

Responses to Comments from Reviewer #1

General comments

In this manuscript, the authors present three different study sites in Alaska where NASA collected LiDAR snow depth data during SnowEx-campaigns in 2023. The aim is to examine a) how well these measurements correspond to reality in the area's various land cover types, particularly in forests, and b) utilize machine learning to address which terrain features specifically affect the accurate measurement of snow depth. One goal is to provide quality control tools for remote sensing snow depth products. I believe this article has much to offer in this field of research, as similar studies specifically involving LiDAR data collected from an aircraft with flight height of 700m and field measurements on this scale at boreal sites are scarce. In this regard, systematic reporting of the data, methods and results would be essential. The research design is simple, which means time could be devoted to more detailed reporting.

At this stage, it is not entirely clear to the reader how the data has been processed, how the analysis was structured and conducted, and the results are not reported consistently or clearly. The figures could be combined, and I recommend creating a summary table to support the findings presented in the figures. Additionally, the manuscript contains claims for which I would like to see more references, such as the effect of ground frost on snow depth. In general, I find the references to be quite insufficient—yet there have been quite a few studies on this topic that could be cited and compared. In the "Discussion" section, I would definitely like to see more reflection on the analysis and its limitations, together with comparison to similar studies.

I believe that clarifying the text, clearly stating and describing the data and methods, adding more recent publications to the sources, and streamlining or combining the figures—as well as possibly presenting data in tables—would significantly improve the article's readability and scientific value.

We sincerely thank Referee #1 for the careful and constructive evaluation of our manuscript. We appreciate the referee's recognition of the value of evaluating aircraft-based airborne LiDAR snow-depth measurements in boreal forest environments using the NASA SnowEx 2023 Alaska dataset. We agree that the original manuscript did not describe several aspects of the data processing, analysis workflow, and model evaluation with sufficient clarity. We also agree that the organization of the results, figures, literature review, and discussion can be substantially improved.

In response to these comments, we have revised the manuscript to more clearly distinguish the processing performed by the SnowEx data providers from the analyses performed in this study, provided additional methodological details and model-validation information, improved the organization and presentation of the results, consolidated figures where appropriate, added summary tables, expanded the relevant literature, and strengthened the discussion of uncertainty and limitations.

Specific comments

Comment 1:

The abstract could do a better job of highlighting why this study is so important and what it actually found. After all, quite a few studies like this have already been conducted. Perhaps it could be marketed better: why is this study so important?

Response:

We agree that the original Abstract does not sufficiently emphasize the specific contribution of this study relative to previous LiDAR snow-depth studies.

The main contribution is that this study evaluates fixed-wing aircraft-based airborne LiDAR observations collected over multiple boreal forest sites during NASA SnowEx 2023 and examines how LiDAR snow-depth errors vary systematically with canopy and terrain characteristics using both statistical analyses and a machine-learning framework.

We have revised the Abstract to more clearly emphasize this context and to report the principal quantitative findings, including the overall LiDAR–in situ agreement, the systematic underestimation observed in more challenging forested conditions, and the dominant importance of canopy height in the error analysis.

Revised manuscript (Page 1) with key changes in bold italic

Abstract. Quantifying spatial distribution of the snowpack is crucial for hydrological, ecological, and climate research, as well as their applications. Due to the high spatial resolution and extensive coverage, Airborne Light Detection and Ranging (LiDAR) has emerged as an effective tool for large-scale snow depth estimation. However, discrepancies between LiDAR-derived and manually measured snow depth values exist across areas

influenced by topographical and vegetation characteristics such as canopy height, slope, and roughness. *This study examines how these errors vary systematically with canopy and terrain characteristics across multiple boreal forest sites, using both statistical analyses and a machine learning framework.* This study aims to 1) evaluate the performance of the airborne LiDAR snow depth measurements compared to Magnaprobe in-situ data and 2) identify key factors affecting the accuracy of airborne LiDAR snow depth measurements focusing on the boreal forest environment. We utilize airborne LiDAR data and ground-based snow depth observations collected in the Fairbanks region of central Alaska during NASA SnowEx 2023 Alaska Campaign. The study focuses on three subregions: Bonanza Creek Experimental Forest (BCEF), Farmers Loop Creamers Field (FLCF), and Caribou-Poker Creeks Research Watershed (CPCRW). The results showed that the LiDAR snow depth data has a reasonable agreement with in-situ observations (R: 0.61, Mean Absolute Error: 18.8 cm) but exhibits varying levels of errors across the three subregions, *with LiDAR increasingly underestimating snow depth in more densely forested and topographically complex areas.* By applying regression analysis and machine learning, we quantify the contribution of individual factors to measurement discrepancies and determine which factors are most influential. We employed Gradient Boosting Machine (GBM) model using five LiDAR-derived environmental variables—canopy height, elevation, slope, roughness, and ground point density—as predictors of relative error. Across all subregions and models, canopy height consistently emerged as the most important factor of LiDAR snow depth error. These findings provide a framework for assessing LiDAR snow depth reliability across boreal forest environments in North America.

Comment 2:

Magnaprobe is mentioned in the abstract, but never afterwards in the text. In the methods section, a manual probe is described. It should be correctly stated which instrument has been used.

Response:

Thanks for identifying this inconsistency. According to the official SnowEx 2023 field documentation, the in-situ snow depth measurements used in this study were collected using a Magnaprobe. Therefore, the term “Magnaprobe” in the Abstract is correct, while the more general term “manual snow-depth probe” used in the Methods section was imprecise.

We have revised the manuscript to consistently refer to the instrument as a Magnaprobe throughout the Abstract, Methods, Results, and Discussion, and we have added appropriate references to the SnowEx 2023 field documentation in the Methods section.

Comment 3:

When describing the previous LiDAR studies in the introduction, I would recommend clarifying the context at the beginning. LiDAR can easily achieve centimeter-level accuracy, and point density way higher than 1 point/m² depending on the vehicle and flight height and speed, overlap etc. I assume the authors are referring specifically to LiDAR mapping done from an airplane, not a drone. This could be clarified for the reader to make it clear that the study specifically refers to high-altitude ALS devices, not drones. Later in the introduction UAS are introduced, so I would restructure the chapter to clarify different devices and what accuracies there are in different ALS platforms.

Response:

We thank the reviewer for this comment. Upon review, we found that the "about 1 point/m²" figure was drawn from the general characterization in Deems et al. (2013), whose abstract describes typical airborne lidar datasets as having "decimeter-scale with order 1 m point spacings." We agree that presenting this figure as representative of "modern crewed-aircraft ALS systems" in general is misleading, both because it reflects a broad, decade-old characterization rather than current practice, and because it does not reflect the actual point density of the dataset used in this study. We have revised the text to remove this generic density claim and instead report the native point density of the SnowEx23 airborne LiDAR dataset used here, approximately 20 points per square meter, as documented in Fair et al. (2025), which independently characterizes the same UAF SnowEx Alaska lidar dataset.

Revised manuscript (Page 2 - 3) with key changes in bold italic

Airborne LiDAR (Light Detection and Ranging) has emerged as a promising solution for high-resolution snow depth mapping in complex terrains, including forests (Painter et al., 2016; Mazzotti et al., 2019; Deschamps-Berger et al., 2020; Currier et al., 2019, Deems et al., 2013). ***LiDAR actively emits laser pulses and measures their returns, capable of penetrating gaps in canopy to reach the snow surface or ground (Lefsky et al., 2002).***

Modern crewed-aircraft Airborne Laser Scanning (ALS) systems, the focus of this study, typically achieve vertical accuracies on the order of decimeters (Deems et al., 2013), while achievable point density varies considerably with system specification, flight altitude, and survey objective. The airborne LiDAR dataset used in this study was collected at an average density of approximately 20 point/m² (Fair et al., 2025), consistent with surveys designed to resolve forest structure and terrain. This makes it possible to detect snow depth by comparing snow-on and snow-off surface elevations. Notably, LiDAR in near-infrared wavelengths can even map sub-canopy topography to some extent, which is a major advantage over passive optical methods (Clark & Clark, 2004). Over the past decade, numerous studies have demonstrated LiDAR's utility for snow depth estimation. For example, Deems et al (2013) provided an extensive review of LiDAR snow measurement, noting typical airborne LiDAR vertical uncertainties of ~0.10–0.25 m in open areas and its ability to generate continuous, high-resolution maps of snow depth rather than the sparse point measurements provided by traditional in situ methods. Unpiloted Aerial Systems (UAS) have recently enabled the use of drone-based LiDAR for snow depth mapping (Harder et al., 2020; Cho et al., 2025). Because UAS platforms typically fly at lower altitudes and slower speeds, they can achieve much higher point densities and finer spatial resolution than crewed ALS systems, offering greater flexibility for surveying small or difficult-to-access regions and making them increasingly popular for snow research in forested environments. More recent work with UAS LiDAR in New Hampshire showed that high-density LiDAR point clouds could measure a shallow 10 cm snowpack with ~1 cm error in open fields, though errors increased to ~10 cm under forest canopy (Jacobs et al., 2021). This highlights that while LiDAR substantially improves snow mapping, its accuracy can degrade in the presence of dense vegetation, where fewer laser pulses reach the ground and the snow surface itself may be harder to discern (Dharmadasa et al., 2022).

Comment 4:

In my opinion, the "Introduction" section could include additional sources that evaluate the accuracy of LiDAR in measuring snow depth, through different platforms (airplanes, helicopters, drones, terrestrial). Currently, there is an entire paragraph there that cites only NASA 2018 as its source—but that section discusses, for example, mapping alpine environments and remote sensing of snow in general.

Response:

We agree that the literature review should be more comprehensive. The original manuscript relies too heavily on a limited number of studies and does not sufficiently distinguish results obtained using aircraft, UAS, terrestrial, and other LiDAR platforms.

We have expanded the Introduction and Discussion to include additional studies evaluating snow-depth accuracy across different LiDAR platforms, with particular emphasis on aircraft-based airborne LiDAR, forested environments, and previous SnowEx campaigns. We have also revised statements that previously rely primarily on the NASA SnowEx Science Plan so that methodological or scientific claims are supported by peer-reviewed literature whenever possible.

Revised manuscript (Page 3 - 4) with key changes in bold italic

Despite the recognized potential of airborne LiDAR, there is a need for systematic assessment of its accuracy in forested regions and an understanding of what conditions lead to larger errors. *Accuracy evaluations of LiDAR-based snow depth mapping have been conducted across a range of platforms and forest conditions, though predominantly outside boreal environments. Airplane- and unpiloted aerial system-based studies have validated LiDAR snow depth performance in alpine and temperate forests (Hopkinson et al., 2004; Jacobs et al., 2021), helicopter-based systems have been used to assess LiDAR performance in steep, forested terrain (Bühler et al., 2015; Wilder et al., 2025), and terrestrial laser scanning studies have specifically examined the influence of forest canopy structure on snow depth retrieval accuracy (Prokop, 2008; Revuelto et al., 2015; Hojatimalekshah et al., 2021). Relatively few of these evaluations, however, have focused on boreal forests, which are characterized by cold climates, discontinuous permafrost, and dense conifer canopies.* These unique features strongly influence both snow accumulation and LiDAR performance, making boreal forests a distinct and understudied environment in snow remote sensing (NASA, 2018). The NASA SnowEx 2023 campaign in Alaska provides an ideal opportunity to evaluate LiDAR snow depth retrievals in a boreal forest setting across varying conditions. By combining airborne LiDAR data with an extensive set of ground truth measurements and environmental variables, we can quantify how well LiDAR performs and what drives any discrepancies.

Comment 5:

There are no sources in the section describing the study sites, yet it describes weather and snowfall amounts. I would recommend including information on the years for which the averages were calculated, as well as the data source and location.

Response:

We thank the reviewer for this comment. We agree that the climate and snowfall figures in this section lacked proper sourcing, and we have revised the paragraph to specify the averaging period, data source, and location for each figure, as requested.

Years for which averages were calculated: All regional climate normals (mean January temperature, average winter snowfall) now reference the standard 1991–2020 30-year averaging period.

Data source: These normals are cited from the NOAA National Centers for Environmental Information (NCEI) U.S. Climate Normals dataset.

Location: The specific station used is Fairbanks International Airport (Station ID: USW00026411; elevation 432 ft / 132 m; 64.8039°N, 147.8761°W), Fairbanks International Airport was selected as a long term regional climate reference station for the Fairbanks area.

Based on this source, mean January air temperature is –22.4 °C (–8.3 °F), and average winter (October–April) snowfall totals approximately 156 cm (61.4 in).

We have also separated the winter 2022–2023 snow depth figures (previously stated without a clear source) from these long-term climate normals. These now explicitly reference the manual, in-situ snow depth measurements collected during the March 2023 Intensive Observation Period (IOP) of the NASA SnowEx 2023 Alaska field campaign, used as part of this study, which recorded depths ranging from approximately 15–120 cm across the three sites (site medians: 67 cm at BCEF, 89 cm at CPCRW, 67 cm at FLCF).

Revised manuscript (Page 3 - 4, Lines 95 – 105) with key changes in bold italic

BCEF is located ~20 km southwest of Fairbanks and includes both lowland floodplain forests and upland rolling hills within the Tanana River valley. Vegetation at BCEF is dominated by mature spruce and birch forest in the uplands and mixed hardwood-conifer stands in lowlands; the area includes some experimental plots managed by

the Bonanza Creek Long Term Ecological Research program. *Within our study subregion, terrain is gently to moderately sloping, though BCEF spans a broader elevation range, with uplands rising to a ridge crest of approximately 470 m (Bonanza Creek LTER, 2002).* CPCRW is a 104 km² research watershed northeast of Fairbanks in the Caribou-Poker Creeks area, featuring more complex topography and higher elevation forests *(Bonanza Creek LTER, 2002).* Elevations in CPCRW range from ~200 m to over 800 m. *Vegetation follows a strong aspect-driven pattern: black spruce dominates north-facing slopes and poorly drained flats, while south-facing, well-drained slopes support aspen, birch, and alder, with white spruce occurring only occasionally. Tree density decreases toward treeline at approximately 600 m elevation, where open forest gives way to dense willow, alder, and dwarf birch shrub. Approximately 30.7% of the watershed is underlain by permafrost (Haugen et al., 1982), and the basin contains steep, narrow stream channels and a mosaic of permafrost and permafrost-free terrain, making it a challenging but representative site for studying snow in forested terrain. FLCF refers to Farmer Loop/Creamer's Field, one of three boreal forest study sites near Fairbanks used in the NASA SnowEx 2023 Alaska campaign, alongside BCEF and CPCRW. The FLCF study area sits at approximately 130–180 m ASL and includes open meadow/pasture areas adjacent to birch-spruce woodlands, representing a more open forest environment within the discontinuous permafrost zone. The regional climate of interior Alaska is subarctic continental, with cold, relatively dry winters. Based on the 1991–2020 U.S. Climate Normals at the Fairbanks International Airport station (NOAA NCEI, Station USW00026411), mean January air temperature is approximately –22.4 °C, and average winter (October–April) snowfall totals approximately 156 cm. Snow depth on the ground by late winter is typically lower than seasonal snowfall totals due to settling and sublimation. The winter of 2022–2023 during SnowEx featured sustained snow cover from October through April. In forested areas, snow depths tend to be lower under canopy than in open areas due to interception; field observations in the region have shown significant differences in snow depth between forested and open areas in Interior Alaska (May et al., 2025), although the specific range of snow depth variability may vary regionally and interannually. Manual, in-situ snow depth surveys from the NASA SnowEx 2023 Alaska field campaign recorded snow depths typically ranging from approximately 61 to 85 cm across the three sites (interquartile range), with the full observed range spanning 15–120 cm. This range of conditions across the three study sites provides a robust testbed for airborne LiDAR*

snow depth estimation: BCEF and CPRW include dense canopy and sloping terrain, whereas parts of FLCF offer open, flat reference areas.

Comment 6:

Section 3.1: In this section, it would be helpful to explain in more detail: where the point clouds were georeferenced, what was done with them afterward, what algorithm was used to calculate the DEMs, and in which software program. Later, there's a reference to NASA. This could be clarified for readers who haven't worked with this data before. Since metadata is available for Larsen 2024, one could, for example, note that additional information on the data processing can be found in this source.

Response:

We thank the reviewer for this suggestion. We have expanded Section 3.1 to specify: (1) the georeferencing system used (WGS 84 / UTM Zone 6N, EPSG:32606); (2) the full processing chain, including Applanix and Riegl software for initial point cloud generation, BayesMap StripAlign for flight-line alignment, and Quick Terrain Modeler for point cloud classification and DTM/DSM extraction; and (3) the specific method used to calculate snow depth (direct elevation differencing between snow-on and snow-off DTMs, with no reprojection or warping). We also note, as the reviewer suggested, that additional processing details and data accuracy information are available in the Larsen (2024) data set documentation, which we now cite directly alongside the NSIDC-hosted user guide.

Revised manuscript (Page 6 - 7):

Raw point clouds were georeferenced to WGS 84 / UTM Zone 6N (EPSG:32606). Raw data were first processed using Applanix and Riegl software to generate individual flight-line point cloud swaths. These swaths, together with the aircraft trajectory solution, were further processed in BayesMap StripAlign software to improve positional accuracy and reduce swath-overlap offsets. The aligned swaths were then merged and classified using the Quick Terrain Modeler software classification and extraction tool to produce gridded GeoTIFF terrain models: a bare-earth digital terrain model (DTM) from the snow-off survey, and a canopy digital surface model (DSM) from first returns. Canopy height was calculated as the DSM minus the DTM. Snow depth was calculated as the

direct elevation difference between the snow-on DTM and the snow-off DTM at each 0.5 m grid cell, with no reprojection or warping applied between the two epochs (Larsen, 2024). Individual DTM accuracy is estimated at approximately 5 cm, with an internal precision of approximately 1 mm (Larsen, 2024).

Comment 7:

Section 3.1: Areas where there were no valid return signals were removed or extrapolated only in March—was there no need to do this in the May data? Why?

Response:

Thank you for pointing this out. The statement regarding the treatment of areas with no valid returns was inadvertently retained from an earlier exploratory workflow and does not reflect the final analysis.

In the final workflow, we used the official SnowEx point clouds and did not apply separate masking or extrapolation to the March data. Therefore, no corresponding treatment was required for the May data. We have removed this statement from Section 3.1 and clarified the processing procedures used in the final analysis.

Revised manuscript (Page 7-8, Lines 204 – 211):

The snow-on (March) point cloud was processed to obtain the elevation of the snow surface, using the same ground-classification procedure applied to the snow-off (May) point cloud. The resulting snow depth raster was produced at 0.5 m resolution (chosen to balance detail and noise) as the difference between the snow-on surface and the snow-off ground surface at each grid cell. Areas with no valid returns in either epoch (e.g., extremely dense canopy with no ground or snow returns) were flagged as no-data in the provided products and excluded from the snow depth calculation; this treatment was applied equally to the March and May acquisitions and was determined by the data provider's processing rather than by any separate procedure adopted in this study.

Comment 8:

The text makes extensive use of parentheses () to supplement sentence structures. While their use may be justified in some cases, I think there are far too many of them, which makes the text confusing. I recommend significantly reducing the use of parentheses and considering how the ideas could be expressed without them.

Response:

We agree and we have edited the manuscript throughout to reduce unnecessary parenthetical statements.

Explanatory information that is important to the scientific argument is incorporated directly into the sentence structure, whereas nonessential details are removed.

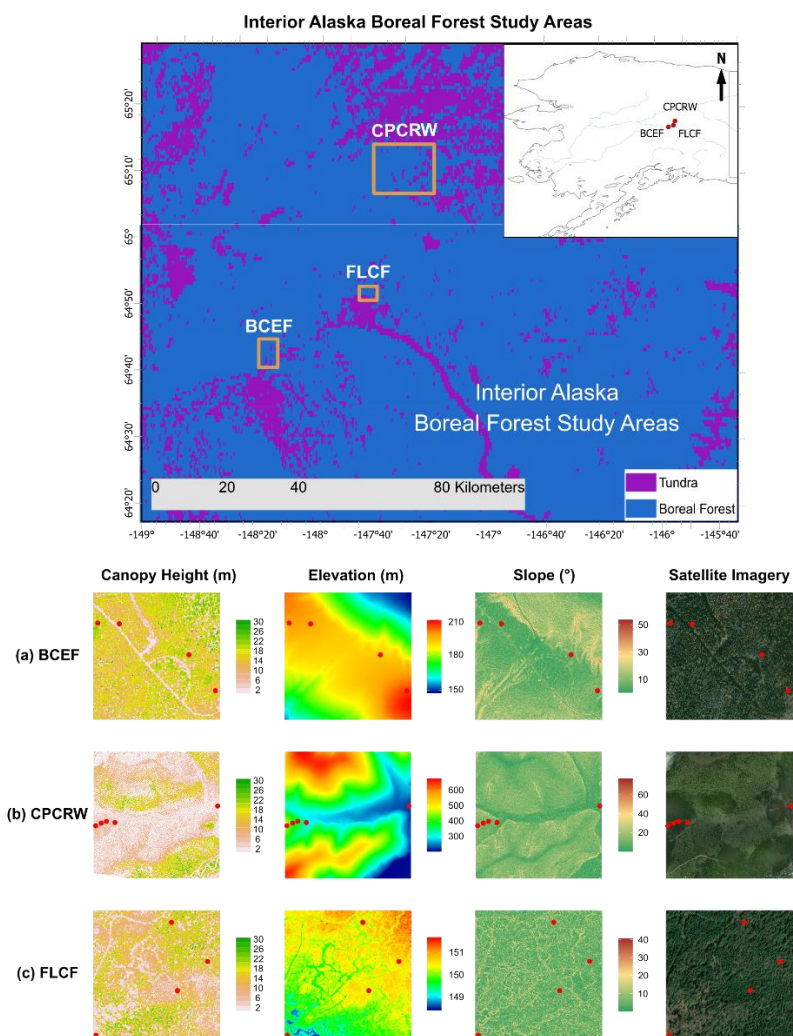
Comment 9:

Figures 1 & 2 could be combined, now the figure 1 is a broad figure without too much information in practical, meanwhile figure 2 is giving a lot of important background information of the area itself. Figure 2 is not referred in any part of the main text.

Response:

We thank the reviewer for this suggestion. We have combined Figures 1 and 2 into a single figure, retaining the broad site-location map (with orange boxes indicating the zoomed sub-regions) and removing the previously redundant site-scale elevation panels. Below the location map, we retain the canopy height, elevation, slope, and aspect panels for the zoomed sub-region of each site. We have also added explicit reference to this figure in Section 2, where we now describe the topographic and vegetation contrasts between the three sites with direct reference to the figure.

286 The revised figure now appears as follows:



287
288 Figure 1. Location of the three study areas within interior Alaska's boreal forest and tundra zones, with orange
289 boxes indicating the zoomed-in sub-regions used for detailed environmental characterization. Inset shows the
290 regional context within Alaska. Below, canopy height, elevation, slope, and RGB imagery are shown for the
291 selected subregion of each site: (a) BCFE, (b) CPCRW, and (c) FLCE. The RGB imagery was accessed through
292 Esri World Imagery. The BCFE and CPCRW panels use 0.5 m WorldView 2 imagery from the Vantor Vivid
293 basemap, while the FLCE panel uses 0.34 m WorldView Legion 1 imagery from Vantor Vivid Advanced. These
294 images provide visual context only and were not used in the quantitative analyses. Canopy height was calculated
295 by differencing the DSM and DTM; elevation and slope were extracted from the SnowEx-derived DTM. Red

dots indicate the locations of manual snow depth measurement sub-regions within each study site.

Comment 10:

Section 3.3: The sources are poorly cited. As I noted earlier in comment 5, in some sections they are not cited at all. But in this section, for example, Larsen 2024a and Larsen 2024b are not cited in the reference list in a way that makes it clear to the reader which source is being referred to. Overall, section 3.3 could be more concise without subheadings. References to Larsen could be cited more clearly. This section discusses relevant points regarding why these parameters were chosen, but these could perhaps be addressed elsewhere rather than in this "Data and Methods" section. Variables in equation 2&3 are not explained.

Response:

We thank the reviewer for these detailed comments. We have made the following revisions to Section 3.3:

1. We corrected the citation inconsistency: both canopy height and elevation are derived from the same SnowEx23 Airborne Lidar-Derived 0.5M product. We have revised the manuscript to consistently cite this derived product as Larsen (2024b) and the related raw lidar scan dataset as Larsen (2024a) throughout the text and the reference list, so that each in-text citation and its corresponding reference list entry are unambiguously matched.
2. We removed the five subheadings and condensed Section 3.3 into a single concise paragraph describing the five predictor variables.
3. We moved the rationale for why each variable was expected to influence lidar snow depth accuracy to the Introduction/Discussion, retaining only objective variable definitions in the Methods section.
4. We added explicit definitions for all variables in Equations (2) and (3).

Revised manuscript (Page 8):

$$\text{Slope} = \arctan\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right) \quad \text{eq. (2)}$$

where z denotes the DTM elevation at a given pixel, and x and y denote the horizontal grid coordinates in the east-west and north-south directions, respectively. The partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ represent the elevation gradient in each direction, estimated from the elevation differences among the neighboring pixels within the 3×3 moving window. The resulting angle was converted from radians to degrees.

$$\text{Roughness} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2} \quad \text{eq. (3)}$$

where n is the number of DTM pixels within the 3×3 moving window ($n = 9$ for interior pixels), z_i is the elevation of the i -th pixel within the window, and $\bar{z}$ is the mean elevation of all pixels within that window.

Comment 11:

Section 3.5: To support the choice of this method, we would need references explaining why it is so well-suited for a dataset of this size (300 points per site?). Additional information about the method, such as training data, testing, and validation sample sizes, would be important.

Response:

We thank the reviewer for this comment and will add the requested information to Section 3.5. After quality control, a total of 1,012 observations were retained for the GBM analysis. Of the original 1,060 in-situ measurements, 48 records were excluded because of missing or incomplete ancillary information in the field records, such as missing elevation values or unidentified measurement-device information. The remaining data at each site were split 80/10/10 into training, test and validation sets. We have corrected the earlier, imprecise statement of “roughly 300 points per site” to reflect these site-specific totals, which vary considerably from 207 to 555 observations because of differences in field sampling coverage and the availability of complete predictor information.

For each candidate hyperparameter combination (interaction.depth, n.trees, shrinkage, n.minobsinnode), a model was fit using the training subset, and its performance was evaluated on the validation subset; the combination yielding the highest validation R^2 was selected as the final model for each site. The selected model was then evaluated on the independent test subset.

We selected GBM not because of any inherent suitability for small sample sizes, but because of its ability to capture nonlinear relationships and interactions while providing interpretable measures of variable importance and partial dependence. Given the limited sample size, we constrained model complexity through hyperparameter tuning, including tree depth, learning rate, number of trees, and minimum node size, to reduce the risk of overfitting (Friedman, 2001).

Revised manuscript (Page 10 - 11):

The response variable for modeling was the relative error in snow depth, defined as (LiDAR depth – in-situ depth) / in-situ depth. We chose relative error (dimensionless, expressed as a fraction) to normalize for the fact that deeper snow measurements could tolerate a larger absolute error. This also helps stabilize variance if, for example, one site had generally deeper snow than another. Points with very shallow snow (< 10 cm) were treated carefully in relative error computation to avoid disproportionately amplifying relative error magnitudes (in such cases, we also considered absolute error models for comparison). *After quality control, a total of 1,012 of the original 1,060 in-situ measurements were retained for the GBM analysis; 48 records were excluded due to missing or incomplete ancillary information, such as missing elevation values or unidentified measurement-device information. We trained separate models for each site. Site-specific sample sizes varied considerably, from 207 to 555 observations, depending on differences in field sampling coverage and the availability of complete predictor information; training separate models by site allows us to see if the importance of factors changes with location. At each site, the retained data were split 80/10/10 into training, validation, and test sets. For each candidate hyperparameter combination (interaction depth, number of trees, shrinkage, and minimum number of observations per node), a model was fit using the training subset, and its performance was evaluated on the validation subset; the combination yielding the highest validation R2 was selected as the final model for that site.* The selected model was then evaluated on the independent test subset to obtain an unbiased estimate of predictive performance.

Comment 12:

Figures 4-6: It is not clear to the reader what these four sub-areas actually are. The captions are repetitively unclear; it consists of four lines that repeat the same information twice. I'd cut the number of these maps in half—there are currently three, and each one has four areas that don't really provide any context for the maps. It would work just fine to have, say, two examples per page.

Response:

We agree that the current presentation contains too many similar maps and that the purpose of the numbered subregions is not adequately explained.

We have simplified these figures substantially. The revised manuscript combined representative examples across the three sites. The captions are shortened and rewritten to describe what is displayed rather than repeat the Results text.

The revised figure now appears as follows:

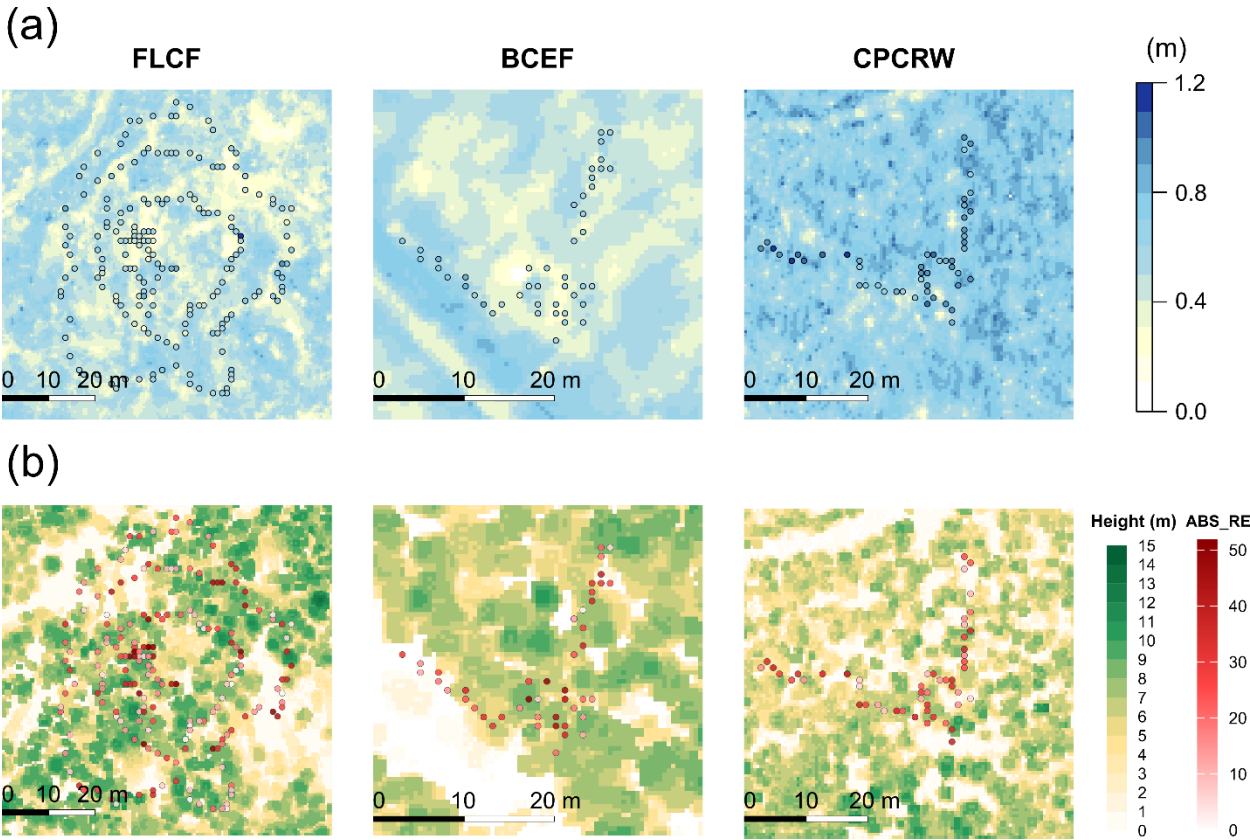


Figure 4. LiDAR-derived snow depth, canopy height, and error distribution across the three boreal forest study sites: Farmers Loop and Creamer's Field (FLCF), Bonanza Creek Experimental Forest (BCEF), and Caribou-Poker Creek Research Watershed (CPCRW). (a) LiDAR-derived snow depth for a representative region within each study site, overlaid with in-situ manual snow depth measurement locations (black circles). (b) Canopy height and absolute relative error (ABS_RE) between LiDAR-derived and manually measured snow depth for the same regions.

Comment 13:

Section 4.2: It might be a good idea to create a summary table of the results, so you can see, for example, the sample sizes across the different groups. Right now, it's not very easy for the reader to understand the differences

between the groups. Once again, parentheses are used, which clutter up the text and bring up points that have already been explained. The phrase "markedly higher proportions" is used, even though the differences are not necessarily that large; there is also variation within them, which may be related to the method itself. For example, the largest proportion of very high relative error is in the 10–15 m class, but the error decreases in the other, larger tree size classes. Text is also referring to figure 8, but figure 8 has also a, b and c sub-figures. The figure captions are showing information that should be explained in the text rather than in the captions. Overall, when looking at this analysis, one might wonder why elevation was even included as an explanatory factor in the first place. The reasons behind these choices need to be explained more clearly. The section also claims that “Higher density areas tended to have smaller errors, suggesting that sufficient ground returns improve terrain characterization beneath the snowpack”—of course, more points produce a more accurate terrain model. And in that case, there’s no need to even see beneath the snow. It sees the surface it’s mapping. The statement doesn’t really make sense.

Response:

We thank the reviewer for these detailed suggestions and will revise Section 4.2 accordingly:

1. We have added the summary table reporting sample sizes for each group discussed in this section.
2. We have removed redundant parenthetical asides that restate points already explained earlier.
3. "Markedly higher proportions" was replaced with precise, quantified language reflecting actual within-group variation.
4. We have clarified the non-monotonic pattern across the higher canopy classes: the proportion of high-error cases peaks at 15.2% in the 10–15 m class, decreases to 9.3% in the 15–20 m class, and then rises again to 14.0% in the ≥ 20 m class, rather than following a simple increasing or decreasing trend with canopy height.
5. In-text references to Figure 8 have specified sub-panels (a, b, c).
6. Descriptive content have moved from the figure caption into the main text.
7. We have clarified that Elevation was included as an a priori topographic covariate to evaluate whether within site topographic position and associated environmental gradients were related to LiDAR snow depth error, not as a variable with an assumed direct causal effect.

8. We agree the "beneath the snowpack" phrasing was imprecise and does not make physical sense as written. Ground point density is calculated from the March (snow-on) point cloud using ground-classified points (Classification = 2, ASPRS standard); these points represent snow-surface returns reaching through canopy gaps, not returns penetrating to bare earth beneath the snow. We have revised the text to state that ground point density instead reflects how effectively lidar pulses penetrated the canopy to reach the snow surface, with sparser returns implying greater reliance on interpolation and thus higher uncertainty.

Revised manuscript (Page 15, Lines 366 – 376) with key changes in bold italic

We next analyzed how the LiDAR snow depth error (LiDAR minus ground measurements) varies with the environmental factors of interest (canopy height, slope, surface roughness, elevation and ground point density). Figure 7 shows that higher canopy classes tend to have a larger proportion of high-error cases. In low-canopy areas (0–5 m), high errors (Absolute relative error ≥ 0.5 , 50%) account for only about 3% of the points, with most observations in the low-error range (Absolute relative error ≤ 0.25 , 25%). *In contrast, canopy heights of 10–15 m, 15–20 m, and ≥ 20 m show a substantially higher proportion of high-error cases relative to the two lowest canopy classes, rising from 3.4% at 5–10 m to 15.2% at 10–15 m, 9.3% at 15–20 m, and 14.0% at ≥ 20 m, though the proportion does not increase monotonically across these three higher bins. This shift is accompanied by a corresponding reduction in the proportion of low-error cases (Absolute relative error ≤ 0.25), which falls from around 42–67% in the two lowest canopy classes to roughly 41–51% in the three higher classes.*

Table 1. Sample size and distribution of LiDAR snow depth error across canopy height categories.

Canopy Height	n	High Error, n (%)	0–0.25	0.25–0.5	0.5–0.75	≥ 0.75	OR (95% CI)
0–5 m	366	10 (2.7%)	67.5%	29.8%	2.2%	0.5%	1.00 (ref)
5–10 m	234	8 (3.4%)	41.9%	54.7%	3.4%	0.0%	1.26 ns (0.47–3.24)
10–15 m	165	25 (15.2%)	41.2%	43.6%	14.5%	0.6%	6.36 *** (3.06–14.20)

Canopy Height	n	High Error, n (%)	0–0.25	0.25–0.5	0.5–0.75	≥0.75	OR (95% CI)
15–20 m	204	21 (10.3%)	48.0%	41.7%	9.3%	1.0%	4.09 *** (1.93–9.23)
≥20 m	43	6 (14.0%)	51.2%	34.9%	14.0%	0.0%	5.77 ** (1.87–16.48)

439

440 **Revised manuscript (Page 17, Lines 394 – 414) with key changes in bold italic**

441 Overall, the relationship between roughness and error was less consistent than that observed for canopy height,
442 suggesting that roughness is a secondary factor influencing LiDAR performance (Figure 9b).

443 *Elevation was included as an a priori topographic covariate to evaluate whether within-site topographic*
444 *position and associated environmental gradients were related to LiDAR snow depth error, rather than as a*
445 *variable expected to have a strong direct effect on measurement accuracy on its own.*

446 As a standalone factor, no consistent monotonic relationship between elevation and error was observed within
447 individual sites. Although CPRW exhibited both higher elevations and larger errors, this likely reflects site-
448 specific conditions rather than a direct elevation effect. Within each site, elevation differences did not
449 noticeably impact LiDAR performance after accounting for local slope and canopy. Therefore, elevation is
450 acting as a proxy for site characteristics rather than a direct influence on error.

451 Ground point density, defined as the density of points classified as ground returns (ASPRS Classification 2)
452 within the March snow-on point cloud, showed a moderate relationship with LiDAR snow depth error.

453 Although labeled "ground" under the standard LiDAR classification scheme, under snow-on conditions these
454 returns represent laser pulses that reached through gaps in the canopy to the snow surface itself, rather than
455 pulses penetrating to the bare ground beneath the snowpack. Locations with lower ground point density
456 generally exhibited larger errors, particularly in CPRW and FLCF (Figure 9c), while higher-density areas
457 tended to have smaller errors. This pattern suggests that sparser returns provide less direct information about
458 the snow surface at a given location, requiring greater reliance on interpolation from surrounding points and
459 thus increasing the uncertainty of the derived snow depth estimate. Many low-density observations occurred
460 beneath taller canopies, indicating that canopy occlusion reduces the number of laser pulses reaching the snow

surface.

Comment 14:

Section 4.3: It says here that this is a reasonable result. But what is that result it is referring to? The CPCRW region is also significantly larger in area than the others, which might also affect the analysis?

Response:

We thank the reviewer for this comment. We agree the phrase "this is a reasonable performance" was vague and appeared before the specific R^2 values were given. We have restructured the paragraph to first report the site-specific R^2 values (CPCRW = 0.53, BCEF $\approx$ 0.33, FLCF $\approx$ 0.18), then discuss their interpretation.

We agree that the larger spatial extent of CPCRW may have influenced the analysis. A larger study area is more likely to encompass greater topographic complexity and a broader range of environmental conditions. At the same time, when the number of observations is limited, sampling a larger and more heterogeneous area may result in relatively sparse representation of this environmental variability, which could negatively affect model performance and increase uncertainty. We therefore acknowledged spatial extent, together with terrain heterogeneity and sample size, as a potential factor affecting differences in model performance among the three sites.

Revised manuscript (Page 3 - 4) with key changes in bold italic

To quantify the control drivers of snow depth errors, we used a GBM model to predict spatial errors and identify the contributing input variables. *Model performance varied by site: the CPCRW GBM model performed best ($R^2 = 0.53$, explaining about 53% of error variance), followed by BCEF ($R^2 \approx 0.33$) and FLCF ($R^2 \approx 0.18$).* Given the inherent variability, unmodeled factors, and possible measurement noise in LIDAR snow depth errors, *these values, particularly at CPCRW, represent a reasonable level of explained variance using just five predictors.*

The stronger performance at CPCRW may partly reflect its wider range of terrain and slope conditions, which could provide the model with more signals to learn from. However, CPCRW also covers a substantially larger and more heterogeneous spatial extent than BCEF or FLCF, and with a fixed number of ground observations, a

larger and more variable area could equally result in sparser sampling of that environmental variability, which would tend to reduce rather than improve model performance. We therefore interpret the observed differences in R^2 across sites cautiously: spatial extent, terrain heterogeneity, and sample size likely interact to influence model performance in this dataset, and the current results cannot fully disentangle their individual contributions.

Comment 15:

Figure 11: In this figure, the captions don't explain differences between a/b/c/d. It is not clear, what are the two lines labeled 2 and 10 in c-plot, or what are those lines at the bottom of each subplot? For example, in that c-plot, they look completely different from the ones in the others. It would probably be enough if that labels were just on the X-axis and in the figure caption – currently each subplot has a title.

Response:

We thank the reviewer for this helpful suggestion. We have revised Figure 11 and its caption to make the panels self-explanatory without needing separate titles. Specifically:

1. We have removed the individual panel titles (a) Slope, (b) Roughness, (c) Groundpoint Density, (d) Canopy Height and instead label these directly on each x-axis, with the panel letters (a–d) retained only as small in-panel labels for cross-referencing in the text.
2. We have expanded the caption to clarify that shaded bands represent 95% confidence intervals and rug marks indicate the distribution of observations.
3. We have removed the two dashed reference lines at 2 and 10 pts m^{-2} from panel (c). These values were originally added as visual reference points based on apparent changes in the partial dependence curve rather than independently determined or statistically tested thresholds. We agree that their inclusion could imply specific threshold values that are not supported by the present analysis. The revised panel therefore presents the partial dependence relationship without these reference lines.

The revised figure now appears as follows:

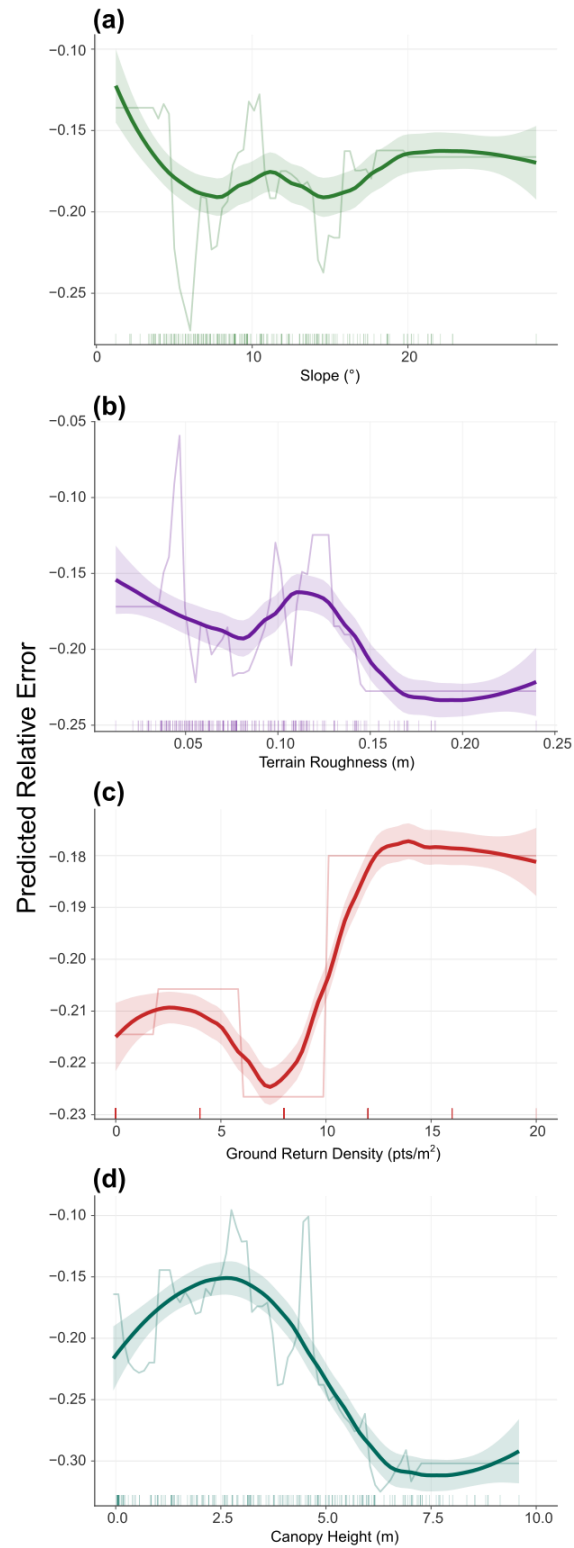


Figure 11. Partial dependence plots illustrating the influence of terrain slope, surface roughness, ground return density, and canopy height on LiDAR snow depth relative error. Shaded bands represent 95% confidence intervals, and rug marks indicate the distribution of observations.

Comment 16:

Section 5: In the Discussion section, I would like to see references to other studies conducted using aircraft or other SnowEx-campaigns. Currently, the accuracy measurements are being compared to the study by Jacobs et al., which was conducted using a drone. Once again, there are paragraphs without a single source. Section 5.2 begins with the claim: “Relatively few studies have directly quantified airborne LiDAR snow depth accuracy in dense forests,” which is not necessarily entirely true, as studies have indeed been conducted—admittedly not with a specific focus on densely forested areas—but there is nevertheless evidence that accuracy is reduced due to the influence of vegetation. In my opinion, this section lacks essential studies conducted after the 2020s. In the discussion section, I think the role of terrain variables in the analysis, as well as the chosen modeling method and its limitations, could be examined in greater detail. For example, regarding slope, there is no mention at all of the fact that slope direction is a proven significant factor in snow distribution. Furthermore, the small sample size is not addressed, nor is GBM, for that matter.

Response:

We agree with this comment and have revised the Discussion section.

1. Regarding the request for comparisons with other aircraft-based studies and SnowEx campaigns, we have added a comparison with Holland-Goon et al. (2026), who evaluated airborne lidar snow depth accuracy against in situ transects in boreal forest and tundra environments during the same NASA SnowEx 2023 Alaska campaign. They reported a mean bias of -0.16 m in boreal forest, consistent with the systematic underestimation under canopy that we observe here. We have also revised the opening statement of Section 5.2, which previously stated that relatively few studies have quantified airborne LiDAR snow depth accuracy in forests; this was overly broad, as several studies (e.g., Currier et al., 2019; Hopkinson et al., 2004) have addressed this question in other forested settings. The revised text now clarifies that relatively few such studies have been conducted specifically within boreal forest during dedicated remote-sensing campaigns comparable to ours, and incorporates this and other recent literature, including studies published after 2020 (e.g., Cho et al., 2021; May et al., 2025; Holland-Goon et al., 2026).

Several studies have quantified airborne LiDAR snow depth accuracy in forested environments (Currier et al., 2019; Hopkinson et al., 2004), but relatively few have done so specifically within boreal forest during dedicated remote-sensing campaigns comparable to ours. Using the same NASA SnowEx 2023 Alaska campaign, Holland-Goon et al. (2026) evaluated airborne lidar snow depth against in situ transects in boreal forest and tundra sites and reported a mean bias of -0.16 m in boreal forest, consistent with the systematic underestimation under canopy that we observe here. Our results align with these and other related findings.

2. We agree that the roles of terrain and canopy variables warranted clearer treatment in the Discussion. We have revised the text to distinguish between factors that control the natural spatial distribution of snow and factors that more directly affect LiDAR measurement accuracy. For terrain slope, we now note that it affects both processes: slope influences snow redistribution and retention (a natural-distribution effect), while it also directly affects LiDAR measurement geometry and can amplify elevation errors (a measurement-accuracy effect).

Revised manuscript (Page 36)

Terrain slope being a significant factor also echoes findings from previous topographic studies of snow distribution. While slope mainly influences snowpack through redistribution (steep slopes hold less snow and lose snow more easily), it also geometrically affects LiDAR accuracy. It is worth noting that slope aspect, rather than slope magnitude alone, is also a well-established driver of snow accumulation, with south-facing slopes typically retaining less snow than north-facing slopes at similar elevations due to differential solar radiation (Varhola et al., 2010; Guan et al., 2023). Aspect was not included as a predictor in our GBM model, so any influence it has on LiDAR error would be captured only indirectly through its correlation with slope and canopy structure. In photogrammetry, steep slopes cause larger errors due to viewing angle; similarly with LiDAR, the pulse footprint elongates on slopes and small GPS/INS errors translate to vertical errors on steep inclines. Our data showed $\sim 5\text{--}15$ cm worse errors on slopes $>20^\circ$. Jacobs et al. also noted that areas with steep slopes had higher snow depth confidence intervals and somewhat reduced LiDAR performance. Thus, both our study and others highlight slope as a risk factor for remote sensing of snow.

3. The referee also raises an important point regarding slope aspect. We agree that aspect is an important topographic control on the spatial distribution of snow, as differences in solar radiation, wind exposure, and energy balance can substantially affect snow accumulation and melt. However, the objective of the present analysis is to identify factors associated with the accuracy of LiDAR snow depth measurements rather than factors controlling the natural distribution of snow itself. Unlike slope, which can directly affect LiDAR measurement geometry, aspect does not have an equally direct relationship with LiDAR measurement accuracy; its influence is more likely to be indirect, acting through its effects on snow accumulation, melt, and vegetation conditions. We have therefore revised the Discussion to acknowledge the importance of aspect in controlling snow distribution while clarifying why it was not included as a predictor in the present analysis.

Revised manuscript(Page 36-37)

It is worth noting that slope aspect, rather than slope magnitude alone, is also a well-established driver of snow accumulation, with south-facing slopes typically retaining less snow than north-facing slopes at similar elevations due to differential solar radiation (Varhola et al., 2010; Guan et al., 2023). Aspect was not included as a predictor in our GBM model, so any influence it has on LiDAR error would be captured only indirectly through its correlation with slope and canopy structure.

4. We have expanded the discussion of modeling limitations in Section 5.3. The revised manuscript now acknowledges the modest site-level sample sizes relative to the number of predictors considered, the likely correlation among some predictors (e.g., canopy height and ground point density, or elevation and slope), and the sensitivity of GBM-derived variable importance to the specific predictor set used. We also clarify that variable importance should not be interpreted as evidence of a direct causal effect on LiDAR error, but rather as an indicator of a predictor's usefulness within the modeled feature space.

Revised manuscript (Page 30)

Fourth, several aspects of our modeling approach warrant caution in interpretation. Site-level sample sizes, particularly at FLCF, were modest relative to the number of predictors considered, which limits the precision of the site-specific GBM fits and may partly explain the lower R^2 obtained at that site. Several predictors are also likely correlated with one another in practice (for example, canopy height and ground point density, or elevation and slope), and this collinearity can redistribute importance among correlated variables in ways that do not necessarily reflect their true individual contributions. Finally, variable importance derived from GBM reflects the relative contribution of a predictor within the specific set of variables and feature space we selected, and it should not be interpreted as evidence of a direct causal effect on LiDAR error; a variable's high importance instead indicates that it is a useful marker for the underlying, and possibly confounded, drivers of measurement error.

Comment 17:

Section 5.3: Here is where I would expect the issues of the comment 16 to be discussed, but it focuses on addressing the accuracy of field measurements. Of course, there is uncertainty involved, but take, for example, the statement: "For example, under a tall tree, snow depth can vary within a few meters due to drip zones, microtopography, or vegetation cover. A single probe might read 40 cm, but just a few meters away the depth could be 30 cm."—it is true, but that is precisely why it is important to take many manual measurements, locate them precisely, and exactly why a combination of field + remote sensing data collection is important. Furthermore, at this point in the discussion, new information is presented regarding GPS accuracy, which should have been highlighted earlier in the "Data and Methods" section: I find it strange that GPS accuracy varies between 2.5 m and 15 m. Manual measurements should be very precise to serve as "ground truth," but here it seems that LiDAR-based data with 1-meter resolution is significantly more accurate than manual measurements with a range of 2–15 meters horizontal accuracy.

Response:

We thank the reviewer for raising this point. We have made two sets of revisions in response.

1. To ensure Section 5.3 aligns with our response to Comment 16, the expanded model-limitation discussion described above (Comment 16) covering sample size, predictor correlation, GBM importance sensitivity, and the causal-attribution caveat, will be presented as a dedicated subsection in Section 5.3, alongside the existing measurement-error content.
2. We agree that GPS horizontal accuracy should not be introduced for the first time in the Discussion. We have moved the statement on magnaprobe GPS accuracy (± 2.5 m in open areas, 10–15 m in dense forest, citing May et al., 2025) into Section 3.2 (Ground-based Snow Depth Measurements), where the ground-based measurement protocol is first described, and will retain a brief cross-reference to it in Section 5.3 where its implications for the comparison are discussed.

Regarding the reviewer's broader concern about whether manual measurements can serve as reliable "ground truth" given this positional uncertainty: we have clarified in the revised text that the depth recorded by the magnaprobe at each point is a direct physical measurement, independent of GPS accuracy, and is not degraded by positional error. What the GPS uncertainty affects is the spatial collocation between a given depth measurement and its corresponding LiDAR pixel. i.e., whether we are comparing the manual reading to the correct pixel, not whether the manual reading itself is accurate. We have stated this distinction explicitly, and we have discussed it as a contributor to the reduced correlation (R^2) between ground-based and lidar-derived snow depths, consistent with the sub-meter variability discussed elsewhere in Sect. 5.3. We have also noted that improving GPS accuracy (e.g., using RTK-GPS) in future field campaigns would reduce this source of mismatch and is a practical recommendation for future snow survey design.

Revised manuscript (Page 8) with key changes in bold italic

Extensive in-situ snow depth measurements were collected to validate and evaluate snow remote sensing datasets. NASA SnowEx23 field teams visited each study area around the time of the late winter LiDAR flight (early to mid March 2023) and measured snow depth using Magnaprobos. Within each site, multiple plots and transects were established to capture local variability. Measurement locations were recorded with Magnaprobe GPS units. *The GPS horizontal positioning accuracy for these units is on the order of ± 2.5 m in open areas and can increase to 10–15 m under dense forest canopy (May et al., 2025).*

Revised manuscript (Page 25-26, Lines 608 – 616) with key changes in bold italic

Third, *the GPS horizontal positioning accuracy of the Magnaprobe units (Sect. 3.2) may also degrade the comparison between in-situ and LiDAR data. It is important to note that this positional uncertainty does not affect the accuracy of the depth reading itself, which is a direct physical measurement taken independently of the GPS; rather, it affects the spatial collocation between a given depth measurement and its corresponding LiDAR pixel*, i.e., whether we are comparing the manual reading to the correct pixel rather than a nearby one. Given the high spatial variability of snow depth at sub-meter scales discussed above, this collocation mismatch is a plausible contributor to the reduced correlation (R2) between ground-based and lidar-derived snow depths. Improving GPS accuracy in future field campaigns, for example through real-time kinematic (RTK) GPS, would reduce this source of mismatch and is a practical recommendation for future snow survey design.

Technical corrections (non-exhaustive)

We thank the referee for the detailed technical corrections. We have addressed these items throughout the manuscript, including clarification of ambiguous sentences, removal of repeated abbreviation definitions, correction of duplicated sentences, consistent use of “Fig.”, clarification of map units and symbols, improved equation references, corrected axis labels and legends, removal of redundant caption text, and revision of informal expressions including at Lines 72-87, 111, 125, 131, 163, 165, 200, 217, 296, 311, 313, 334, 336, 342, 344, 349, 361, 407, 448, and 601..

We also carefully review the manuscript for similar issues beyond those specifically identified by the referee.

Line 39-40: I would clarify this sentence, for example, how do the terrain and vegetation affect the snow accumulation and therefore streamflow timing?

Response:

We thank the reviewer for this comment and we have clarified this sentence. Terrain influences snow accumulation through wind redistribution and through aspect-driven differences in solar radiation, with south-facing slopes typically experiencing earlier melt than north-facing slopes at similar elevations. Vegetation

influences accumulation primarily through canopy interception, which reduces the snow reaching the forest floor relative to open areas, and through canopy shading, which delays melt beneath dense forest cover relative to open or sparsely vegetated areas. These terrain- and vegetation-driven differences in the timing and spatial pattern of snow accumulation and melt desynchronize meltwater release across a watershed, which in turn lengthens and flattens the resulting streamflow hydrograph relative to a scenario of spatially uniform, synchronous snowmelt.

Revised manuscript (Page 2, Lines 34 – 50) with key changes in bold italic

Snow plays a critical role in global hydrological and ecological processes. It provides a crucial source of water for streamflow and reservoirs and controls the land-atmosphere energy fluxes (Li et al., 2017; Siirila-Woodburn et al., 2021; Sturm et al., 2017). Seasonal snowpack acts as a natural aquifer, releasing water during the snowmelt season and influencing global climate by its high albedo and insulating effect on the ground (Xu & Dirmeyer, 2013; Blau et al., 2024). In snow-dominant watersheds, a substantial portion of annual streamflow comes from snowmelt up to 75% in some Western North American basins, underscoring the importance of accurately measuring snow volume (Stewart et al., 2004; Fritze et al., 2011). Beyond total snow water storage, the spatial distribution of snow depth is also critical; heterogeneous snow cover (patchy or nonuniform snowpacks) strongly controls melt patterns, soil freeze/thaw dynamics, and ecological processes at the landscape scale (Lundquist & Dettinger, 2005). *For instance, terrain influences snow accumulation through wind redistribution and aspect driven differences in solar radiation, with south facing slopes typically experiencing earlier melt than north facing slopes at similar elevations, while vegetation influences accumulation primarily through canopy interception, which reduces the snow reaching the forest floor relative to open areas, and through canopy shading, which delays melt beneath dense forest cover. These terrain and vegetation driven differences in the timing and spatial pattern of snow accumulation and melt desynchronize meltwater release across a watershed, affecting streamflow timing and habitat conditions (Varhola et al., 2010; Guan et al., 2023).*

Line 49-51: What do you mean by “ground-based sensors cannot see the snow under the trees”?

Response:

We thank the reviewer for this comment. This was a typographical error: the sentence was intended to read "space-based" rather than "ground-based." Space-based and airborne optical and passive microwave sensors cannot directly observe the snow surface beneath dense forest canopy, because canopy elements obstruct the sensor's line of sight and dominate the measured signal, leading to systematic underestimation of snow depth or complete data gaps in densely forested areas. Ground-based point measurements are not affected by this limitation, since they physically measure snow depth directly at each sampled location rather than relying on a remote line of sight. We have revised this sentence in the Introduction to correct this error.

Line 52: Sources?

Response:

We thank the reviewer for this comment. This claim is supported by the citations that appear later in the same paragraph (Painter et al., 2016; Mazzotti et al., 2019; Deschamps-Berger et al., 2020; Currier et al., 2019; Deems et al., 2013), but their placement at the end of the following sentence made the attribution unclear. We have moved this citation to immediately follow the introductory sentence so that the source supporting this statement is unambiguous.

Line 91: What does it mean, when discontinuous permafrost affects snow accumulation and LiDAR performance?

These sentences could be clarified, and they would need decent sources.

Response:

We thank the reviewer for this comment and will revise this sentence for clarity. Cold climate, discontinuous permafrost, and forest structure collectively influence snow accumulation patterns in boreal forests. Dense conifer canopy specifically, rather than permafrost itself, is the factor that limits LiDAR's ability to penetrate to the underlying snow surface, consistent with the canopy-related mechanisms already discussed earlier in this section. We have revised the sentence to attribute these two effects to their respective causes rather than implying that all three characteristics equally affect both snow accumulation and LiDAR performance.

Revised manuscript (Page 3-4, Lines 100 – 110) with key changes in bold italic

Despite the recognized potential of airborne LiDAR, there is a need for systematic assessment of its accuracy

in forested regions and an understanding of what conditions lead to larger errors. Accuracy evaluations of LiDAR-based snow depth mapping have been conducted across a range of platforms and forest conditions, though predominantly outside boreal environments. Airplane- and unpiloted aerial system-based studies have validated LiDAR snow depth performance in alpine and temperate forests (Hopkinson et al., 2004; Jacobs et al., 2021), helicopter-based systems have been used to assess LiDAR performance in steep, forested terrain (Bühler et al., 2015; Wilder et al., 2025), and terrestrial laser scanning studies have specifically examined the influence of forest canopy structure on snow depth retrieval accuracy (Prokop, 2008; Revuelto et al., 2015; Hojatimalekshah et al., 2021). Relatively few of these evaluations, however, have focused on boreal forests, which are characterized by cold climates, discontinuous permafrost, and dense conifer canopies. *Cold climate and discontinuous permafrost, together with the associated mosaic of forested and open ground cover that they support, shape snow accumulation patterns across these landscapes (Jorgenson and Osterkamp, 2005; Bouchard et al., 2024).* Dense conifer canopy, specifically, limits LiDAR's ability to penetrate to the underlying snow surface (Gustafson et al., 2010; Dadic et al., 2010). Together, these characteristics make boreal forests a distinct and understudied environment in snow remote sensing (NASA, 2018; May et al., 2025).

Line 93: Since not all readers may know what the NASA SnowEX 2023 campaign is, I think it would be helpful to describe it here, at least briefly.

Response:

We thank the reviewer for this suggestion and we have added a brief description here. NASA SnowEx is a multi-year airborne and ground-based snow measurement campaign designed to evaluate and improve snow remote sensing methods across diverse snow classes and land cover types; the 2023 Alaska campaign specifically targeted boreal forest and tundra environments in interior Alaska.

Revised manuscript (Page 4, Lines 111 – 116) with key changes in bold italic

NASA SnowEx is a multi-year airborne and ground-based snow measurement campaign designed to evaluate and improve snow remote sensing methods across diverse snow classes and land cover types. The 2023 Alaska campaign, which specifically targeted boreal forest and tundra environments in interior Alaska, provides an

ideal opportunity to evaluate LiDAR snow depth retrievals in this setting across varying conditions. By combining airborne LiDAR data with an extensive set of ground truth measurements and environmental variables, we can quantify how well LiDAR performs and what drives any discrepancies.

Line 116: What are those little red dots in the maps? What is the m? Elevation (from sea level?)? I'd also recommend removing the repetition from the image and the caption—right now the names are written in both.

Response:

We thank the reviewer for this comment and will address these points in the revised figure. The red dots in each panel mark the individual field survey subregions within each study site where snow depth measurements were collected. The "m" in the elevation legend refers to elevation in meters above sea level; we have labeled this as "Elevation (m a.s.l.)" in the revised figure to avoid ambiguity. We have also removed the redundant site-name labels currently duplicated between the image panels and the figure caption, retaining the names in only one location.

The revised figure now appears as follows:

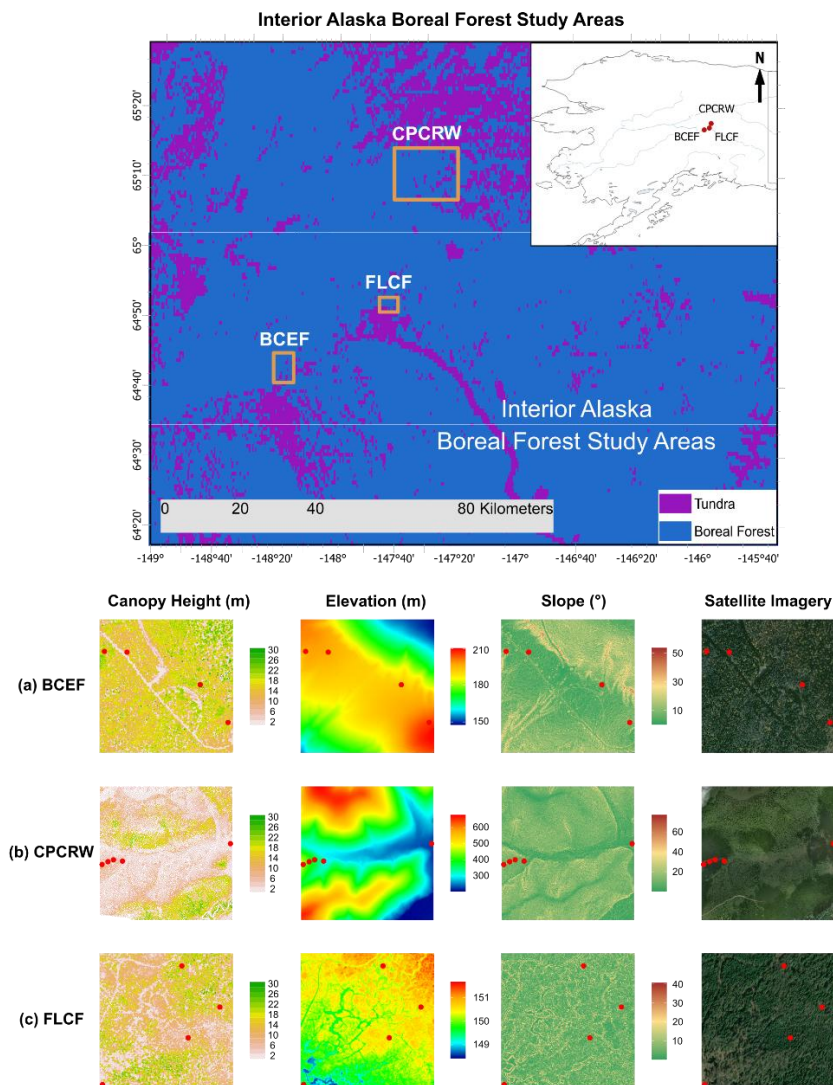


Figure 1. Location of the three study areas within interior Alaska's boreal forest and tundra zones, with orange boxes indicating the zoomed-in sub-regions used for detailed environmental characterization. Inset shows the regional context within Alaska. Below: canopy height, elevation, slope, and satellite imagery within the zoomed sub-region of each site (a) BCF, (b) CPCRW, (c) FLCF derived from NASA SnowEx23 airborne LiDAR datasets. Canopy height was calculated by differencing the DSM and DTM; elevation and slope were extracted from the SnowEx-derived DTM. Red dots indicate the locations of manual snow depth measurement sub-regions within each study site.

Line 135: Maybe a new section for the climatic conditions?

Response:

We thank the reviewer for this suggestion and will add a new subsection for the climatic conditions.

788 **Line 160:** dates for May 2022?

789 **Response:**

790 We thank the referee for this comment. Upon review of the NSIDC SnowEx23 airborne lidar data records
791 (https://nsidc.org/data/snex23_lidar_raw/versions/1), we confirm the precise dates of the snow-off lidar
792 acquisition: BCEF and FLCF were surveyed on 28 May 2022, and CPCRW was surveyed on 29 May 2022. We
793 have replaced the general "May 2022" reference in the manuscript with these precise acquisition dates.

794

795 **Revised manuscript (Page 7)**

796 Two LiDAR acquisitions were used: a snow-off flight conducted on 28 May 2022 at BCEF and FLCF and 29
797 May 2022 at CPCRW (bare ground baseline) and a snow-on flight on 11 March 2023 (late winter, peak
798 snowpack).