

Responses to Comments from Reviewer #2

Liu et al. investigate lidar snow depth error from lidar and manual samples collected during the NASA SnowEx 2023 campaign. The authors investigate the sources of lidar error using canopy height, surface roughness, point density, elevation, and slope. They also generate a machine learning framework to estimate expected error using these variables. This work is an important contribution to the field given the emergence of lidar used for validation, while lidar itself is difficult to validate.

Prior to publication, the presentation of the results should be improved and several conclusions should be softened. The manuscript also currently lacks any discussion of in-situ sampling bias, which may be influencing the reported results. I believe this manuscript will be a valuable contribution to the field, but it requires significant revision before publication.

We sincerely thank Referee #2 for the thorough and constructive review of our manuscript, and for recognizing the value of this study given the growing use of LiDAR for snow depth validation. We agree that the presentation of the results should be strengthened, that several conclusions should be stated more cautiously, and that the manuscript does not yet adequately discuss the potential for bias in the in situ measurements used as ground truth. We address each point below.

General comment

General comment 1: A critical piece missing from this manuscript is a discussion of potential bias in the in situ measurements used as ground truth. Given that this manuscript reports a systematic negative bias in the LiDAR estimates, the positive bias reported by Stuefer et al. (2025) using the same SnowEx dataset should be discussed in much greater detail, and the authors should evaluate how this known sampling bias affects their conclusions, including by repeating the analyses after correcting for the reported 10 cm bias.

Response:

We thank the referee for this comment. To directly test how sensitive our results are to the known measurement bias in the Magnaprobe ground truth, we repeated our primary accuracy assessment after subtracting 10 cm from

all in-situ measurements at BCEF, CPCRW, and FLCF, consistent with the positive Magnaprobe bias reported by Stuefer et al. (2025) in boreal forest plots. After this correction, LiDAR bias improved substantially but did not disappear. Pooled bias decreased from about -0.17 m to about -0.07 m ($N = 1,012$); bias also improved at BCEF (-0.18 to -0.08 m, $N = 207$), CPCRW (-0.19 to -0.09 m, $N = 250$), and FLCF (-0.16 to -0.06 m, $N = 555$). MAD and RMSE decreased proportionally at all sites, while Pearson correlation coefficients were unaffected. These results indicate that roughly half of the LiDAR's apparent negative bias at BCEF and CPCRW is attributable to the known positive bias in the Magnaprobe ground truth rather than to LiDAR measurement error, while a residual bias of 6 to 9 cm remains at all three sites after correction. We have added this sensitivity analysis to Section 5.3 and the Supplementary Material (Figure S6) and now cite the Stuefer et al. (2025) results explicitly as a quantified source of ground-truth uncertainty in section 3.2 Ground-based Snow Depth Measurements, per the referee's suggestion.

Revised manuscript (Page 8)

Stuefer et al. (2025) quantified Magnaprobe measurement uncertainty at the same SnowEx Alaska sites by comparing 1,070 co-located Magnaprobe and excavated snow-depth-profile measurements in boreal forest plots, finding a systematic positive bias of approximately 10 cm and a root-mean-square error of 12.1 cm ($R^2 = 0.46$). We adopt these boreal forest values as the formal measurement uncertainty of our in-situ ground truth throughout this study.

Revised manuscript (Page 30)

To evaluate how this known ground-truth bias affects our conclusions, we repeated the accuracy assessment reported in Section 4.1 after subtracting 10 cm from every in-situ measurement at all three sites, consistent with the boreal forest bias reported by Stuefer et al. (2025). This correction reduced the pooled bias from about -0.17 m to about -0.07 m, with MAD (median) decreasing from about 0.17 m to about 0.10 m and RMSE from about 0.22 m to about 0.16 m; correlation coefficients were unchanged, as expected for a uniform additive shift. At the site level, bias decreased from about -0.18 m to about -0.08 m at BCEF and from about -0.19 m to about -0.09 m at CPCRW, indicating that approximately half of the apparent LiDAR underestimation at these two sites is attributable to the known positive bias in the Magnaprobe ground truth rather than to LiDAR measurement error. A residual negative bias of 6 to 9 cm remains at all three sites after correction, indicating that LiDAR

underestimation under canopy is not fully explained by this single, known source of ground-truth bias. Full results of this sensitivity analysis are reported in Supplementary Table S6.

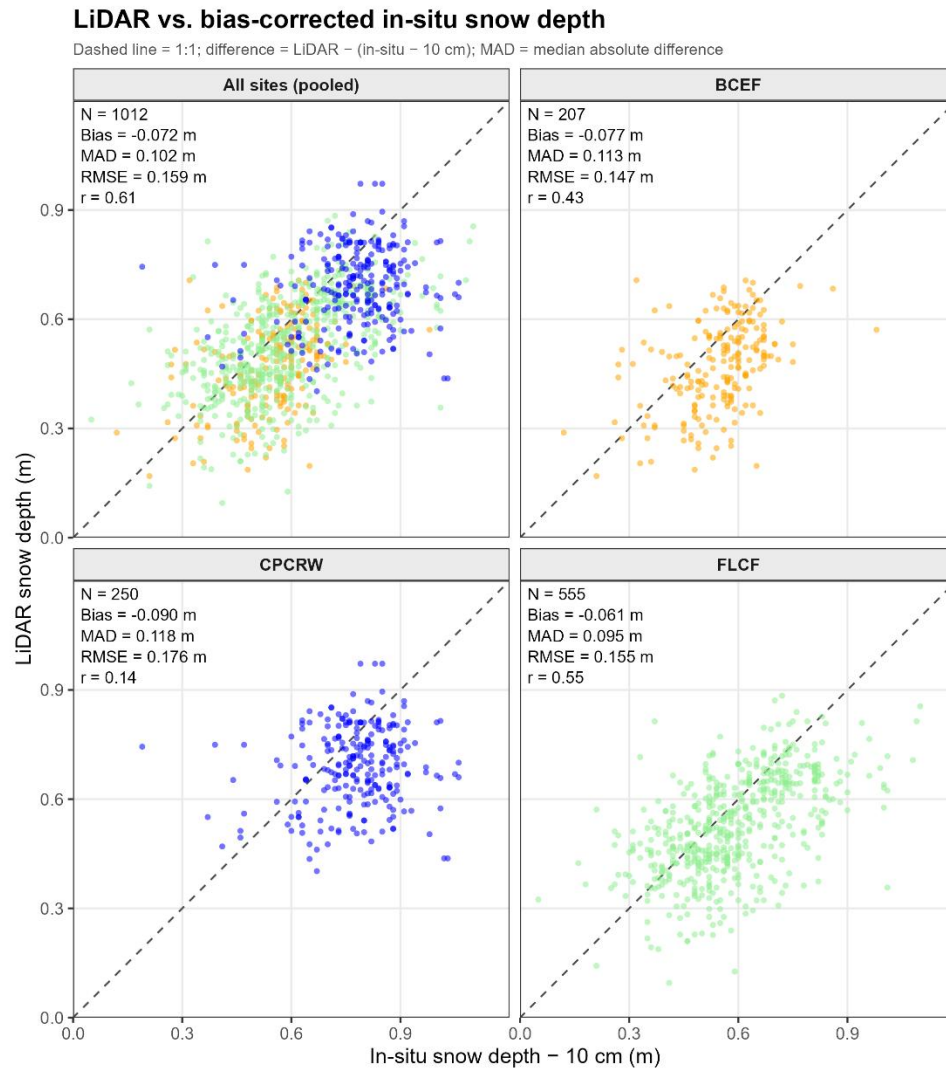


Figure S6. Comparison of LiDAR-derived and Magnaprobe-derived snow depth for all sites pooled and for each site individually (BCEF, orange; CPCRW, blue; FLCF, green), after subtracting 10 cm from each in-situ measurement to account for the positive Magnaprobe bias reported by Stuefer et al. (2025) in boreal forest plots. The dashed line is the 1:1 line; points falling below the line indicate LiDAR underestimation. Each panel reports the bias (mean error, LiDAR minus bias-corrected in-situ), root mean square error (RMSE), median absolute difference (MAD), Pearson correlation coefficient (r), and the number of paired observations (N).

General comment 2: Several conclusions are stronger than the presented evidence supports. Figures 4 to 6 do not clearly show that larger errors occur beneath dense canopy, and the evidence for increasing error with slope is inconsistent across sites. The discussion should better acknowledge that these variables explain only part of the observed error and that other factors may contribute substantially.

Response:

We agree and have revised the Discussion to specify that canopy height is the dominant predictor among the five variables tested rather than the dominant control on LiDAR error in general, and to emphasize that the site-level R^2 values (0.18 to 0.53) leave a substantial fraction of the error unexplained. We have expanded the discussion of other possible contributors, including in situ measurement uncertainty, snow-off LiDAR uncertainty, sub-canopy vegetation effects, and GPS geolocation error. We have also revised the slope discussion, which previously compared gentle-slope conditions at CPCRW to steep-slope conditions at FLCF, to present the slope–error relationship separately for each site and state plainly that the increase in error with slope is clear at FLCF but not consistently observed at BCEF or CPCRW. We have reduced the number of maps in Figures 4 to 6 and added a quantitative panel showing canopy height plotted directly against absolute error at each site (see response to the reviewer's comment on Figures 4–6 below).

General comment 3: If the predictive error of the GBM model is similar to the observed LiDAR error itself, the practical utility of the model is unclear. The more important conclusion may be that LiDAR errors are difficult to predict using the available environmental variables. Testing snow depth as an additional predictor is also suggested.

Response:

We have tested snow depth as an additional predictor for both the relative- and absolute-error GBM formulations and report the results in our response to Comment 11 below; in situ snow depth emerged as by far the most important predictor at all three sites. We will directly compare the RMSE of the GBM-predicted error to the magnitude of the observed LiDAR error at each site and reframe the conclusion to state that LiDAR errors are difficult to predict from the available environmental variables, rather than emphasizing predictive skill.

General comment 4: The manuscript currently focuses on maps of the study areas, which are not especially effective at supporting the conclusions. Several additional summary figures, such as site specific error statistics, probability density functions, and predictor distributions, would help. The systematic negative bias appears to be one of the primary results, yet it receives much less emphasis than correlation or MAD.

Response:

We have added a summary table reporting sample size, bias, MAD, RMSE, and correlation for each site and the pooled dataset, with a breakdown of bias and MAD for open versus forested conditions, so that bias is reported alongside correlation and MAD throughout the manuscript rather than as a secondary statistic, consistent with the referee's assessment that the systematic negative bias is one of the primary results of this study.

Comment 1:

Title: Consider adding in situ / manual sampling to the title, since this is the other key dataset aside from LiDAR.

Response:

We appreciate this suggestion. We have revised the title to reflect that the manuscript relies equally on in situ manual snow depth measurements and airborne LiDAR data.

Proposed revised title:

Performance and Controlling Factors of Airborne LiDAR Snow Depth Estimates Evaluated against In Situ Measurements in Boreal Forests: Insights from the NASA SnowEx 2023 Alaska Campaign

Comment 2:

19: Change ‘showed’ to present tense

Response:

We thank the reviewer for pointing this out. We have changed “showed” to “show” to maintain the present tense in the Abstract.

Comment 3:

41-51: Starting a paragraph with ‘however’ feels a bit awkward. Also, the focus of this paragraph could be improved. The first half is about snow sampling methods and the second half is about difficulty of sampling in forests. Maybe split into two paragraphs or improve the topic sentence / transitions.

Response:

We thank the reviewer for this helpful suggestion. We have revised and divided the original paragraph into two paragraphs. The first paragraph now discusses the limitations of in situ and conventional remote sensing methods for measuring snow depth distribution, while the second focuses on the specific challenges of measuring snow in forested environments. We have also removed the awkward opening transition and added a clearer topic sentence to introduce the shift in focus.

Revised manuscript (Page 2)

Measuring snow depth distribution across large and remote areas remains challenging with existing methods. In situ methods, such as snow probes and snow courses, provide accurate point measurements but are labor-intensive and limited in spatial coverage (Sturm et al., 2015). Remote sensing provides an effective means of extending snow depth measurements over broader areas, and recent advances have enabled the use of spaceborne sensors such as ICESat-2 altimetry and Sentinel-1 C-band SAR for snow depth mapping. These approaches have demonstrated promising performance across a range of environments (Deschamps-Berger et al., 2023; Dunmire et al., 2024; Hoppinen et al., 2024; Fair et al., 2025). However, challenges remain in forested environments, where vegetation can interfere with laser returns and SAR signals, increasing uncertainty in snow depth retrievals (Hoppinen et al., 2024; Fair et al., 2025).

Measuring snow depth in forested environments is particularly challenging because tree canopies intercept laser returns and obscure the snow surface beneath vegetation (Lefsky et al., 2002). In addition, tree canopies can intercept a substantial portion of snowfall (30–80% in dense conifer forests) before it reaches the ground, contributing to shallower and more spatially variable snow depths than in open areas (Lumbrazo et al., 2022; Dharmadasa et al., 2024). These challenges highlight the need for high-resolution observations that can better characterize snow depth variability beneath and around forest canopies. Airborne LiDAR provides such an approach by enabling dense, spatially continuous measurements of snow surface elevation over relatively large areas and has been increasingly used to characterize and validate snow depth in forested environments (Deschamps-Berger et al., 2023; Hoppinen et al., 2024; Fair et al., 2025).

Comment 4:

56-57: Does this need a citation?

Response:

We thank the reviewer for pointing this out. We have added Lefsky et al. (2002), a foundational reference in LiDAR remote sensing.

Revised manuscript:

Forested environments pose particular difficulties for remote sensing because tree canopies intercept laser returns and complicate the characterization of surfaces beneath vegetation (Lefsky et al., 2002).

Comment 5:

62-63: ‘ability to capture snow depth spatial variability that is unattainable with traditional methods’ is vague.

Response:

We thank the reviewer for this suggestion. We agree this phrase was vague and have revised it to describe the specific advantage of LiDAR more concretely.

Revised manuscript (Page 3, Lines 72 – 75) with key changes in bold italic

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.*

Comment 6:

82: Awkward transition between sentences. Introduce that you are now talking about vegetation effects.

Response:

Thank you for pointing this out. We have revised the transition into this paragraph to explicitly introduce that the discussion is turning to the effect of vegetation on LiDAR performance.

Revised manuscript (Page 3, Lines 86 – 89) with key changes in bold italic

Beyond the effects of platform type and flight altitude discussed above, vegetation cover is a key additional concern for LiDAR snow depth accuracy in boreal forest environments (May et al., 2025). Tall canopy and complex understory can cause LiDAR to underestimate snow depth because some fraction of laser pulses are intercepted by branches or attenuated, resulting in fewer ground returns.

Comment 7:

86-87: ‘even small absolute errors can be a large relative fraction’. I would remove this. This is an error metric / statistics phenomenon, not an issue with snow sampling.

Response:

We agree this statement describes a property of the error metric rather than an issue with snow sampling itself, and we have removed it from the Introduction. The underlying issue, that relative error can become unstable for very shallow snow depths, is addressed directly in the Methods (Section 3.5) and is discussed further in our responses below to the related comments on lines 269 to 270 and 273.

Revised manuscript (Page 3, Lines 86 – 89) with key changes in bold italic

In relatively open environments, elevation and wind redistribution lead to deeper snow at higher elevations or in lee aspects, whereas forests reduce snow accumulation via canopy interception and sublimation losses (Gustafson et al., 2010; Dadic et al., 2010). *These same factors likely influence the agreement between LiDAR and ground measurements: where terrain is steep or highly irregular, co-registration errors can occur.*

Comment 8:

Figure 1: label the color bars as elevation

Response:

We thank the reviewer for this suggestion. We have combined Figures 1 and 2 into a single figure. In the revised figure, the elevation panels have their color bars explicitly labeled.

The revised figure now appears as follows:

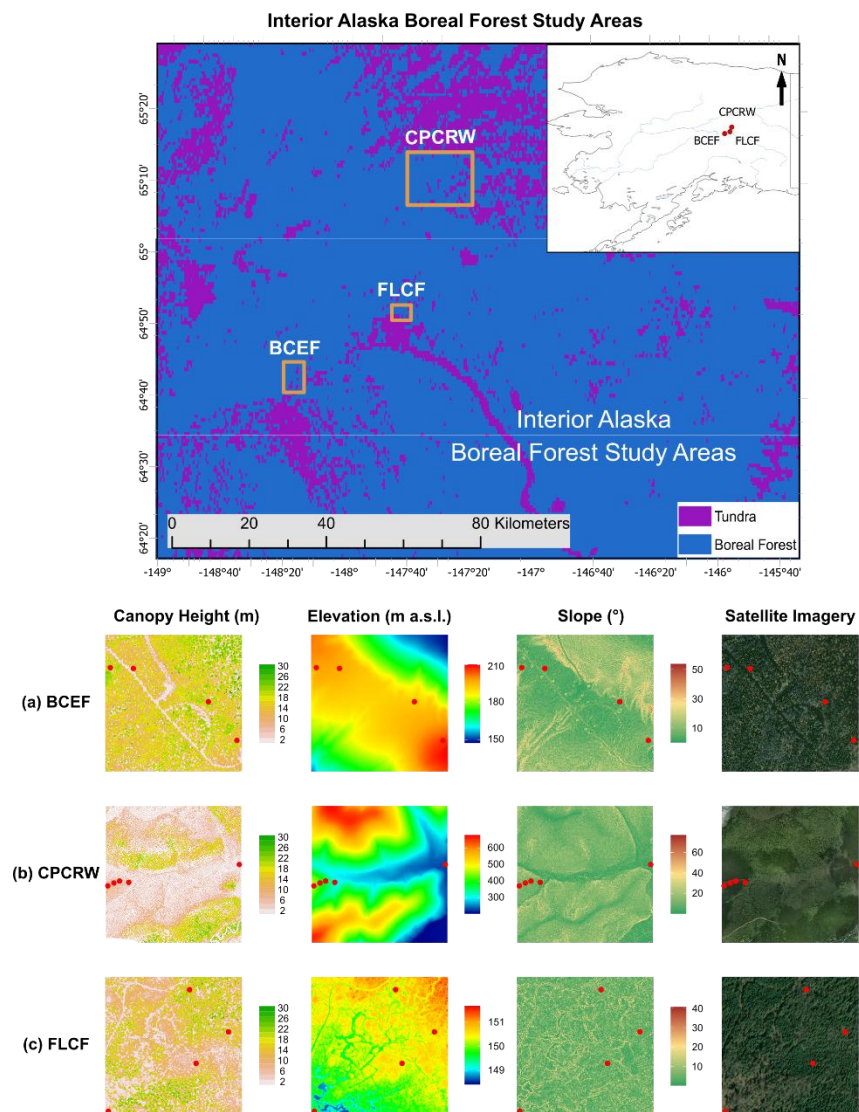


Figure 1. Location of the three study areas within interior Alaska's boreal forest and tundra zones, with orange boxes indicating the zoomed-in sub-regions used for detailed environmental characterization. Inset shows the

regional context within Alaska. Below: canopy height, elevation, slope, and satellite imagery within the zoomed sub-region of each site (a) BCEF, (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.

Comment 9:

Figure 2: Could you add a satellite image of the sites here or in figure 1? Would be helpful to conceptualize this sites more.

Response:

We thank the reviewer for this helpful suggestion. As described in our response to Comment 8, we have merged the original Figure 2 into the revised Figure 1 and added satellite imagery panels for each study site to provide clearer visual context. The revised Figure 1 is presented in our response to Comment 8 above and is not repeated here.

Comment 10:

250: Suggest adding bias as a metric.

Response:

We thank the reviewer for this suggestion. Bias (mean error) is already computed as part of the accuracy assessment described in Section 3.4. We have revised the text in Section 4.1 to report bias directly alongside Pearson r , RMSE, and MAD for the pooled dataset and each site (pooled bias about -0.17 m; BCEF about -0.18 m; CPCRW about -0.19 m; FLCF about -0.16 m), rather than only for BCEF and CPCRW as in the original submission. As described in our response to General comment 4 above, we have additionally included these values, together with a breakdown of bias and MAD for open versus forested conditions, in an expanded summary table.

Revised manuscript (Page 11 - 15, Lines 320 – 366) with key changes in bold italic

The airborne LiDAR surveys successfully mapped snow depth across all three areas, but with some differences in accuracy when validated against in-situ Magnaprobe observations. Overall, considering all available paired observations (1,012 points) pooled from the three study sites, the LiDAR-derived snow depths showed a moderate correlation with in-situ measurements (Pearson $R = 0.61$). This pooled correlation was slightly higher than the strongest site-specific correlation observed in FLCF ($R = 0.55$), likely reflecting the larger sample size and broader range of environmental conditions represented in the combined dataset. ***The MAD and RMSE were 0.17 m and 0.22 m, respectively, and the overall bias was about -0.17 m, indicating that LiDAR generally underestimated snow depth.*** These underestimations were consistent across three areas (Figure 3).

When examining the sites individually, some differences emerge. Among the three sites, FLCF exhibited the strongest agreement between LiDAR and ground measurements. ***The correlation at FLCF was around 0.55, MAD was 0.16 m and RMSE was 0.22m and bias was about -0.16 m.*** Points at FLCF were almost all located in areas characterized by a mixture of open and wooded cover over flat to gently rolling terrain where LiDAR performed very well, raising the site's overall accuracy. Under forested areas identified by higher canopy heights in Figure 4b, errors were generally larger than those observed in adjacent open areas. (Figure 4).

In BCEF, LiDAR tended to underestimate snow depth under mature spruce stands. The correlation was lower ($R = 0.43$), and MAD was 0.18 m. We observed that in some forest plot locations at BCEF, ground-based snow depths around 50 cm corresponded to LiDAR estimates of only ~30 cm, a substantial underestimation. In more open parts of BCEF, LiDAR was accurate. This variability lowered the overall correlation. ***The bias at BCEF was about -18 cm*** (LiDAR underestimating on average) (Figure 5).

CPCRW had the lowest raw correlation ($R = 0.14$) and the largest errors in relative terms ($RMSE = 0.24$ m). This site had the most challenging conditions: steep slopes and dense spruce forest in many areas. Notably, a number of CPCRW points on slopes under dense canopy showed LiDAR snow depths less than half of the actual measured depth (e.g., LiDAR ~57 cm vs actual ~110 cm). ***The bias at CPCRW was around -19 cm (underestimate). Despite these errors, LiDAR still captured general patterns: Errors tended to be larger in areas characterized by dense canopy and complex terrain. (Figure 6).***

Comment 11:

269-270: Did you attempt the ML framework with absolute errors instead of relative? I think this could potentially perform better, especially if providing snow depth as a predictor.

Response:

We thank the reviewer for this suggestion. As described in our response to General comment 3 above, we ran the GBM framework using absolute error (LiDAR minus in situ, in meters) as the response variable, in addition to relative error, and included in situ snow depth as an additional sixth predictor alongside canopy height, slope, roughness, elevation, and ground point density.

We plan to include the feature importance results (Figures S1–S2) in the Supplementary Material. In the main text, we will briefly reference this analysis in Section 5.1, immediately following the discussion of canopy height as the dominant predictor, noting that in situ snow depth emerges as an even stronger predictor of absolute LiDAR error when included in the model (see Supplementary Fig. S2), consistent with the negative correlation between canopy height and snow depth reported above.

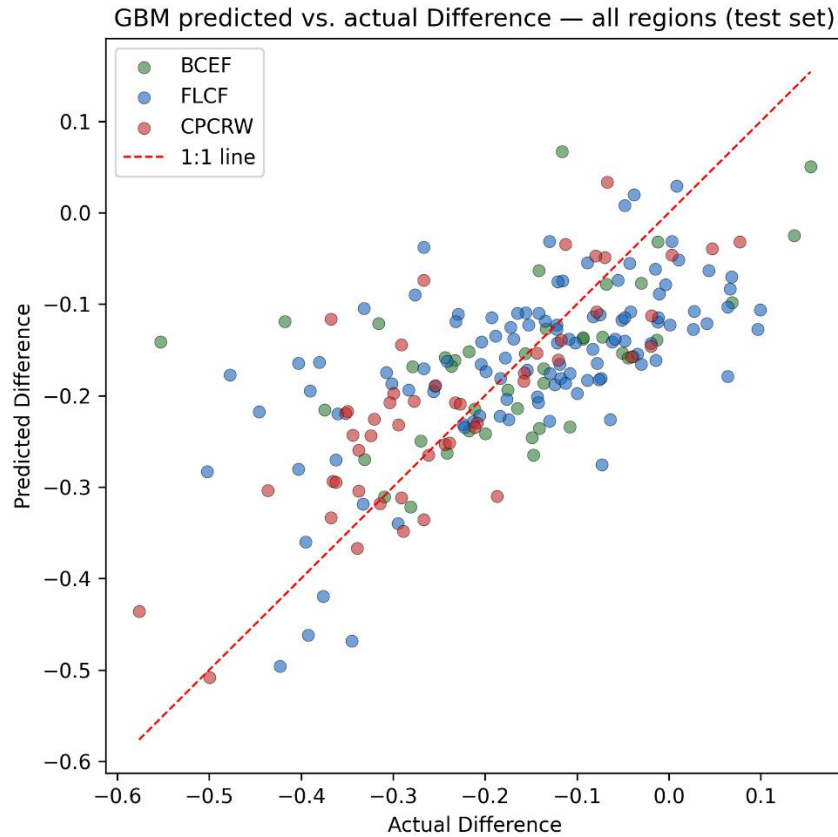


Figure S1. Predicted versus actual LiDAR error (Difference = LiDAR minus in situ snow depth, in meters) from the absolute-error GBM model, pooled across all three sites (test set only). The dashed line indicates 1:1 agreement between predicted and actual values.

Model performance changed unevenly across sites relative to the original relative-error models: R2 improved at CPRW (0.527 to 0.599) and improved substantially at FLCF (0.18 to 0.368), but declined at BCEF (0.33 to 0.257). We therefore do not find that the absolute-error formulation uniformly outperforms the relative-error formulation, though it does perform better at two of the three sites.

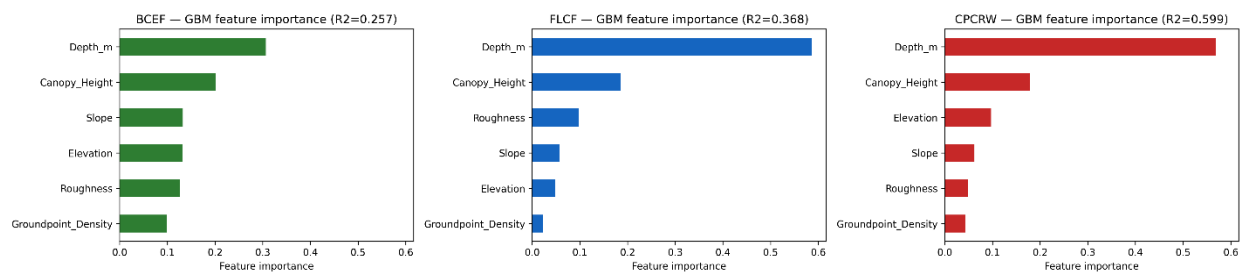


Figure S2. Relative importance of environmental predictors in the absolute-error GBM model for (a) BCEF, (b)

FLCF, and (c) CPCRW. Predictors include in situ snow depth (*Depth_m*), canopy height, elevation, slope, roughness, and ground point density. The response variable is absolute LiDAR error (LiDAR minus in situ snow depth, in meters). *R*² values are test-set coefficients of determination for each site-specific model.

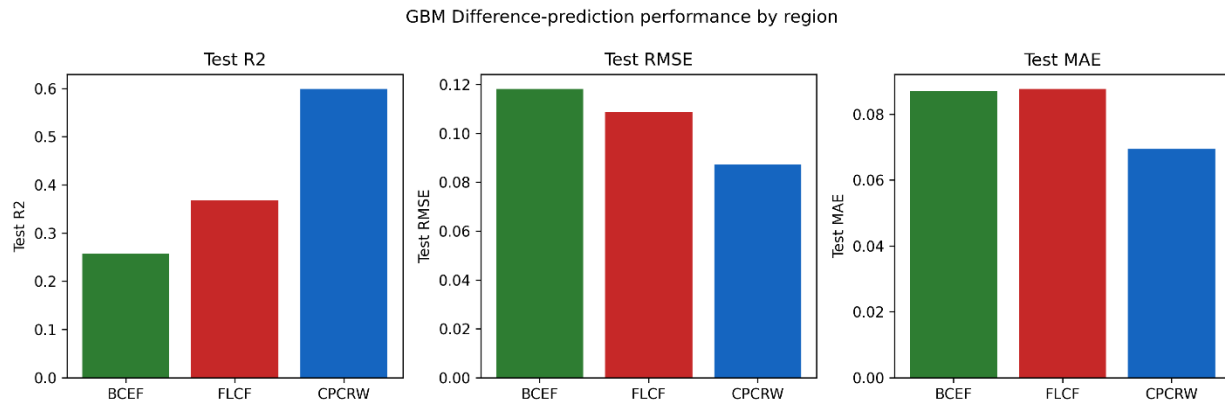


Figure S3. Test-set performance of the absolute-error GBM model by site, showing the coefficient of determination (*R*²), root mean square error (RMSE, *m*), and mean absolute error (MAE, *m*).

Including snow depth as a predictor had a much larger effect: at all three sites, in situ snow depth emerged as by far the most important predictor of absolute error, with an importance roughly two to three times that of canopy height, which remained the second most important predictor at every site. We interpret this as consistent with our finding (reported in response to Comment 15 above) that canopy height and in situ snow depth are themselves correlated (Pearson $r = -0.52$), such that canopy height's influence on LiDAR error may act partly through its relationship with the underlying snow depth. We have included the feature importance results from this analysis as a supplementary figure, and have discussed this finding in the revised Discussion, specifically noting that snow depth itself is the strongest predictor of absolute LiDAR error once it is included in the model.

Comment 12:

273: ‘blowing up the fraction’ seems informal. How about ‘disproportionately amplifying relative error magnitudes’

Response:

We thank the reviewer for this suggestion. We have replaced this phrase as suggested.

Comment 13:

Figure 3: Consider reporting statistics separately for each site, either by splitting the figure into three panels or by adding a table. Bias should also be included as a summary statistic. The systematic negative bias appears to be one of the primary findings.

Response:

We thank the reviewer for these suggestions. We have revised Figure 3 by splitting it into four panels: the pooled comparison across all sites, followed by separate panels for BCEF, CPCRW, and FLCF. Bias is now reported alongside Pearson r , RMSE, MAD, and sample size in each panel. The revised figure makes clear that a systematic negative bias (about -0.16 to -0.19 m) is present across all three sites, consistent with the referee's assessment that this is one of the primary findings of the study.

The revised figure now appears as follows:

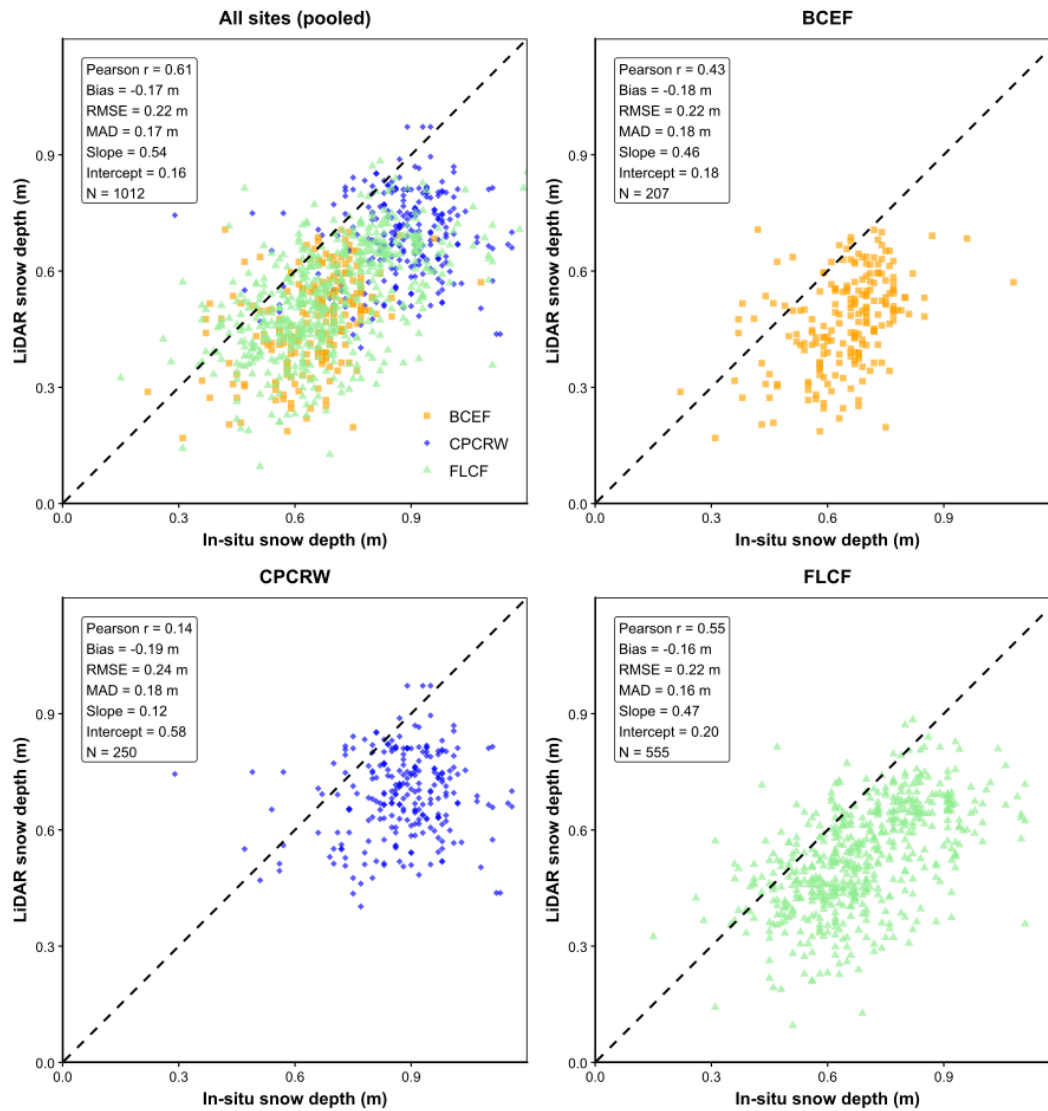


Figure 2. Comparison of LiDAR-derived and in-situ measured snow depth for all sites pooled and for each site individually (BCEF, orange squares; CPCRW, blue diamonds; FLCF, green triangles). The dashed line is the 1:1 line; points falling below the line indicate LiDAR underestimation, and points above indicate LiDAR overestimation. Each panel reports the Pearson correlation coefficient (r), bias (mean error, LiDAR minus in-situ), root mean square error (RMSE), mean absolute difference (MAD), the slope and intercept of the least-squares regression line, and the number of paired observations (N).

Comment 14:

326-327: Figure 6 doesn't explicitly support this claim

Response:

We agree that Figure 6 did not explicitly demonstrate this claim, and have revised the text to more accurately describe what the figure shows, without asserting a general pattern that the figure does not support.

The revised passage reads:

CPCRW had the lowest raw correlation ($R = 0.14$) and the largest errors in relative terms ($RMSE = 0.24$ m). This site had the most challenging conditions, with steep slopes and dense spruce forest present in many areas. A number of CPCRW points located on slopes beneath dense canopy showed LiDAR snow depths less than half of the corresponding measured depth. The overall bias at CPCRW was around -19 cm (an underestimate). As shown in Figure 6, the largest errors at CPCRW were concentrated among a relatively small number of points rather than distributed evenly across the site, and several of these high-error points coincide with locations of dense canopy and steep terrain. However, similarly forested and sloped locations elsewhere in the study area did not consistently show elevated errors, so this correspondence should be read as illustrative of specific problem locations rather than as evidence of a general relationship. A more systematic, quantitative assessment of how canopy density and slope relate to error magnitude across all three study sites is presented in Section 4.2.

Comment 15:

Figure 4-6. These figures provide useful context but do not clearly support the conclusions regarding canopy effects. Consider combining them into a single summary figure and replacing the remaining space with quantitative analyses, such as canopy height versus error.

Response:

We thank the reviewer for this suggestion. We have combined Figures 4, 5, and 6 into a single summary figure (new Figure 4), presenting the first subregion of each site across BCEF, CPRW, and FLCF in a unified multi-panel layout, rather than three separate figures. Section 4.1 focuses on evaluating overall LiDAR accuracy, while the quantitative relationship between canopy height and LiDAR snow depth error is addressed in detail in Section 4.2 (Physical Influence on Error), including logistic regression results (Figure 8), the distribution of relative error across canopy height categories (Table 1), and GBM-based variable importance rankings (Figure 10).

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

Canopy Height	<i>n</i>	<i>High Error, n (%)</i>	<i>0–0.25</i>	<i>0.25–0.5</i>	<i>0.5–0.75</i>	<i>≥0.75</i>	<i>OR (95% CI)</i>
<i>0–5 m</i>	366	10 (2.7%)	67.5%	29.8%	2.2%	0.5%	1.00 (<i>ref</i>)
<i>5–10 m</i>	234	8 (3.4%)	41.9%	54.7%	3.4%	0.0%	1.26 <i>ns</i> (0.47–3.24)
<i>10–15 m</i>	165	25 (15.2%)	41.2%	43.6%	14.5%	0.6%	6.36 *** (3.06–14.20)
<i>15–20 m</i>	204	21 (10.3%)	48.0%	41.7%	9.3%	1.0%	4.09 *** (1.93–9.23)
<i>≥20 m</i>	43	6 (14.0%)	51.2%	34.9%	14.0%	0.0%	5.77 ** (1.87–16.48)

The revised figure now appears as follows:

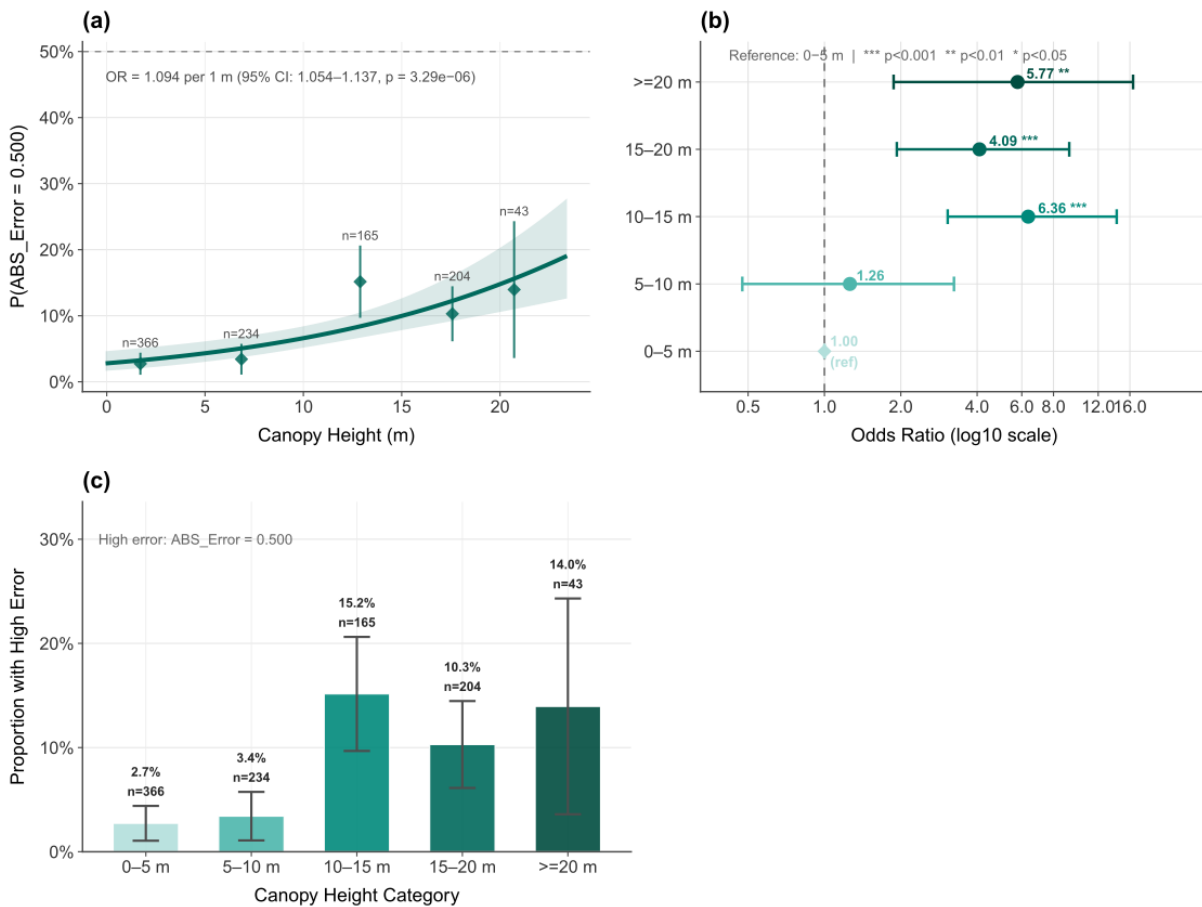
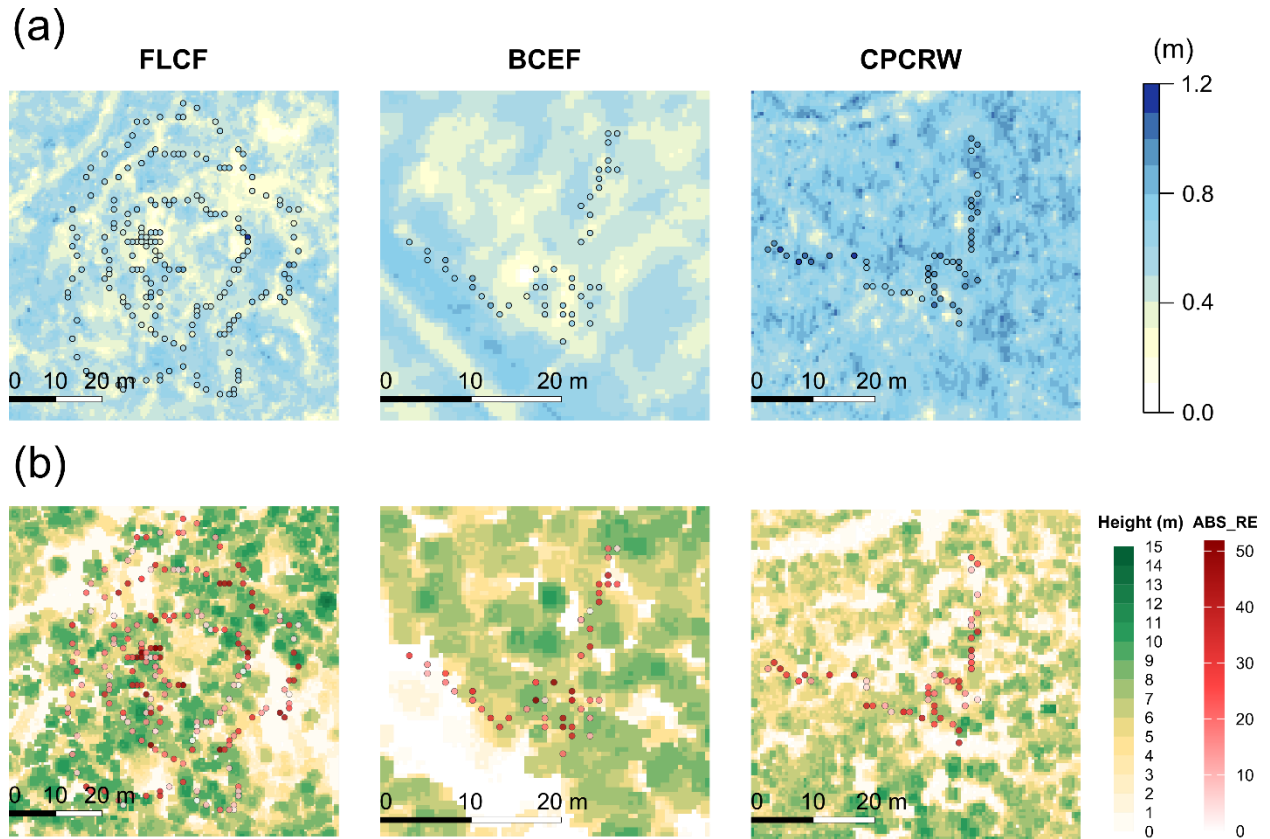


Figure 1. Relationship between canopy height and high LiDAR snow depth error (ABS_Error ≥ 0.5). (a) Predicted probability of high error as a function of continuous canopy height, from a logistic regression. (b) Odds ratios of high error by canopy height category, relative to the 0–5 m reference group. (c) Observed proportion of high-error cases by canopy height category. Error bars indicate 95% confidence intervals.



NEW 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 16:

Figure 7: I would also be interested in seeing this analysis repeated using absolute error rather than relative error. Some of the apparent increase in relative error under taller canopy may simply reflect shallower snowpacks.

Response:

We agree that the apparent increase in relative error under taller canopy may partly reflect shallower snowpacks beneath dense canopy. We have repeated the analysis using absolute error (LiDAR minus in situ, in meters) and find that the pattern differs from the relative error version: the proportion of high absolute error (≥ 0.3 m) peaks in the 5–10 m canopy class rather than the 10–15 m class seen with relative error. We also examined the relationship between in situ snow depth and canopy height directly and find a moderate negative correlation (Pearson $r = -0.52$), consistent with canopy interception reducing snow accumulation beneath taller canopy. Because relative error is calculated by dividing the absolute error by in situ snow depth, and because snow depth itself decreases with canopy height, relative error is inherently more sensitive to canopy height than absolute error is; the same absolute LiDAR error corresponds to a larger relative error where the underlying snowpack is shallower. We report both error formulations, together with the snow depth–canopy height relationship, as supplementary figures, and we have revised the Discussion to note this sensitivity explicitly and to state the canopy height effect on relative error more cautiously as a result.

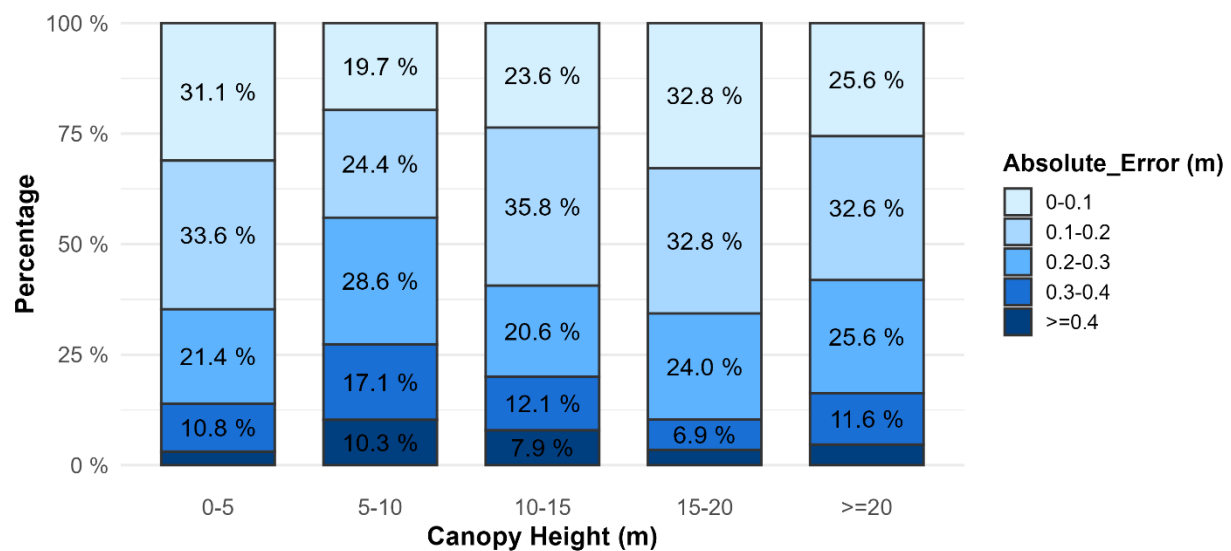


Figure R1. Distribution of absolute LiDAR error (m) across canopy height categories.

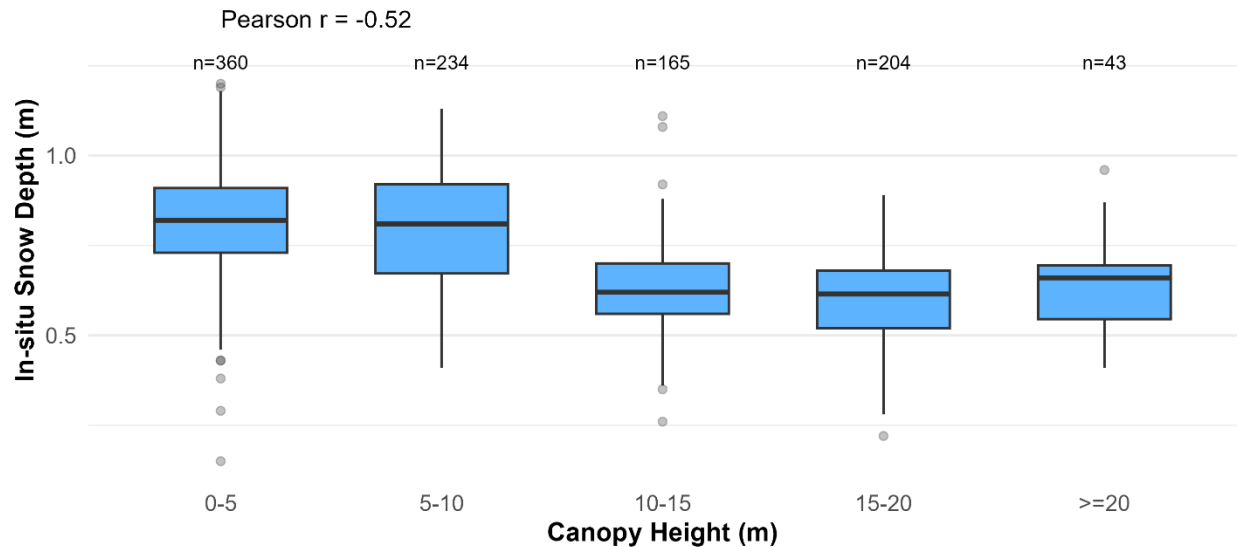


Figure R2. In-situ snow depth versus canopy height category (Pearson $r = -0.52$).

Revised manuscript (Page 29)

We found that canopy height was the most frequently top-ranked factor in predicting relative error, followed by slope, while other variables played minor roles; this ranking is sensitive to how error is defined, as canopy height is also correlated with shallower snow depth (Pearson $r = -0.52$), which inflates relative error independent of any LiDAR-specific effect.

Revised manuscript (Page 31)

We investigated the accuracy of airborne LiDAR for measuring snow depth in the boreal forests of interior Alaska and analyzed the environmental factors influencing any measurement errors. Using data from the NASA SnowEx 2023 campaign, we compared LiDAR-derived snow depths to extensive ground truth and found that LiDAR can capture snow depth patterns in forested terrain with moderate success: overall correlation ~ 0.60 and mean absolute error ~ 0.19 m. However, accuracy is not uniform – LiDAR performed best in open or lightly forested areas (often within $\sim 20\%$ of truth) and worst under dense canopies and on steep slopes (errors of $\sim 30\%$ or more, typically underestimating the snow). By applying machine learning models, *we identified canopy height as the most frequently dominant driver of LiDAR relative error, with taller vegetation associated with greater underestimation, though part of this relationship reflects the tendency for snow to be shallower beneath tall canopy rather than a purely LiDAR-specific effect.* Terrain slope was the second most important factor, also contributing to underestimation on steeper grades. Other factors like surface roughness and elevation

played secondary roles, and low ground point density was associated with larger errors but largely because it occurs in areas of tall canopy. Our results underscore that while airborne LiDAR is a powerful tool for snow depth mapping over broad areas, users must be aware of its limitations in forested landscapes. In particular, depths retrieved beneath mature boreal forest can significantly under-represent true snow accumulation. *The recurrence of a canopy effect across three distinct sites suggests that this is a general phenomenon in boreal forests and likely in other forest types as well, though its magnitude should be interpreted alongside the residual, unexplained error variance (R^2 of 0.18 to 0.53) and the metric-dependence discussed in Section 5.3.*

Comment 17:

364-367: I'm not convinced that the results demonstrate that errors increase with slope. Yes, there is a maximum error for FLCF, but not the two other sites.

Response:

We agree and have revised this passage to present the slope-error relationship separately for each site rather than comparing gentle slopes at one site to steep slopes at another. The revised text states that median error rises with slope at FLCF (from about 22 to 24% below 10° to about 33% at 15 to 25°), but that this trend is not consistently observed at BCEF or CPCRW, where error does not increase monotonically with slope. We interpret the slope effect as clearest at FLCF and note explicitly that it does not hold uniformly across all three sites.

Revised manuscript (Page 17-18, Lines 403 – 410) with key changes in bold italic

Terrain slope showed a site-dependent relationship with LiDAR snow-depth error (Figure 9a). *At FLCF, median absolute relative error increased from around 22 to 24% on slopes below 10° to about 33% on slopes of 15 to 25°, consistent with LiDAR measurements becoming more challenging on steeper terrain at this site.* This pattern was less consistent at the other two sites: at BCEF, median error increased modestly with slope up to the 10 to 15° class but did not continue to rise at steeper slopes, and at CPCRW, median error did not increase monotonically with slope, with comparable or lower values observed at the steepest slope classes. *We therefore interpret the positive relationship between slope and error as clearest at FLCF and note that slope alone does not consistently predict elevated error across all three sites.*

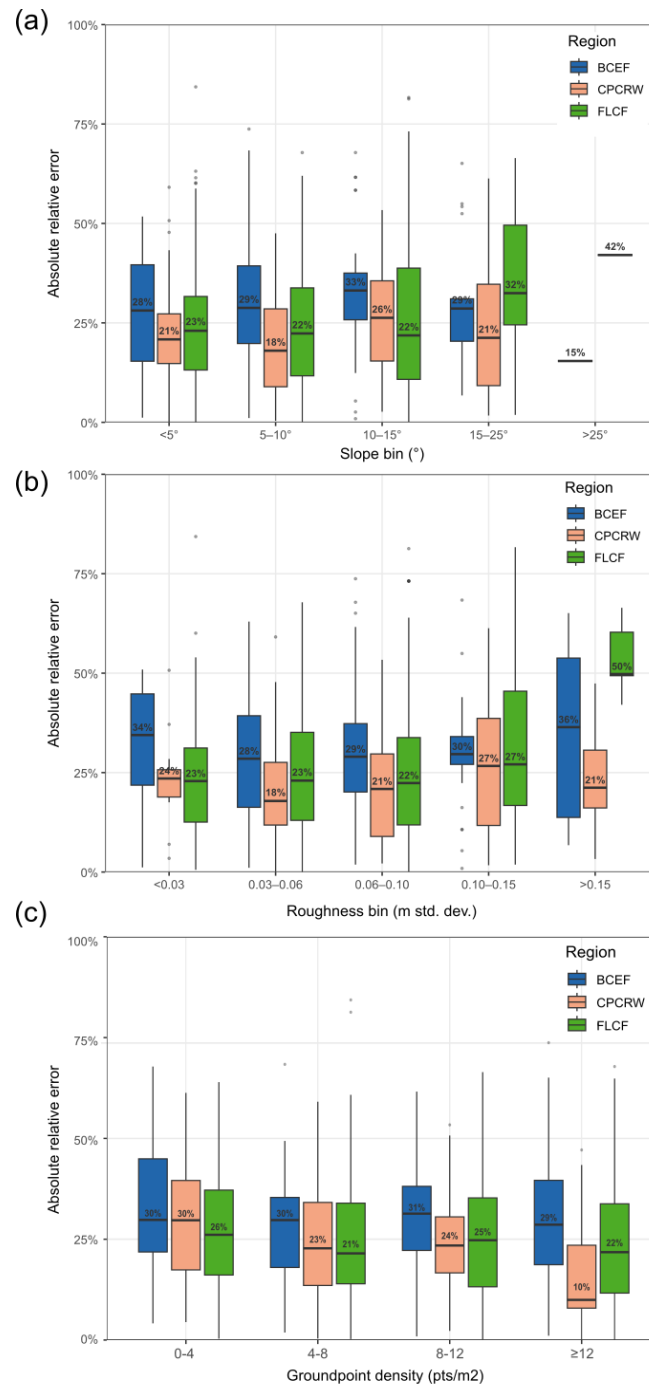
Comment 18:

Figure 9: label the x-axis for point density

Response:

We thank the reviewer for pointing this out. We have added an x axis label to the last panel.

The revised figure now appears as follows:



Comment 19:

426-427: ‘given the inherent variability and some unmodeled factors... this is a reasonable performance.’ This sentence is vague. Also, stating that it’s a reasonable performance is subjective.

Response:

We thank the reviewer for raising this point. We have restructured this paragraph to first report the site level R squared values (CPCRW about 0.53, BCEF about 0.33, FLCF about 0.18) and then discuss their interpretation, without characterizing the result as subjectively reasonable. We now state plainly that even at CPCRW, where the model performed best, nearly half of the error variance remains unexplained by the five predictors, and that this unexplained share is larger still at BCEF and FLCF.

Revised manuscript (Page 21, Lines 476 – 484) 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 ($R^2 = 0.527$, explaining about 53% of error variance), followed by BCEF ($R^2 \approx 0.33$) and FLCF ($R^2 \approx 0.18$). Even at CPCRW, where the model performed best, nearly half of the variance in LiDAR snow depth error remains unexplained by the five predictors considered, and this unexplained share is substantially larger at BCEF and FLCF. ***Given the inherent variability, unmodeled factors, and possible measurement noise in LiDAR snow depth errors, this outcome is not unexpected, but it indicates that canopy height, slope, roughness, elevation, and ground point density together capture only part of the drivers of LiDAR error.***

Comment 20:

449: 'the modeling results reenforce that canopy height is the dominant control on lidar snow depth error'. I would specify that canopy is the dominant control out of the tested variables. The R^2 values for the model demonstrate that much of the error is not accounted for by the variables.

Response:

We agree and have revised this statement to specify that canopy height is the dominant predictor among the variables tested in this study, rather than the dominant control on LiDAR error in general. We have also revised the following sentence, which previously stated that the models could "reasonably estimate error," to instead explicitly report the site level R squared range (0.18 to 0.53) and state that a substantial share of the error variance remains unaccounted for by the available variables, even at the best performing site.

Revised manuscript (Page 23, Lines 511 – 518) with key changes in bold italic

In summary, the modeling results indicate that canopy height is the dominant predictor among the five variables tested for LiDAR snow depth error across these boreal forest sites, with slope as a secondary contributor. However, the site level R^2 values (0.18 to 0.53) indicate that a substantial share of the error variance is not accounted for by these variables, even at CPRW where the model performed best. Within this limitation, the predicted error patterns were useful for identifying where LiDAR measurements are more likely to be unreliable, and this approach of error prediction could potentially be applied to flag or filter LiDAR-derived snow products by accounting for these variables, while recognizing that other unmeasured factors also contribute to LiDAR error.

Comment 21:

465-466: Providing the MAD of 19 cm for all points does not demonstrate that performance decreases under forest canopies. This statement would require two different errors for open and forested errors showing that.

Response:

We agree and have examined this using an open-versus-forest comparison in our response, though we note that the NASA SnowEx23 dataset provides canopy height only as a continuous variable rather than a discrete land-cover classification, so open and forested conditions cannot be independently verified against an authoritative land-cover product. Given this limitation, together with the modest and unevenly distributed sample across the canopy height range, we defined open locations as those with canopy height below 2 m ($n = 198$) and forested locations as those at or above this threshold ($n = 814$). Under this classification, open locations show a MAD of 15 cm and a bias of about -14 cm, while forested locations show a somewhat larger MAD of 17 cm and a bias of about -18 cm, supporting the point that performance is somewhat reduced under forest canopy rather than relying on the single pooled MAD value for all points. We consider the 2 m threshold a pragmatic choice given the available data rather than an established ecological definition of open versus forested cover.