Reply to Reviewer 1

We thank the Reviewer for insightful comments and suggestions. Our item-by-item replies are in blue.

1. The transfer of ground parameters from open areas to forested pixels (Section 3.2.2, lines 230-238) currently depends on low spatial variability among open-area estimates alone, as stated in the text - but this does not establish that the same values hold under canopy. A sensitivity test perturbing the interpolated ground parameters to see how much the final SWE retrieval changes would help address this, given that in-situ validation data under canopy may not be available.

We thank the reviewer for this suggestion. We conducted a sensitivity analysis in which the forward model used in Step IV (MEMLS + WCM) is run one parameter at a time across the range of values retrieved over a forest pixel with deep snow (new Appendix F, Fig. F.1). Soil moisture in the context of the retrieval algorithm is an effective variable used to infer the dielectric properties at the soil-snowpack interface (Rai and Barros, 2026). Unconstrained soil moisture changes X-VV by up to 3.9 dB. For ground roughness changes, sensitivity for both bands is less than 0.4 dB, and Ku-VV is controlled mainly by snow microstructure (e.g. snow correlation length). If 100% of the maximum error from ground parameters in Step II of the algorithm propagated to SWE retrievals, the local uncertainty for the specific forested deep snow pixel used for sensitivity analysis would be 8%. For comprehensive sensitivity analysis of the ground parameter retrieval step using snow pit data, please see Rai and Barros (2026). Nevertheless, the errors from Step II are upper bounds, because effective soil moisture is re-optimized in the final Bayesian step (Step IV) using the kriged value as the prior. A full perturbation and re-inversion following the three-step Bayesian sequence (Step II, III and IV) is beyond the scope of this study. A short summary has been added at the end of Sect. 3.2.2, and a note is added to Conclusions.

2 The WCM parameter-fixing scheme (Section 3.2.3, line 254; discussed further via retrieved A/B in Appendix D) is presented without addressing how sensitive retrieval skill is to which parameter is held constant at each step.

The sensitivity analysis (Appendix F, Fig. F.1) shows how strongly the backscatter responds to each WCM parameter over its retrieved range. The parameter A changes the backscatter by 0 dB for the a median density forest pixel and by at most 0.3 dB under dense canopy ($Mv \sim 12$ kg m^{-2}), thus the observations cannot constrain it. (see also Appendix D). B and Mv change the backscatter by only 0.1-0.5 dB for the median density forest pixel and up to 1.1-1.4 dB under dense canopy, so they are retrieved without problem. The prior range of B is bounded by canopy closure ($\tau^2 \geq 1 - Cc$, Eqs. 5-6). Fixing A therefore has little effect on retrieval skill, whereas fixing B would matter for the densest forest pixels. Ground roughness, which is held constant in Step IV, behaves the same way: it changes both bands by less than 0.4 dB within its retrieved range. We have added

this explanation to Sect. 3.2.3 and revised the argument given there for holding roughness constant for consistency with the sensitivity analysis.

3 The 30% RRB threshold used to define a "successful" retrieval (Section 3.3, line 295; Conclusion, line 409) is described as chosen "as a balance between physical quality and data retention" rather than from an independent physical or statistical justification — it would help to know why 30% specifically, rather than, say, 20% or 40%. Pixels that fail this threshold (along with the 30°-50° incidence angle criterion) are excluded from the reported skill statistics (e.g., the ARD percentages, Figures 6/8/9) — please report what fraction of all forested pixels are excluded; to help readers judge how representative the reported statistics are of overall retrieval performance.

We thank the reviewer for this suggestion.

First the incidence angle criterion is only necessary for airborne SAR due to the viewing geometry issues. In practice, this criterion eliminates the pixels along the edges of the flights. This issue would not arise for the case of satellite-based measurements. Table E.1 summarizes the numbers of pixels for each flight that satisfy the viewing angle criterion.

The RRB criterion is not met for 3-8% of the pixels depending on the flight path. Overall the total number of retrieved pixels is 5582, and 5135 meet the RRB 30% criteria. Note that the 30% RRB threshold is an upper bound. It was selected to balance the agreement between simulated and observed backscatter with the retention of observations in heterogeneous forested environments. Figure E1 shows that most pixels exhibit residuals substantially below this threshold, while larger residuals primarily occur where landcover heterogeneity is not well described at the spatial resolution of the retrievals, such as forest transitions and mixed land-cover regions.

We agree that the choice of threshold and its influence on retrieval coverage require further clarification. A new Table E.2 is added to show retrieval acceptance rates for different RRB thresholds, including 20%, 30%, and 40% for each flight at 90m resolution. Table E.1 reports the fractions of pixels excluded by the residual and incidence-angle criteria separately. The difference in pixel numbers between 20% and 40% thresholds varies between 4% and 8% of all retrieved forested pixels depending on the flight. Overall the difference is small, and it would be even smaller if retrievals were conducted at higher spatial resolution with better resolved land-cover heterogeneity.

4. The snow pit validation (Table 7, lines 315-323) pools pits with only one collocated retrieved pixel within 100 m at equal weight with pits that have many more (e.g., N=14) - a weighted or stratified error metric, or a check on how the reported 0.033 m RMSE changes if the single-pixel pits are excluded, would clarify how much those pits are driving the result.

We thank the reviewer for this suggestion. We added a weighted ARE column to Table 7. That table gives the error metrics for all pits and for pits with more than one collocated pixel, with each pit weighed either equally or by its number of collocated pixels (N). Weighting by N lowers the RMSE from 0.056 m to 0.048 m and the MARE from 8.9 % to 7.1 %. Excluding the three single-pixel pits gives an RMSE of 0.049 m and a MARE of 7.4 %. The single-pixel pits therefore do not

drive the result. Pit 32S has the largest error (ARE 24 %): it lies at the forest edge, 71 m from its only collocated pixel. Its weighted ARE is only 8.5 %. While making this check, we found that the RMSE reported in the manuscript (0.033 m) is not correct if all values in Table 7 are included in the estimation. The correct value is 0.056 m, which is consistent with the stated 8 % of maximum pit SWE (0.69 m), and it has been corrected throughout. The number of collocated pixels per pit ranges from 1 to 6. The value 14 in Table 7 is the MARE of Pit 26W and Pit 71W, not a pixel count. In the revised Fig. 6b, marker size is proportional to N.

5. The fourth flight's poor performance (Table 8; lines 348-354, 369-377) is attributed to retrieval error, but the authors' own citation of LiDAR underestimation under dense canopy and steep terrain (Jacobs et al., 2021) is never used to separate how much of the disagreement is LiDAR bias instead.

The Reviewer's point is well taken. We agree that the poorer agreement in the fourth flight (20:23:38 UTC) assumes LiDAR estimates are perfect and attributes the error to the retrieval alone. LiDAR tends to underestimate snow depth under dense canopy and on steep slopes (Jacobs et al., 2021). The retrieval bias in this flight is positive (+0.18 m, Table 8), which is the direction that LiDAR underestimation would produce. The MSHM prior, which is independent of the radar data, also has its largest positive bias against LiDAR in this flight (+0.05 m). This flight's short track crosses the steep northern slope and mixed forest-wetland edges (Figs. 6 and 13), where such LiDAR errors are expected. We control for this effect in part by excluding LiDAR pixels with sub-pixel SD greater than 0.3 m or snow depth below 20 cm. However, without independent snow depth observations at these locations, the LiDAR and retrieval contributions cannot be separated quantitatively.

6. Canopy closure and the interception model were both developed at other sites and applied to Grand Mesa without site-specific validation. The manuscript states that uncertainty from these inputs has not been propagated into the final retrievals (line 428, line 484). Short of a full perturbation-and-rerun test, the authors could instead (a) discuss, even qualitatively, the expected direction and magnitude of error these relationships might introduce given known differences between Grand Mesa and the sites where they were originally developed, or (b) report any uncertainty bounds already published alongside the original Pomeroy et al. (2002) and Hedstrom and Pomeroy (1998) parameterizations, to give readers some sense of the error these inputs could be carrying forward

We thank the reviewer for this suggestion. We now describe in the Methodology (Sect. 3.2.1) how canopy closure introduces error, and we report the published uncertainty bounds in Appendices A and B. The parameter Cc affects the retrieval through three pathways: two concern the MSHM priors (canopy shading of radiation, snowfall interception), and one concerns the WCM as it establishes the bounds on canopy transmissivity in the retrieval ($\tau^2 \geq 1 - Cc$). An overestimated Cc lowers the prior of SWE under the canopy through increased interception. It also allows stronger canopy attenuation and therefore a larger posterior SWE, and the two effects partly offset each other. An underestimate has the opposite effect. The sign of the Cc error at Grand Mesa cannot be

independently determined. MODIS LAI accounts for foliage clumping, while the LAI-2000 values used to fit the empirical relationship do not, and this can increase the bias. Snow interception and storage on the canopy and vegetation heterogeneity within 500 m pixels can decrease the bias. Pomeroy et al. (2002) fitted the Cc-LAI relationship with $R^2 = 0.97$ for an effective winter LAI of 0.2-4.1. Because the relationship is logarithmic, a factor-of-two error in LAI changes Cc by about 0.2. The Hedstrom and Pomeroy (1998) interception model reproduced measured interception with standard errors of 0.27-1.00 mm SWE. At the seasonal scale, it predicted forest snow accumulation within 9.4 mm SWE (Pomeroy et al., 2002), which is about 2 % of the median forest SWE at Grand Mesa. The sensitivity analysis (Appendix F) shows that canopy parameters matter mainly for densely forested pixels.

Line 75–77: The reported RMSE (<7%) is presented as a favorable result — can the authorsclarify how this threshold was established, e.g., whether it reflects accuracy standardsaccepted elsewhere in the snow remote-sensing literature (e.g., reiterate NASEM DecadalSurvey requirements)?

We thank the reviewer for this comment. The RMSE of less than 7 % is the error reported by Singh et al. (2024) for open grasslands, not a threshold defined by us. We have revised the Introduction (l.76–79) to place it against accepted requirements: the 30 mm (SWE < 300 mm) and 10 % (SWE > 300 mm) accuracy proposed for satellite SWE retrieval (Rott et al., 2010), and the priority given to high-resolution SWE in mountain areas by the NASEM Decadal Survey (2018).

Line 109: Which HRRR forecast hour was used? Given that forcing data are subsequentlymodified within the workflow, it would help to note whether forcing-data limitations areaddressed later in the manuscript.

The HRRR forcing follows Cao and Barros (2020), who used the hourly forecast from disaggregated to 30 min Grand Mesa for 2016–2017; this is now stated at l.109. Errors in the forcing propagate into the MSHM priors, as now noted at l.118–119. They are discussed in Sect. 3.2.1 (l.233–239) for the canopy adjustments and in Sect. 4 (l.419–421) for the coarse-resolution HRRR precipitation. the prior uncertainties in Tables 3–6 allow the Bayesian update to correct part of these errors.

Line 126: A reference is needed here.

We thank the reviewer for pointing this out. The placeholder has been replaced with Singh et al. (2024), which describes the quality control applied to the same SnowEx'17 SnowSAR data (l.127).

Line 127–128: Please cite the prior analyses referenced here.

The prior analyses are now cited at l.128–129. Singh et al. (2024) retained only the co-polarized SnowSAR measurements for the same dataset, and Tsang et al. (2022) note that X- and Ku-band SWE retrieval algorithms have so far relied on the two co-polarizations.

Table 1 / Line 147: How many snow pits were evaluated, and on what basis were thesespecific selected? This information (number of sites) appears in Table 7 but is not statedin the main text.

The number of pits and the selection criterion are now stated in the main text (l.150–152). Of the 72 SnowEx'17 snow pits sampled on Grand Mesa between 20 and 24 February 2017, the 12 pits from 21–23 February that have at least one successful forested 90 m retrieval pixel within 100 m are used for evaluation (Table 7).

Line 167–168: Please specify how the snow hydrology model was modified to simulatesnowpack physics under canopy, or reference where this modification is described.

The canopy modifications of MSHM follow Cao and Barros (2023): canopy shading of shortwave and longwave radiation, snowfall interception, and unloading of intercepted snow. These are described in Sect. 3.2.1 and Appendices A–B, and the sentence at l.170–172 now names these processes and gives the references.

Figure 4: Please clarify what the two columns containing "Veg," "Grd," and "n" outside ofthe IV steps represent.

Added the explanation in the caption

References

Cao, Y., Barros, A.P., 2020. Weather-dependent nonlinear microwave behavior of seasonal high-elevation snowpacks. Remote Sens. 12, 3422. https://doi.org/10.3390/rs12203422

Cao, Y., Barros, A.P., 2023. Topographic controls on active microwave behavior of mountain snowpacks. Remote Sens. Environ. 284, 113373. https://doi.org/10.1016/j.rse.2022.113373

Hedstrom, N.R., Pomeroy, J.W., 1998. Measurements and modelling of snow interception in the boreal forest. Hydrol. Process. 12, 1611–1625.

Jacobs, J.M., Hunsaker, A.G., Sullivan, F.B., Palace, M., Burakowski, E.A., Herrick, C., Cho, E., 2021. Snow depth mapping with unpiloted aerial system lidar observations: a case study in Durham, New Hampshire, United States. The Cryosphere 15, 1485–1500.

National Academies of Sciences, Engineering, and Medicine (NASEM), 2018. Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. The National Academies Press, Washington, DC. https://doi.org/10.17226/24938

Pomeroy, J.W., Gray, D.M., Hedstrom, N.R., Janowicz, J.R., 2002. Prediction of seasonal snow accumulation in cold climate forests. Hydrol. Process. 16, 3543–3558.

Rai, A., Barros, A.P., 2026. Quantifying snow-ground backscatter uncertainty: a Bayesian approach using multifrequency SAR and in-situ observations. Remote Sens. 18, 634. https://doi.org/10.3390/rs18040634

Rott, H., Yueh, S.H., Cline, D.W., Duguay, C., Essery, R., Haas, C., Hélière, F., Kern, M., Macelloni, G., Malnes, E., Nagler, T., Pulliainen, J., Rebhan, H., Thompson, A., 2010. Cold Regions Hydrology High-Resolution Observatory for Snow and Cold Land Processes. Proc. IEEE 98, 752–765. https://doi.org/10.1109/JPROC.2009.2038947

Singh, S., Durand, M., Kim, E., Barros, A.P., 2024. Bayesian physical–statistical retrieval of snow water equivalent and snow depth from X- and Ku-band synthetic aperture radar – demonstration using airborne SnowSAR in SnowEx'17. The Cryosphere 18, 747–773. https://doi.org/10.5194/tc-18-747-2024

Tsang, L., Durand, M., Derksen, C., Barros, A.P., Kang, D.-H., Lievens, H., Marshall, H.-P., Zhu, J., Johnson, J., King, J., Lemmetyinen, J., Sandells, M., Rutter, N., Siqueira, P., Nolin, A., Osmanoglu, B., Vuyovich, C., Kim, E., Taylor, D., et al., 2022. Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing. The Cryosphere 16, 3531–3573. https://doi.org/10.5194/tc-16-3531-2022