March 26, 2025

Manuscript: "Analyzing vegetation effects on snow depth variability in Alaska's boreal forests with airborne lidar," by Lora D. May, Svetlana Stuefer, Scott Goddard, Christopher Larsen

Responses to Reviewer #1

We want to thank the reviewer for their time and effort reading our manuscript and providing comprehensive and constructive comments. The reviewer's valuable insight and suggestions have enhanced the manuscript for clarity, depth, and overall quality. We are grateful for your contribution to refining our work. The manuscript has been revised and updated based on the reviewer's comments and suggestions. The following pages contain comments that appear exactly as they were received. Our responses are inserted next to each comment in blue text. Italicized sentences are directly from the edited manuscript.

Specific Comments

The manuscript presents analysis of snow depth measuments made with airborne lidar in Alaska in boreal forest region. Lidar measurements were compared with in-situ snow depth measurements. Snow depth was estimated for three different canopy height classes. The study resulted statistically significant differences in snow depth between forest and lower vegetation.

General comments:

The manuscript is well prepared and easy to read. Figures and tables support the text well. Title is clear and abstract gives good summary for the manuscript. I have some general questions:

- How SWE values are calculated? Have you used density measurements? Please describe it with more details.

Response 1: Thank you for this comment. We clarified SWE value calculations by revising the following sentences.

In methods: *SWE values were calculated using the end of winter snow density statistic for the boreal forest climate class (217 kg m^{-3}) as reported in Sturm and Liston (2021).*

- Could you estimate how large area was covered with the magnaprobe measurements?

Response 2: We estimated the area covered with the magnaprobe measurements to be approximately 0.4 km^2 and revised the sentence in question to now state the area.

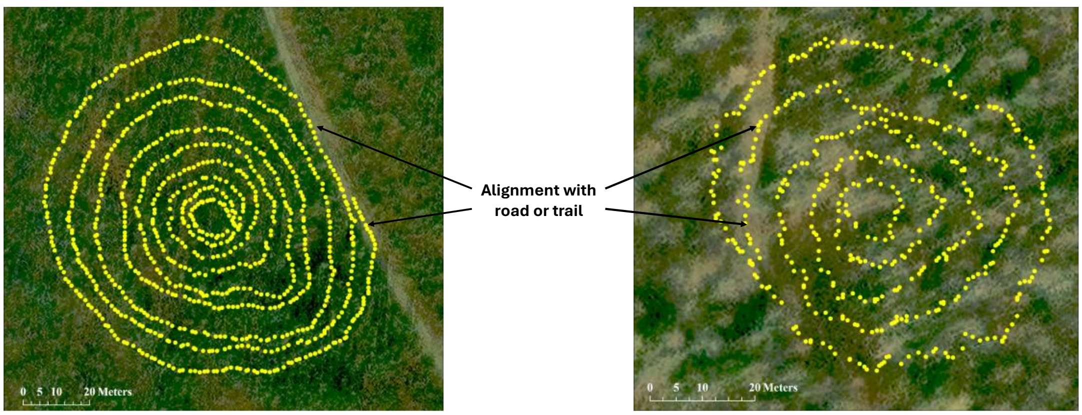
In results: *Ground-based snow depth measurements cover approximately 0.4 km^2 of the watershed area.*

- Have you considered accuracy of magnaprobe GPS in the analysis?

Response 3: Thank you for addressing this topic. The magnaprobe uses a Garmin GPS16x-HVS which has a horizontal accuracy of ± 2.5 m as reported by the manufacturer. The magnaprobe GPS accuracy was considered as part of the preliminary data analysis by performing the following steps:

a) Magnaprobe measurements occurred in forested locations as shown in the photographs below. Magnaprobe data logger was programmed to provide a signal when GPS positioning was acquired and recorded at each snow depth measurement point.
b) Plotting magnaprobe data in the GIS and visually inspecting maps of the magnaprobe measurements for any unusual outliers in the spiral patterns (see yellow points in figures below).

c) Two of the sites had magnaprobe measurements that crossed a fixed road or trail. Maps and photographs were inspected for magnaprobe measurement alignment with the corresponding road or trail (see photographs and figures below).



As you can see from the figures, the spiral patterns and alignment with the road or trail look accurate. We acknowledged in the discussion that magnaprobe GPS accuracy is a potential source of uncertainty. We revised the manuscript to state the following:

In discussion: *The magnaprobe GPS units have a horizontal positioning error of ±2.5 m in open areas and 10–15 m in dense forest which could impact the correlation of ground-based snow depths with collocated lidar-derived snow depths. Efforts to minimize positioning errors involved taking ground-based measurements approximately 1–3 m apart, followed by visually inspecting the measurement pattern for outliers and ensuring alignment with a road or a trail.*

- How subsets were chosen?

Response 4: We used the USGS national land cover and airborne lidar canopy height maps to calculate the percentage of land cover and canopy height for the entire study site. We then created three smaller spatial subsets from the study area, using GIS software, that each contained the same land cover and canopy height percentages as the study site as a whole.

The methods section was revised to clarify this question as follows: *Each spatial subset was created to exhibit land cover and canopy height percentages comparable to those of the entire study site.*

- What was final resolution of the lidar data set after the reduction?

Response 5: The final resolution of the lidar data sets after the reduction is 1.5 m resolution.

The following sentences in results clarify: *The lidar snow depth map was resampled to a 1.5 m spatial resolution to eliminate spatial autocorrelation. The lidar canopy height map was resampled to a 1.5 m spatial resolution to correspond to the lidar snow depth map pixel size.*

- If I understood correctly, in-situ measurements were compared with lidar snow depths from the whole area. Why comparison was not made only with the lidar measuerments at the same area as the in-situ observations?

Response 6: Thank you for this comment. We clarified how in situ measurements were compared with lidar measurements by adding a new section (3.2) in the Methods.

3.2: Collocated Ground-Based and Lidar-Derived Snow Depths

Ground-based measurements were used as validation points to evaluate the accuracy of the lidar-derived snow depth map. GPS coordinates for every magnaprobe measurement were extracted from the magnaprobe data logger, QA/AC, and uploaded into a GIS software program. The corrected lidar-derived snow depth map, containing only plausible snow depth values, and before any reduction techniques, was uploaded at the same time. Using the nearest neighbor sampling technique, a single ground-based measurement was collocated to a corresponding lidar pixel (0.5 m resolution) and a lidar-derived snow depth value was extracted for each ground-based measurement. Error statistics were calculated using the ground-based and lidar-derived snow depth measurements to evaluate the accuracy of the lidar-derived snow depth map.

Technical Corrections

Line 17: "and standard deviation SD= 15cm" so that SD is not mixed to "snow depth"

Response 7: Abbreviation for SD is now spelled out as standard deviation.

Line 21: Typically SWE is presented in mm, 200-300 mm (same comment for the chapter 5.2)

Response 8: SWE is now reported in mm.

Line 147: Consider is it necassary to use DSM and DTM abbreviations instead of writing them open

Response 9: We prefer using the abbreviations after the full names and abbreviations are initially stated.

Line 356: "Ten previous studies"

Response 10: “previous” was added to the sentence.

Lines 410 and 414: Is "this study" referring to your study or previous studies by others?

Response 11: The lines in question have been revised in the discussion to now state “our study.”

Our study used ground-based measurements to validate an airborne lidar snow depth map in a boreal forest ecotype.

Our study produced an RMSE of 12.5 cm when validating airborne lidar with ground-based snow depth measurements.

Responses to Reviewer #2

We thank the reviewer for their comprehensive comments and suggestions. Their comments regarding the uncertainty assessment of low R^2 values when comparing lidar-derived and in situ snow depth measurements are greatly appreciated and have prompted the authors to improve the clarity and overall presentation of the manuscript. The reviewer's technical comments were valuable in guiding us toward a more polished and comprehensive representation of our research. We have carefully considered and addressed each of the reviewer's recommendations, and we believe that the revisions have strengthened our work. The following pages contain comments that appear exactly as they were received. Our responses are inserted next to each comment in blue text. Italicized sentences are directly from the edited manuscript.

Specific Comments

This is a review of "Analyzing vegetation effects on snow depth variability in Alaska's boreal forests with airborne lidar".

Overall, this is a well written manuscript that was easy to read.

My main concern is with the very low r^2 values reported for the comparison of lidar SD to in situ SD. I am unclear on how a couple pieces of the analysis were done, and coupled with this low r^2 makes me think that perhaps there is a post processing error in the analysis.

Response 1: Thank you for your concern with this particular topic. We, as the authors, have struggled with the very low R^2 value in this study as well. Our responses to your comments below, and subsequent changes to the manuscript, will hopefully convince the reviewer and reader that we have analyzed and scrutinized the data from all possible angles and reviewed post processing methods for unintended bias or errors. For standard lidar processing procedures please reference the following user guide:

Larsen, C.F. 2024. SnowEx23 Airborne Lidar-Derived 0.5M Snow Depth and Canopy Height, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/BV4D8RRU1H7U.

How, exactly, are the observed SD compared to the lidar SD? On L124-125 are the authors comparing the magnaprobe transect mean to a single lidar point? Or is the lidar transect averaged over the same length as the magnaprobe?

Response 2: To provide greater clarity on how observed snow depths were compared to the lidar snow depths we have added a new section (3.2) in the Methods.

3.2 Collocated Ground-based and Lidar-Derived Snow Depths

Ground-based measurements were used as validation points to evaluate the accuracy of the lidar-derived snow depth map. GPS coordinates for every magnaprobe measurement were extracted from the magnaprobe data logger, QA/QC, and uploaded into a GIS software program. The corrected lidar-derived snow depth map, containing only plausible snow depth values, and before any reduction techniques, was uploaded at the same time. Using the nearest-neighbor sampling technique, a single ground-based measurement was collocated to a corresponding lidar pixel (0.5 m resolution) and a lidar-derived snow depth value was extracted for each ground-based measurement. Error statistics were calculated using the ground-based and lidar-derived snow depth measurements to evaluate the accuracy of the lidar-derived snow depth map.

If it's the individual magnaprobe point to a lidar point then there is a substantial scale mismatch. If it's an average of the magnaprobe transect to a lidar cell, then it's also a scale mismatch. The authors allude to this a bit in the discussion but it needs to be front and centre. And I think, will be a clue to exactly what is going wrong with the r2 values.

Response 3: We acknowledge that there is a scale mismatch by comparing an individual point snow depth measurement with a corresponding collocated 0.5 m lidar snow depth pixel, and that the difference in scale could be contributing to the low R^2 value.

A recent paper (Komarov and Sturm 2023, https://doi.org/10.1080/15230430.2023.2170086) highlights the local variability of boreal snow cover at a smaller scale (5 m trench excavations with snow depth measurements occurring every 10 cm). They found snow depth variability within the 5 m trench is 6% in a birch forest site, 35% near a spruce tree with tree wells, and 4% in a spruce forest site without tree wells. We expect that variability in point snow depth measurements within 0.5 m horizontal distance is equal or less than reported by Komarov and Sturm (2023). Please see Response 4 and Response 6 for more details on the dense understory effect on lidar DTM's.

Given both the forest (5b) /and/ the open (5d) have such poor R^2 values, I am skeptical it's magnaprobe GPS uncertainty. In figure 3, the treeless in the top left is very different than treeless in the bottom left, especially w.r.t GPS signal and SD uncertainty. Because both the forest and treeless has as poor a R2, this feels like maybe a post processing step. The SD is deep enough that I'd expect a lot of the small rocks/grass/etc to be buried in the treeless (as per fig3) and thus result in a pretty clean lidar vs in situ comparison.

Response 4: While we acknowledge that magnaprobe GPS uncertainty could influence low R^2 values we believe that the greatest source of error is coming from the dense understory that exists in both the treeless and forested environments. In the bottom left picture of Figure 3 the treeless vegetation is adjacent to a gravel road and does not represent the majority of the treeless vegetation found throughout the site. The top left and right pictures are more representative of the dense understory (treeless) vegetation in the study area. The dense understory and tussocks found throughout this boreal forest impede lidar signals in the spring (impacting airborne lidar DTM's) and elevates snowpacks in the winter (impacting ground-based snow depth measurements). An example is taken from the field notes of an observer performing snow pit measurements at one of the study sites at the same time the magnaprobe measurements were occuring. The observer noted: "Snow is packed in on top of ~10 cm of brush. Snow pack height is starting above the brush, soil is frozen below the brush."

In Harder, et al (2020; https://doi.org/10.5194/tc-14-1919-2020) which is one of the few drone-lidar forest papers, the observation vs lidar is much more 1:1. The sub plots in Figure 5 might be interesting to look at on a per site (L124) basis, to see if there is a single site that is causing the major outliner points, or if this is from all (treeless / tree) sites.

Response 5: We agree with your suggestion and as part of our analysis we did calculate error statistics for each of the four sample locations separately. We did not find that any single site contributed significantly to the low R^2 value. Even with potential outliers removed, R^2 values did not improve. Below is a table of error statistics calculated for each site as a reference but is not included in the manuscript.

Site	Bias (cm)	RMSE (cm)	R^2	In situ HS mean±SD (cm)	Lidar HS mean±SD (cm)	n
WB920	-0.4	14	0.001	91±10	90±10	899
WN921	4	9	0.0003	87±7	91±5	278
DN922	3	12	0.023	88±9	91±9	479
DN923	4	13	0.0004	82±8	86±10	461

The OK aggregation step is, I think, correct, but the length scale is very short – 1 m. In canopy this seems reasonable but in the open it feels a bit short. Is there anything in this aggregation step that might be causing a bias in the lidar SD? e.g., it should be possible to extract the pre-OK high resolution lidar SD and see how the data compares to that to rule out any post processing of the OK aggregation step. Or look at larger averaging lengths to further remove sub-grid scale impacts.

Response 6: Ground-based snow depths were compared to high resolution (0.5 m) lidar snow depths, before any reduction techniques were applied. To provide greater clarity on how observed snow depths were compared to the lidar snow depths we added a new section (3.2) in the Methods (see Response 2).

We believe that the 1 m semivariogram range is correct for both treeless and forested areas. In our study area the treeless (or open) vegetation is composed of very dense understory and tussocks which can produce greater variability in snow depths compared to forested areas. We calculated forest HS standard deviation to be 8.7 cm, Shrub HS standard deviation is 10.6 cm, and treeless HS standard deviation is 9.3 cm.

This is all to say, it is difficult to tell from the figures here, but I would urge the authors to take another look at the lidar and observation data, and really diagnose if there is an unintended bias or /something/ in the post processing that has caused this very low r^2. Break it up by site, etc and convince the reader this is a correct result.

As is, the in situ vs lidar looks like random values centered around a mean.

Response 7: I hope our replies to the reviewer's comments above have sufficed to show that we have looked thoroughly and objectively at the data, results, and processing techniques to ensure no unintended bias and to acknowledge where uncertainties could arise.

Technical Corrections

L64 Include the above mentioned Harder, et al (2020)

Response 8: Harder et al. (2020) has been added as a reference

L75 pose this as a scientific research question

Response 9: Thank you for this suggestion, the following research questions have been added to the paragraph.

The purpose of this paper is to contribute to limited boreal forest snow remote sensing research by analyzing ground-based snow depth measurements and airborne lidar data to improve snow depth estimation at a watershed scale. Specifically, we ask: (1) Can combining ground-based measurements with airborne lidar improve snow depth estimation in the Caribou Poker Creeks Research Watershed (CPCRW)? (2) How does vegetation height influence snow depth variability within CPCRW?

L110 Figure 1, it should be good to show the tree cover in (c)
Response 10: Tree cover will be added to Figure 1 (c).

L115 move ? into the " " and remove from end of the sentence

Response 11: According to Cryosphere guidelines and house standards : It is our house standard to position commas and periods outside the end quotation marks.

This sentence has been revised:

Using field work and various remote sensing technologies, SnowEx Alaska aims to determine (Vuyovich et al., 2024): "How well we can characterize the spatial variability of snow depth and density needed for accurate SWE estimates in the boreal forest by measuring snow depth, density, and vegetation characteristics"?

L118 on tundra

Response 12: We clarified that this was focused on "Arctic tundra."

L163 "plausible SD" where were these plausible ranges? was this a subjective expert decision?

Response 13: The decision on maximum plausible snow depths was made based on available ground-based snow depth measurements. We have clarified this by revising the sentence below:

We referenced the maximum ground-based SnowEx snow depths, the maximum NRCS snow course measurements, and a histogram of the lidar snow depths to determine a plausible snow depth range for CPCRW.

L176 "smaller sections" is section = area? if so, I would use area. otherwise explain section.

Response 13: Thank you for suggesting this clarification. The following two sentences have been revised as follows:

To determine spatial autocorrelation, 15 smaller areas of the lidar snow depth map were partitioned, and a semivariogram was developed for each area.

The semivariogram ranges were then averaged over the fifteen areas, and the spatial resolution of the lidar snow depth map was adjusted based on the semivariogram range, to effectively eliminate spatial autocorrelation from contiguous pixels.

L177 "portioned" -> partitioned

Response 14: portioned has been changed to partitioned.

L183 “resampled” I assume this was via OK method noted, but make it clear

Response 15: The following sentences have been revised to clearly state which resampling technique was applied to the lidar data sets.

The semivariogram ranges were then averaged over the fifteen areas, and the spatial resolution of the lidar snow depth map was adjusted, based on the semivariogram range, to effectively eliminate spatial autocorrelation from contiguous pixels. The resulting snow depth map displays the independent snow depths required for the statistical analysis. The spatial resolution of the lidar canopy height map was altered to the same resolution as the new snow depth map with corresponding pixel sizes. Lidar maps were resampled using the nearest-neighbor technique.

L184 “canopy height resample” how was this resampled? bilinear? cubic? OK?

Response 16: See response 15.

L194 note Arc version

Response 17: the version of ArcGIS Pro has been included.

L204 how were these data thinned? Is this the resampling above?

Response 18: Yes, the thinning of the data is in reference to the resampling of the lidar data sets because of the semivariogram range. The sentence has been clarified to now read:

The assumption of independence was met by thinning the data according to the semivariogram range and applying the range to the resampling technique.

L207 what is the constant 12 in eqn 1?

Response 19: Please reference Kruskal and Wallis (1952) for further clarification.

L216 did the authors consider a 99% (0.01) range?

Response 20: Thank you for the suggestion. The following sentence has been revised to now state:

A low p-value (typically ≤ 0.05) indicates a statistically significant difference between the median snow depths for all groups.

L240 how sensitive are the results to changes in this cutoff value?

Response 21: We believe that if we were to adjust the cutoff values between the treeless and SSS canopy height classes the results would not change. We make this conclusion based on the results from the vegetation analysis which show that the difference in mean snow depths between these two canopy height pairs range between 0.1-4.0 cm. We feel that adjusting the cutoff values would not influence the sensitivity of the results.

L250 Figure 2 great plot

Response 22: Thank you

L255 Would be helpful to draw regions around the areas the labels apply to. Add sub-figure labels (a,b,c).

Response 23: This suggestion will be taken into consideration. Thank you.

Citation: https://doi.org/10.5194/egusphere-2024-4042-RC2