidar flights (bottom-top two panels in each figure) and the ATM flight (bottom panels in each figure). The x axes are lag

distances between the observations, in m, and the y axes are the measure of variance, in m<sup>2</sup>. The red squares represent the observed elevation differences in 50 m bins and the lines represent a semivariogram model fit to these data. The range and the sill of the variograms are interpreted to be where the slope of the model fit to the variance asymptotes toward zero, which is indicated where the lines in the figures change from blue to red. At this distance, or at the range value, the observations are considered to have become independent; from Figures 6 and 7, we estimate that the range at which the variance starts to be relatively unchanging, and the length scale at which measurement errors become uncorrelated, to be ~10 to 50 m. These results are based on 5 km of along-track data; semivariograms based on longer length scales (20 km) had similar results. We attribute this 10 to 50 m length scale to be associated with wind-driven surface processes and overall roughness (sastrugi), as visible in the background of Figure 3. Sastrugi causes noise about the mean surface elevation from a measurement perspective and we assume that this is the largest source of correlated error, given the size of the footprints of the observations (1 to 2 m), the distance criteria associated with the differencing methods (1 m), and the length scale of the surface roughness associated with sastrugi (5 to 10 m).

## 5 Discussion

The 1 $\sigma$  mean elevation residual between the 2 GPS units for this study was  $1.1 \pm 4.1$  cm (n=26,442), with GPS A generally being higher than GPS B. This residual compares favorably to the GPS assessments of Brunt et al. (2017) and Kohler et al. (2013), the studies that most closely match the methods and geophysical setting presented here. Brunt et al. (2017) reported a 1 $\sigma$  mean elevation residual of  $0.7 \pm 5.7$  cm, based on comparisons between 2 different passes of the traverse occurring on the same day and using the same GPS unit (n=710). Kohler et al. (2013) reported a 1 $\sigma$  mean elevation residual of  $0.6 \pm 7.5$  cm, based on crossovers between 2 different GPS units during the traverses (n=1131). We attribute the quality of our GPS data to 1) the long length scale of data collection (relative to Brunt et al., 2017) and 2) the flat surface that defined our traverse route (relative to Kohler et al., 2013). ~~We note from Figures 4 and 5 that no discernable signal is evident in the various comparisons and we therefore attribute the differences to surface measurement noise.~~

While the residual between the 88S Traverse vehicles is low, it is not zero. We attribute the ~1 cm bias between our GPS datasets to uncertainties in the measurements of track depth. From Equation 1 and Figure 2, the 3 terms associated with reducing the GPS measurement to a snow-surface height are the phase center offset (which is static and common between the vehicles), the antenna height (vehicle A  $281.3 \pm 0.9$  cm; vehicle B:  $282.3 \pm 0.4$  cm), and the track depth (vehicle A:  $6.2 \pm 1.6$  cm; vehicle B:  $5.8 \pm 1.2$  cm). Given the uncertainties associated with the 2 field-based measurements (antenna height and track depth), we feel confident that the snow depth is the leading term in the height uncertainty. As stated above, we note that we are blind to errors introduced by the close spatial coincidence of the GPS receivers (~50 m) and to those introduced by the common processing of the GPS data. Errors in ionospheric or tropospheric modeling would impact both GPS-based data sets similarly and would introduce a bias between the GPS measurements and the actual ice sheet surface.

Overall, the quality of the lidar data used in this survey was quite good. While a quantitative assessment could be made for the UAF Lidar, a similar assessment of ATM could not be made in this region, as we were limited to one flight. However, Brunt et al. (2017) analyzed ATM data from 5 different airborne campaigns, which included 5 different versions of the ATM system (including both narrow and wide scanning data) near Summit Station, Greenland, on the relatively flat ice sheet interior, similar to this study. Their results indicated an average ATM elevation bias and surface measurement precision of  $-3.4 \pm 6.5$  cm (based on PPP post-processing, which is the method used here). These results match well with those of Martin et al. (2012), who summarize the vertical accuracy and precision of ATM over ice sheets to be  $6.6 \pm 3$  cm. Given that we are using the same lidar, with similar survey techniques, over a similar surface, we consider ATM to be a stable instrument, with data quality suitable for this application.

We note that there is a slight along-flight signature that is apparent in the UAF Lidar elevation data (Fig. 8). The signature is visible in the southern side of the swaths of both the 30 November and 3 December 2017 datasets. Specifically, there appears to be a trough along the southern edge of the swaths that has anomalously lower elevations, relative to the surrounding edges. The magnitude is variable but based on a nearest-neighbor assessment of the overlapping region in Figure 8, where the flight line from 30 November 2017 intersected itself, the mean residual was  $-9.9 \pm 12.7$  cm. While the source of this artefact is still undetermined, it doesn't appear to be an across-track tilt. This effect on measured elevation is small ( $\sim 5$  cm scale from edge of the trough to the base of the trough) and generally limited to near the edge of the lidar swath (Fig. 8). These data were typically not used for ground survey GPS comparison, as the ground-based data generally intersected the center of the swath, where we believe the data quality is not compromised.

The elevation biases and surface measurement precisions of the 2 OIB lidars presented here are comparable to that of the OIB lidars assessed in Brunt et al. (2017); results based on PPP methods for both studies indicated biases that are less than  $\sim 11$  cm and measurement precisions that are less than  $\sim 15$  cm (Table 1 in this document and Table 2 in Brunt et al., 2017). Brunt et al. (2017) also indicate an average ATM elevation bias and surface measurement precision of  $-3.4 \pm 6.5$  cm. From Table 1, the surface measurement precision associated with ATM over the 88S Traverse ( $\pm 14$  cm; 'lower precision') was poorer quality than the average precision of ATM over the Summit Station Traverse ( $\pm 7$  cm; 'higher precision') as determined by Brunt et al. (2017). These 2 assessments had a similar geophysical setting (i.e., ice-sheet interior) and similar survey strategies (GPS collection and processing methods).

We attribute the poorer surface measurement precision to the time difference between the airborne ATM campaign (October 2014) and the ground-based GPS survey (December 2017 to January 2018). Specifically, we hypothesize that these differences were associated with the transient locations of sastrugi. To assess this hypothesis, we used the same nearest-neighbor approach, described in the methods section, to compare the 2014 ATM elevation data to the 2017 UAF Lidar elevation data (Table 2). Ultimately, the difference between these 2 lidar datasets revealed a signature that was of a similar magnitude (meters) and trend (grid SSE, or  $\sim 150^\circ$ ) as the sastrugi, based on observations of the sub-meter-resolution WorldView-2 satellite imagery, obtained via the Polar Geospatial Center at the University of Minnesota (Fig. 9).

Sastrugi causes noise about the mean surface elevation from a measurement perspective. Sastrugi migration between the 2014 ATM campaign and the 2017/2018 ground-based traverse would not have an impact on the surface elevation bias, as the observed differences would be averaged out and lost in surface measurement noise. The migration of the sastrugi adds components of noise on the mean surface measurement. This effect is evident in the observed larger (poorer) ATM surface measurement precision assessment.

We note that our analysis does not attempt to account for elevation changes due to the temperature- and accumulation-rate-driven effects of firm compaction (Li and Zwally, 2015). In this region, we expect variation in firm compaction rate to be driven by changes in firm temperature, which have a large seasonal amplitude and a much smaller secular trend. As the firm warms each austral spring, the surface elevation along our traverse should decrease. Since the UAF lidar data and ground-based GPS data were collected within a month, we expect firm compaction to have a negligible effect on our results. Conversely, the ~2 month seasonal lag between the ATM and GPS data collection means that we may be sensitive to the seasonality of firm compaction rate, as well as any secular trend over the 4 year interval between these data sets.

Overall, these results suggests that the 88S Traverse route is an ideal setting to assess airborne or satellite absolute elevation accuracy (Brunt et al., 2017), as the surface was relatively unchanged between 2014 and 2018 (i.e., no distinguishable change in bias). Further, our results based on the 2014 ATM elevation dataset suggests that airborne data collected along this route are applicable to absolute elevation validation for a few years. However, results based on our comparisons between our GPS measurements and ATM suggest that when a few years has evolved between the datasets being evaluated, the surface elevation measurements become hard to reproduce; this manifests itself in a higher surface measurement precision assessment.

Data collected from the 88S Traverse (and data collected on subsequent surveys of the same route) will provide 300 km of *in situ* data for direct comparison with ICESat-2 elevation data products. The GPS data collection strategies and post-processing methods presented here provide accurate and precise data for such an assessment. Further, the data analysis presented here provides guidance on how to make similar comparisons between ground-based and satellite elevations, given the satellite footprint size and associated rejection criteria. Approximately 3 to 4 ICESat-2 reference ground tracks will intersect this region daily to produce many statistical crossover points between the GPS and ICESat-2 datasets. While the crossover points represent only a small segment of along-track ICESat-2 data, the analysis will be based on data from many ICESat-2 reference ground tracks over the course of the entire satellite mission. Thus, the analysis of the derived ICESat-2 bias and surface measurement precision relative to these GPS data will provide an assessment ICESat-2 performance through time, independent of errors associated with single orbits or single points in time. Results of Brunt et al. (2017) and results presented here also provide an assessment of the accuracy and surface measurement precision of 3 airborne lidars that NASA has routinely deployed over the ice sheets (ATM, LVIS, and the UAF lidar). With a statistical understanding of how these instruments perform on the relatively flat ice-sheet interiors, longer flight lines can be constructed over similarly flat ice-sheet surfaces to create better statistics associated with comparisons using long length scales of along-track ICESat-2 data. In summary, the strategic location of the ground-based 88S Traverse provides a validation of ICESat-2 that is independent of

the errors that are correlated with respect to most satellite time scales, and these ground-based data provide a better understanding of airborne lidars that will survey longer length scales of data, for better satellite error statistics.

## 6 Conclusions

Here we present a comparison of in situ GPS elevation data and laser altimetry in preparation for ground-based and airborne validation of ICESat-2. We show that the ground-based methods for GPS data collection and processing along the 88S Traverse provide internally consistent results, with accuracies and precisions appropriate for assessing airborne lidar data and ultimately, satellite elevation data. Further, we have shown that airborne lidar data assessed here (ATM and the UAF Lidar), relative to the GPS data, show elevation biases that are comparable to results from similar instruments in similar geophysical settings. However, discrepancies between the ATM surface measurement precisions observed here, and those observed in Brunt et al. (2017) under similar ice-sheet interior conditions, suggest that the migration of sastrugi can have an adverse effect on assessments of surface measurement precision when significant time (on the order of a few years) has elapsed between surveys. Thus, absolute elevation bias can be determined with datasets from this surface that are a few seasons old, but for the best assessment of precision, comparisons need to be made with relatively coincident (spatial and temporal) datasets.

## 7 Data availability

The ground-based GPS data associated with this study are available online, as the supplement related to this article (doi:10.5194/tc-~~NNN160~~-supplement). NASA ATM and the UAF Lidar data are publicly available on the NSIDC Operation IceBridge Data Portal (<http://nsidc.org/icebridge/portal/>). WorldView-2 imagery is available to NSF- and NASA-funded researchers via the Polar Geospatial Center at the University of Minnesota.

Formatted: Font color: Text 1

## Acknowledgements

We thank the NASA ICESat-2 Project Science Office for funding the field component and data analysis associated with this project. We thank the National Science Foundation, Office of Polar Programs for logistical support of the field component of this project. We thank Operation IceBridge for the data collection of the ATM and UAF Lidar datasets. We thank our deep-field mechanic and mountaineer associated with the 88S Traverse (Chad Seay and Forrest McCarthy), for ensuring that we safely completed the full Antarctic ground survey. We thank the many science support staff of the US Antarctic Program that helped make the field component of this project possible. ~~And we~~ We thank the National Snow and Ice Data Center (NSIDC) for IceBridge data distribution. Finally, we thank our editor (Kenny Matsuoka) and two anonymous reviewers for

[constructive comments to earlier drafts of this manuscript](#). WorldView-2 imagery was provided by the Polar Geospatial Center at the University of Minnesota, which is supported by grant ANT-1043681 from the National Science Foundation.

## References

- Bindschadler, R., Vornberger, P., Fleming, A., Fox, A., Mullins, J., Binnie, D., Paulsen, S., Granneman, B., and Gorodetzky, D.: The Landsat image mosaic of Antarctica, *Remote Sens. Environ.*, 112, 4214–4226, doi:10.1016/j.rse.2008.07.006, 2008.
- 5 Brunt, K., Hawley, R., Lutz, E., Studinger, M., Sonntag, J., Hofton, M., Andrews, L., and Neumann, T.: Assessment of NASA airborne laser altimetry data using ground-based GPS data near Summit Station, Greenland, *The Cryosphere*, 11, 681–692, doi:10.5194/tc-11-681-2017, 2017.
- Fricker, H., Borsa, A., Minster, B., Carabajal, C., Quinn, K., and Bills, B.: Assessment of ICESat performance at the salar de Uyuni, Bolivia, *Geophys. Res. Lett.*, 32, doi:10.1029/2005GL023423, 2005.
- 10 Johnson, A., Larsen, C., Murphy, N., Arendt, A., and Zirnheld, S.: Mass balance in the Glacier Bay area of Alaska, USA, and British Columbia, Canada, 1995–2011, using airborne laser altimetry *J. Glaciol.*, 59, 632–648, doi:10.3189/2013JoG12J101, 2013.
- Kamb, B., and Echelmeyer, K.: Stress-Gradient Coupling in Glacier Flow: I. Longitudinal Averaging of the Influence of Ice Thickness and Surface Slope, *J. Glaciol.*, 32, 267–284. doi:10.3189/S0022143000015604, 1986.
- 15 Koenig, L., Martin, S., Studinger, M., and Sonntag, J.: Polar airborne observations fill gap in satellite data, *EOS Trans. AGU*, 91, 333–334, doi:10.1029/2010EO380002, 2010.
- Kohler, J., Neumann, T., Robbins, J., Tronstad, S., and Melland, G.: ICESat elevations in Antarctica along the 2007–09 Norway–USA traverse: Validation with ground-based GPS, *IEEE T. Geosci. Remote*, 51, 1578–1587, doi:10.1109/TGRS.2012.2207963, 2013.
- 20 Krabill, W., Abdalati, W., Frederick, E., Manizade, S., Martin, C., Sonntag, J., Swift, R., Thomas, R. and Yungel, J.: Aircraft laser altimetry measurement of elevation changes of the Greenland ice sheet: Technique and accuracy assessment, *J. Geodyn.*, 34, 357–376, doi:10.1016/S0264-3707(02)00040-6, 2002.
- Larsen, C.: IceBridge UAF Lidar Scanner L1B Geolocated Surface Elevation Triplets, NASA NSIDC DAAC, Boulder, Colo, doi:10.5067/AATE4JJ91EHC, 2010.
- 25 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., Schutz, R., Smith, B., Yang, Y., [and](#) Zwally, H.: The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation, *Remote Sens. Environ.*, 190, 260–273, doi:10.1016/j.rse.2016.12.029, 2017.
- 30 Martin, C.: GITAR Program Documentation, NASA contract #NAS5-31558 program document, Goddard Space Flight Center, Wallops Flight Facility, Wallops Island, VA. 1991.

[Martin, C., Krabill, W., Manizade, S. and Russell, R.: Airborne Topographic Mapper Calibration Procedures and Accuracy Assessment. NASA Technical Memorandum 2012-215891, 2012.](#)

Motyka, R., Fahnestock, M., and Truffer, M.: Volume change of Jakobshavn Isbræ, West Greenland: 1985–1997–2007, *J. Glaciol.*, 56, 635–646, doi: 10.3189/002214310793146304, 2010.

- 5 Rolstad, C., Haug, T., & Denby, B.: Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: Application to the western Svartisen ice cap, Norway, *J. Glaciol.*, 55, 666–680. doi:10.3189/002214309789470950, 2009.

Siegfried, M., Hawley, R., and Burkhart, J.: High-resolution ground-based GPS measurements show intercampaign bias in ICESat elevation data near Summit, Greenland, *Geoscience and Remote Sensing, IEEE Transactions on*, 49, 3393–3400, doi:10.1109/TGRS.2011.2127483, 2011.

10 Studinger, M.: IceBridge ATM L1B Elevation and Return Strength with Waveforms, Version 1. NASA NSIDC DAAC, Boulder, Colorado USA, doi:10.5067/EZQ5U3R3XWBS, 2018.

Webb, C., Zwally, H., and Abdalati, W.: The Ice, Cloud, and land Elevation Satellite (ICESat) Summary Mission Timeline and Performance Relative to Pre-Launch Mission Success Criteria. NASA Technical Report, NASA/TM-2013-217512, 15 2012.

Zwally, H., Li, J., Brenner, A., Beckley, M., Cornejo, H., DiMarzio, J., Giovinetto, M., Neumann, T., Robbins, J., Saba, J., Yi, D., and Wang, W.: Greenland ice sheet mass balance: distribution of increased mass loss with climate warming; 2003–07 versus 1992–2002, *J. Glaciol.*, 57, 88–102, doi:10.3189/002214311795306682, 2011.

**Table 1: Elevation bias and surface measurement precision (in cm), relative to ground-based GPS survey data, for ATM and UAF airborne lidar elevation data. Results are posted as  $GPS_{elevation} - Lidar_{elevation}$ .**

Lidar Survey	PPP bias $\pm$ precision:
	Relative to GPS A (cm) Relative to GPS B (cm)
ATM 26 October 2014	2.8 $\pm$ 14.0
	3.6 $\pm$ 14.1
UAF Lidar 30 November 2017	0.1 $\pm$ 9.7
	0.2 $\pm$ 9.5
UAF Lidar 3 December 2017	-9.5 $\pm$ 9.8
	-8.0 $\pm$ 9.7

5 **Table 2: Elevation bias and surface measurement precision (in cm), between ATM and the UAF lidar. Results are posted as  $ATM_{elevation} - UAF_{elevation}$ .**

Lidar Surveys	mean bias $\pm$ 1 $\sigma$ standard deviation, cm
ATM 26 October 2014 vs UAF Lidar 30 November 2017	0.3 $\pm$ 15.0
ATM 26 October 2014 vs UAF Lidar 3 December 2017	-7.8 $\pm$ 15.1

**Figure 1: Map of the 88S Traverse Route, color coded based on elevation. Locations for Figures 4 – 9 are also shown. The South Pole Operational Traverse (SPoT) Route is indicated in orange. Background is the Landsat image mosaic of Antarctica (LIMA; Bindschadler et al., 2008).**

5 **Figure 2: The GPS antenna configuration on a PistenBully.  $GPS_{PC}$  is the surveyed position solution to the phase center of the antenna,  $h_{NGSmodel}$  is the NGS model distance between the antenna phase center and the antenna base plane,  $h_{AntHeight}$  is the distance between the antenna base plane and the indentation of the tracks in the snow,  $h_{TrackDepth}$  is the depth of the sled runners in the snow surface, and  $h$  is the snow surface (Eq. 1).**

Figure 3: Sample footprint spacing for the UAF Lidar (dark blue), ATM (cyan), and the 88S Traverse ground-based GPS data are in shades of red (~~points in shades of red~~ GPS A is in light red, while GPS B is in dark red). WorldView-2 imagery, copyright 2017, DigitalGlobe, Inc.

5

5 **Figure 4: Along-track elevation and elevation differences associated with GPS A. Panel A: Along-track elevation of GPS A, in m. Panel B: Elevation difference between GPS A and the UAF Lidar (30 November 2017), minus the mean difference. Panel C: Elevation difference between GPS A and the UAF Lidar (3 December 2017), minus the mean difference. And panel D: Elevation difference between GPS A and ATM (26 October 2014), minus the mean difference.**

5 **Figure 5: Along-track elevation and elevation differences associated with GPS B. Panel A: Along-track elevation of GPS B, in m. Panel B: Elevation difference between GPS B and the UAF Lidar (30 November 2017), minus the mean difference. Panel C: Elevation difference between GPS B and the UAF Lidar (3 December 2017), minus the mean difference. And panel D: Elevation difference between GPS B and ATM (26 October 2014), minus the mean difference.**

5

Figure 6: Semivariograms of elevation differences between GPS unit A and elevations derived from ATM (top panel) and the UAF Lidar on 30 November 2017 (middle panel) and 3 December 2017 (bottom panel). The x axes are lag distances between the observations, in m, and the y axes are ~~the measure of~~ variance, in  $m^2$ . The red squares ~~represent are~~ the observed elevation differences in 50 m bins and the lines represent a semivariogram model fit to these data. ~~As the slope of the model fit to the variance asymptotes toward zero, where the lines in the figures change from blue to red, the observations are considered to have become independent.~~

5 | Figure 7: Semivariograms of elevation differences between GPS unit B and elevations derived from ATM (top panel) and the UAF Lidar on 30 November 2017 (middle panel) and 3 December 2017 (bottom panel). The x axes are lag distances between the observations, in m, and the y axes are ~~the measure of variance~~, in  $m^2$ . The red squares ~~represent are~~ the observed elevation differences in 50 m bins and the lines represent a semivariogram model fit to these data. ~~As the slope of the model fit to the variance asymptotes toward zero, where the lines in the figures change from blue to red, the observations are considered to have become independent.~~

Formatted: Normal

Figure 8: Elevation data from the UAF Lidar (30 November 2017), where the flight line crossed itself. The along-track artifact in the data is visible in both passes; UAF Lidar elevations are anomalously lower within the artefact and manifest as a narrow trough, parallel to the direction of flight. 88S Traverse ground-based GPS data are in shades of red. WorldView-2 imagery, copyright 2017, DigitalGlobe, Inc.

5

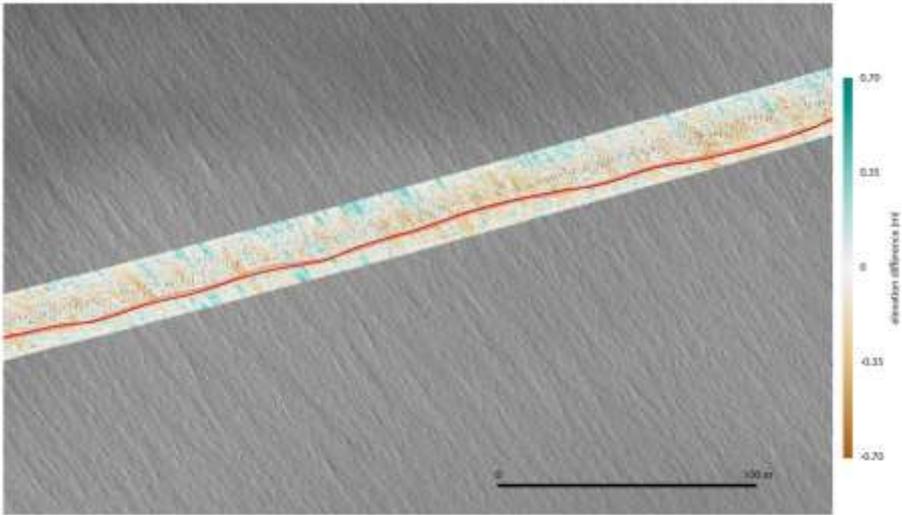


Figure 9: Ground based GPS data (in shades of red) plotted on difference in elevation between ATM (26 October 2014) and the UAF Lidar (30 November 2017). WorldView-2 imagery, copyright 2017, DigitalGlobe, Inc.