Dear Reviewer 1,

The authors greatly appreciate the time you took to review our paper. Your comments and suggestions will greatly improve its overall quality. Please find our responses to each comment below. Reponses are in red. Manuscript changes will be in red italics.

Lines 66–67: "…beyond direct measurements."
It would be helpful to provide the reader with a general overview of the most widely used approaches, including both empirical and machine learning-based methods.

Thank you for this comment for clarity, we add a couple sentences to the paragraph to sort out the differences between the regression and ML based approaches.

To bridge these observational gaps, researchers have increasingly turned to empirical and machine learning approaches to extend snow depth estimates beyond direct measurements. Empirical approaches include terrain and climate based regressions and simple index methods that relate snow depth to predictors such as elevation, temperature, and precipitation, often calibrated from local observations. Terrain and climate-based regressions capture broad accumulation and melt patterns but often fail to represent nonlinear snowpack processes (Pflug et al. 2021, Pflug and Lundquist, 2020). However, machine learning (ML) studies have applied several methods including tree-based models like Random Forest (RF) and gradient-boosted trees, using combinations of in situ, SAR, optical, and lidar data to estimate snow depth across a range of scales (Dunmire et al. 2024).

Chapter 2:
The contrasting characteristics of the two basins are not entirely clear. For readers unfamiliar with these areas, additional information on their topographical and meteorological differences (or similarities) would be beneficial.

We appreciate your comment. To help clarify and contrast the differences between our two study sites we will add an additional paragraph at the end of the section that will directly contrast the two.

Together, these basins span a strong west-east contrast in snow and terrain regimes. Mores Creek is a larger, high-relief, wind-affected western basin with deep, persistent snow and pronounced redistribution. Whereas Hubbard Brook is a smaller, lower-relief, transitional northeastern forested basin with shallower, more spatially uniform snowpacks and milder temperatures.

Lines 131–132:
How were the seasonal phases defined? Were they based on mean or median values?

Thank you for this comment. For both Mores Creek and Hubbard Brook, the “peak SWE date” used to define seasonal phases was based on the mean day of maximum seasonal SWE across the available record at each site. We have clarified this in Section 2.3 by explicitly stating that we use the mean peak SWE date.

Chapter 3.3.3:
What was the rationale for not applying downscaling and bias correction to ERA5-Land data? Do you expect the proposed strategy to perform well at higher spatial resolutions?

We appreciate this comment. We chose not to apply a bias correction because a primary objective of this study is to evaluate whether our model can be run independent of local in situ data. If we ran a bias correction, in situ data would be necessary.

In our framework, spatial detail is provided by the lidar snow depth maps and the ancillary data produced with it (terrain), where as ERA5-Land supplies only a single basin scale temporal snow depth evolution. We anticipate the approach to remain applicable at higher spatial resolutions as long as the lidar and terrain inputs resolve the relevant spatial variability. However, in extremely heterogeneous regions, hyperlocal information could further improve performance.

Chapter 4: “…dynamical daily meteorological data.”
Please clarify what specific data are being referred to here. While they may have been mentioned earlier, the connection is not clearly explained.

Thank you for this comments. We have clarified what we mean by “dynamic daily meteorological data” by explicitly stating that the RF uses daily snow depth, air temperature, and wind time series from either in situ stations or ERA5-Land as the temporal driver.

Lines 302–303:
How were the synthetic snow depth maps generated?

Thank you for suggesting clarification. The synthetic maps are spatially uniform grids where we set the snow depth to zero across the domain. For Mores Creek, we used the high elevation Freeman Station to confidently observe when there is no snow, which turned out to be 1 October and 30 June. So for each year, we created 2 synthetic no snow (all depths = 0) maps for these dates. Without these ‘synthetic’ zero snow depth maps, the RF lacks training data, representing no snow conditions, which can lead to overpredictions of snow at the beginning and end of the snow season. Using the synthetic no snow maps which correspond to the met data for those specific dates (1 October, 30 June), we explicitly train the model to recognize the met conditions associated with a bare-ground state.

We will add the following text:

'…synthetic snow depth maps generated as spatially uniform grids where snow depth was set to zero across the entire domain. These zero-value maps were assigned to specific dates at the beginning (before the first snowfall) and at the end (after snowmelt) of each winter season and paired with the actual meteorological forcing data for those days.'

Lines 322–323:
Is the use of a single site and one winter season sufficient to robustly evaluate the algorithm?

Thank you for this important comment. A single site for one winter alone is not sufficient. At Mores Creek, the single Freeman timeseries is used only as an independent check on the in situ driven model. Model robustness is assessed mostly using multiyear and date lidar-based cross-validation and spatial error analysis in both basins (5.1, 5.2). We do use several in situ stations including SNOTEL at Mores Creek, which has multiple years to assess the ERA5-Land driven model. Conversely, there are several stations in Hubbard Brook where we can assess multiple winters across sites to test the in-situ driven model (see points in Figure 1). We will update this sentence for added clarity:

For the RF model in Mores Creek, only one site (Freeman) and the 2024–2025 winter were used for this time-series evaluation, because SNOTEL observations are used as the temporal driver and cannot simultaneously serve as an independent validation series. Freeman thus provides an independent, high-elevation check on the in situ–driven RF model, while SNOTEL is used over multiple years as an independent reference when evaluating the ERA5-driven RF model (Section 5.3, Table 1). At Hubbard Brook, time-series evaluations were conducted at each of the nine in situ sites across 2023–2025, providing a multi-site, multi-year assessment of temporal performance (Table 1).

Was the sample size sufficiently diverse and representative?

Yes, we use equal allocation stratified sampling across 16 elevation-aspect combinations, with 50k samples at Mores Creek and 10k at Hubbard, ensuring that all elevation and terrain aspect conditions are represented rather than relying on a small purely random subset.

For additional clarity, we will add the following:

At Mores Creek, the stratified sample size represented roughly <0.25% of all valid pixels and spans all combinations of elevation and aspect classes, while at Hubbard Brook, the stratified samples similarly cover the full elevation-aspect range.

Lines 391 and 405: "…with low bias…"
Please quantify the bias.

Thank you for this good observation. We excluded 'with low bias' as we did not include bias for Hubbard Brook.

Lines 465–466:
It would be helpful to explicitly define "Tair fwd5" and "snowdepth lag5."

Thank you for this comment. We will modify this line to the following:

Whereas high 5-day forward-looking air temperatures ('Tair fwd5') generally decreased in modeled snow depth, and deeper 5-day backward-looking snow depths ('snowdepth lag5') slightly increased modeled snow depths, at Hubbard Brook.

Discussion:
In the context of climate change scenarios, do you consider the proposed strategy to be robust and reliable? Would modifications be necessary under changing climatic conditions, and if so, how?

Thank you for this point. Our method is interpolative, relying on the distribution of conditions observed in the training data. Hence, our strategy should remain consistent as long as future snow and met conditions are generally similar to those sampled by the lidar acquisitions and reflected in the driver time series. If climate change/scenarios lead to substantial changes to the snow regime, like large snow events from frequent atmospheric rivers or rain on snow events becoming normal, in that case we would need to consider capturing a few of those events with lidar acquisitions and potentially reevaluate model skill.

We speak to this under 6.6:

Operational use benefits from a few safeguards. Because the method is intentionally interpolative, predictions should be limited to the observed parameter space and accompanied by routine checks for distribution shift and representativeness of the dynamic snow depth forcing. When a shift caused by, new storm types, unusual warmth, or evolving canopy, additional flights can be scheduled during the phases shown to add the most value, using the cadence results to target timing.

References:
Some claims would benefit from more up-to-date and appropriate references:

- Lines 48–49: "…over larger domains"

Added: Gascoin, S., Luojus, K., Nagler, T., Lievens, H., Masiokas, M., Jonas, T., Zheng, Z., and De Rosnay, P.: Remote sensing of mountain snow from space: status and recommendations, Front. Earth Sci., 12, 1381323, https://doi.org/10.3389/feart.2024.1381323, 2024.

• Lines 50–51: "…of sparse ground networks"

Added:
Nolin, A. W.: Recent advances in remote sensing of seasonal snow, J. Glaciol., 56, 1141–1150, https://doi.org/10.3189/002214311796406077, 2010

Qiao, D., Chen, X., Zhou, J., Liang, S., and Liu, G.: Improving the accuracy of gridded snow depth estimation through multi-source data and a machine learning fusion model, Sci Rep, 15, 40917, https://doi.org/10.1038/s41598-025-22347-x, 2025

• Line 68: Please cite the Random Forest algorithm

Added:
Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

• Line 80: "…to sustain frequent airborne surveys"

This sentence will be removed in the revision.

• Line 308: Please cite the Scikit-learn library

Added: Pedregosa, F., Pedregosa, F., Varoquaux, G., Varoquaux, G., Org, N., Gramfort, A., Gramfort, A., Michel, V., Michel, V., Fr, L., Thirion, B., Thirion, B., Grisel, O., Grisel, O., Blondel, M., Prettenhofer, P., Prettenhofer, P., Weiss, R., Dubourg, V., Dubourg, V., Vanderplas, J., Passos, A., Tp, A., and Cournapeau, D.: Scikit-learn: Machine Learning in Python, MACHINE LEARNING IN PYTHON, 12, 2825–2830, 2011.

• Line 332: "…one-way analysis of variance (ANOVA)"
Added: Stahle, L. and Wold, S.: Analysis of variance (ANOVA), Chemometrics and Intelligent Laboratory Systems, 6, 259–272, https://doi.org/10.1016/0169-7439(89)80095-4, 1989.

• Line 692: More recent references should be considered

Added:

Geissler, J., Rathmann, L., and Weiler, M.: Combining Daily Sensor Observations and Spatial LiDAR Data for Mapping Snow Water Equivalent in a Sub-Alpine Forest, Water Resources Research, 59, e2023WR034460, https://doi.org/10.1029/2023WR034460, 2023.

Herbert, J., Raleigh, M. S., and Small, E. E.: Using a random forest model to combine airborne lidar and Snotel data for daily estimates of snow depth across mountain drainage basins of Colorado, https://doi.org/10.22541/essoar.173655460.06498107/v1.