

## **Reviewer 2:**

The reviewers comments are in black and the responses are in blue

We thank Dr. Palomaki for his comments that helped improve the manuscript.

### General comments

The authors present results from an experiment to assimilate modeled SWE (via SVS2/Crocus) and radar backscatter (via SMRT) into a timeseries of predictions of SWE and snow depth. The modeling framework was developed to support the Terrestrial Snow Mass Mission currently under development. Overall this work is relevant for the snow hydrology and remote sensing communities and is appropriate for The Cryosphere.

The manuscript would benefit from more precise language throughout. Specifically, I think it is important to be explicit that no snow or radar observations are actually used in this analysis; everything is a model output. The performance statistics are derived using 10 reference runs which are model output generated with perturbed meteorological data. Given this setup, I do not think it's appropriate to refer to the input data as an "observation". This is fairly inconsistent throughout the paper, e.g. "backscatter observations" (line 268) and "SWE observations" (line 269), but sometimes there are qualifiers like "synthetic SWE observations" (line 276) or "synthetic true states providing reference snow states and backscatter values" (line 182). I suggest using phrases like modeled SWE, simulated backscatter, synthetic data, etc. and being consistent throughout the paper. I think the use of "reference" to describe a model run used as the basis for comparison is effective. Maybe it's possible to use that terminology to refer to variables as well.

Thank you for this comment. We have revised the manuscript to consistently use the term "synthetic observations" throughout, which clearly distinguishes these model-generated observations from real observational data while maintaining standard data assimilation terminology.

My other general comment is about the ensemble generation, specifically the assumption in lines 171-172 that the HRDPS model errors are consistent between the three sites. This assumption could use some additional justification because the three sites are described as having very different characteristics in Section 2.1. In addition to the Powassan site having warmer temperatures and much less snow than Rogers Pass, I assume the rolling terrain around Powassan is much better represented in the 2.5 km HRDPS model than the complex terrain around Rogers Pass. So I am not convinced that the HRDPS errors would necessarily be of similar magnitudes at these sites, especially for precipitation which you

state can be a significant error source for SWE estimates in line 28. I think the perturbation distributions could actually be much wider at Rogers Pass, which would in turn impact most/all of the subsequent results. If you have long-term timeseries of both snow depth and SWE at both sites, one way to check (not the only way!) would be to compare the ensemble spread shown in rows 3 and 4 of Figure 2 to many seasons of observations at the three sites. If a similar percentage of seasonal snow depth/SWE curves fall within the spread of the model at all sites, that could be evidence that the method is fine. If fewer seasons fall within the spread at some sites (e.g. Rogers Pass compared to Powassan), you may need to widen the perturbation distributions for those sites. I'm not sure the best way to perform this analysis quantitatively to look for significant differences between sites, but maybe you can somehow start with the distributions in Table 2.

We acknowledge that using the same perturbations across all three sites may not be optimal for each individual location. It is worth noting that because we apply multiplicative perturbations to the precipitation amounts, the spread in SWE at Rogers Pass is wider than at Powassan and TVC as Rogers Pass experiences larger snowfalls. However, since these are synthetic data assimilation experiments, our primary focus was on evaluating the assimilation method itself, rather than focusing on the model uncertainty specific to the three locations. Larue et al. (2018) used a similar approach when assessing the impact of assimilating synthetic observations of brightness temperature at 14 sites across the province of Quebec (Canada). The meteorological forcing for Larue's simulations was taken from the Canadian Regional Deterministic Prediction System (RDPS) at 10-km resolution, and the same perturbations were applied to all the 14 sites. In the context of a future data assimilation system using the TSMR radar backscatters, the snow ensemble prior to data assimilation will account for uncertainties in the meteorological forcing and uncertainties in the snowpack scheme itself (Lafaysse et al., 2017). Uncertainty in the meteorological forcing will be generated using an ensemble precipitation analysis where uncertainties vary in space and time.

We invested considerable efforts in ensuring that our open-loop ensembles exhibit appropriate spread characteristics, given our assumptions relative to the use of the same perturbations across all three sites. To verify ensemble quality, we calculated the spread-skill ratio (the ratio of ensemble spread to RMSE, Fortin et al., 2014; Dirkson and Buehner, 2025) using the reference runs as truth. The ensemble size was accounted for in the calculation of the RMSE to eliminate the effect of the ensemble size on the spread-skill scores (Dirkson and Buehner, 2025a). We also calculated the climatological variance condition (Johnson and Bowler, 2009; Dirkson and Buehner, 2025b), i.e. the ratio of the mean of the spread of each member by the variance of the truth for each reference run. These two scores, when close to 1, indicate that an ensemble is reliable. Table R1 presents

spread-skill values over the three years and across the 10 reference runs for each site. The results show that for all variables, the average spread-skill and climatological variable condition values are close to 1, indicating that the ensembles are reliable and that the ensemble spread accurately captures the prediction errors. In some cases, our ensembles are slightly under-dispersive (ratio < 1), such as for SWE at Rogers Pass or slightly over-dispersive (ratio > 1) such as for SWE at Powassan.

Table R1: Spread skills and climatological variance conditions (mean of the spread of each member divided by the variance of the truth) over the 10 reference runs and the three winter seasons

|                       | Powassan      |                                   | TVC           |                                   | Rogers Pass   |                                   |
|-----------------------|---------------|-----------------------------------|---------------|-----------------------------------|---------------|-----------------------------------|
|                       | Spread skills | climatological variance condition | Spread skills | climatological variance condition | Spread skills | climatological variance condition |
| Backscatter at 13 GHz | 1.11          | 1.21                              | 0.98          | 1.08                              | 1.02          | 0.87                              |
| Backscatter at 17 GHz | 1.11          | 1.22                              | 0.97          | 1.10                              | 1.00          | 0.89                              |
| SWE                   | 1.22          | 1.10                              | 1.11          | 1.17                              | 0.84          | 0.90                              |
| Snow depth            | 1.16          | 1.07                              | 1.11          | 1.16                              | 0.87          | 0.92                              |

A paragraph and Table R1 were added to Section 2.3.2 of the revised manuscript to present the results of spread-skill that show that our ensembles are reliable. This new paragraph reads:

“The spread-skill (ensemble spread divided by RMSE) of the ensembles and the climatological variance condition (mean of the members spreads by the spread of the truth) were calculated to assess ensemble reliability (Fortin et al., 2014; Dirkson and Buehner, 2025a,b; Johnson and Bowler, 2009). The ensemble size was accounted for in the calculation of the RMSE to eliminate the effect of the ensemble size on the spread-skill scores (Dirkson and Buehner, 2025a). The reference runs (the truths) were used in the calculation of these two scores. Table 4 presents the spread skills over the three winter seasons and the reference runs for backscatter (13 GHz and 17 GHz), snow depth, and SWE. Most values are close to 1, indicating that the ensembles are reliable and that the ensemble spread accurately captures the predicted errors in ensemble mean.”

Additionally, we compared the open-loop ensembles against available snow pit observations at the different sites (Figure R1). In most cases, the field observations fall within or near the ensemble ranges, providing further confidence in our ensemble

generation approach. The measurements at TVC present large spreads due to the variability of the snowpack at that site.

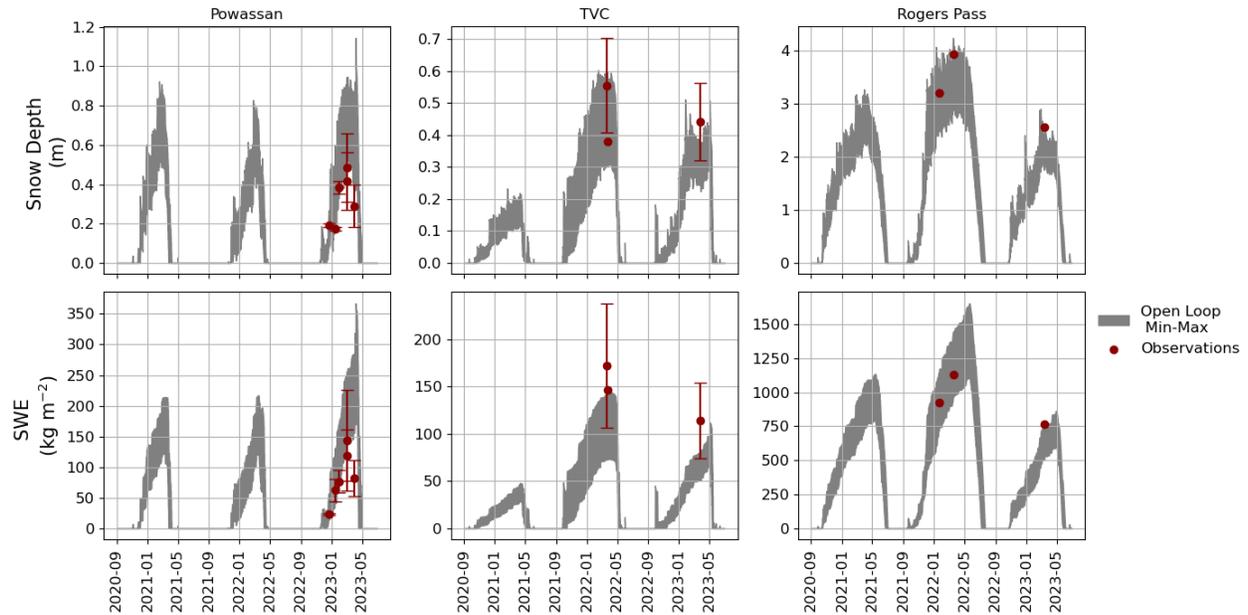


Figure R1: Evolution of the open loop spread (min and max) of snow depth (SD) (top row) and SWE (bottom row) at Powassan (left column), TVC (middle column), and Rogers Pass (right columns). Mean in-situ observations from snowpits are shown by the red dots and the spreads (if several snowpits were done at the same location) are shown by the vertical bars.

While we recognize that site-specific perturbations would be ideal, we are confident that our uniform perturbation approach does not compromise the validity of our synthetic experiment results, given the demonstrated ensemble quality metrics and agreement with available observations.

#### Reference:

Fortin, V., Abaza, M., Anctil, F., and Turcotte, R., 2014: Why Should Ensemble Spread Match the RMSE of the Ensemble Mean?, *Journal of Hydrometeorology*, 15, 1708–1713, <https://doi.org/10.1175/JHM-D-14-0008.1>.

Dirkson, A., and M. Buehner, 2025a: The Effect of Ensemble Size on the Mean Squared Error and Spread–Error Relationship. *Mon. Wea. Rev.*, 153, 1219–1229, <https://doi.org/10.1175/MWR-D-24-0189.1>.

Johnson, C., and N. Bowler, 2009: On the Reliability and Calibration of Ensemble Forecasts. *Mon. Wea. Rev.*, **137**, 1717–1720, <https://doi.org/10.1175/2009MWR2715.1>.

Dirkson, A. and Buehner, M. , 2025b: Are we misdiagnosing ensemble forecast reliability? On the insufficiency of Spread-Error and rank-based reliability metrics, <https://doi.org/10.48550/arXiv.2512.02160>, preprint.

### **Specific comments**

Title: The phrase “Assimilation of synthetic radar backscatters” in the title is a bit unclear, given that TSMR is a synthetic aperture radar mission but here “synthetic radar” is used to refer to the output of the SMRT model. I suggest replacing “synthetic” in the title to simulated or modeled (or another synonym of your choosing).

We have modified the title to “Assimilation of synthetic observations of radar backscatters at Ku-band improves SWE estimates”. We have decided to keep the term “synthetic” in the title since it is widely used in the context of data assimilation studies. We hope the new formulation reduces the potential confusion with synthetic aperture radar.

Line 69: “if retrieval algorithms are able to retrieve a wider range of SWE values” – does this imply the need for backscatter measurements to be obtained over a wide range of snow conditions, or an appropriate algorithm formulation that can produce a large range of SWE values from the backscatter data? Put another way, is this in reference to the backscatter data or the algorithm?

We would like to clarify our interpretation of Cho et al. (2022). While they do highlight the need for additional work to demonstrate the sensitivity of X- and Ku-band backscatter to deep SWE, their primary emphasis is on the need to improve retrieval algorithms for deep SWE conditions. Given this focus, we believe our current citation appropriately reflects the main contribution of their work and have therefore retained the original sentence.

Line 145: Please be more specific with the forecasts and lead times here. For example, with the 7-12 hour lead time, are HRDPS forecasts generated only every 12 hours but with hourly lead times that allow you to fill in the hourly timeseries? Or are there hourly forecasts that allow you to build your timeseries always with 1 hour lead times?

This is made more clear in the revised manuscript. This reads:

“Successive short-term HRDPS forecasts were combined to generate continuous hourly meteorological forcing. HRDPS forecasts initialized every 6 hours (00Z, 06Z, 12Z, 18Z) were used, with forecast lead times of 7-12 hours extracted from each run and concatenated to create a continuous hourly time series.”

Line 150-156: I think I understand that this perturbation method is applied to multiple meteorological variables. I suggest making this more explicit in the text here, either using another subscript in Equation 1 (though this may end up looking too messy) or a sentence before Equation 1. If there are multiple variables considered, do they all have the same decorrelation time length ( $\tau$ )? How is  $\tau$  calculated?

This is rephrased in the revised manuscript. It now reads:

“Separate perturbations are applied to different meteorological forcing. The time evolution of each perturbation follows a first-order auto-regressive model describing the time evolution of an error as in Magnusson et al. (2017)”.

In addition, some sentences were moved before Eq. 1 following the next comments from the reviewer to improve the clarity of this section.

To estimate  $\tau$ ,  $\alpha$  was first calculated and Eq. 2 was used. A sentence was moved after Eq. 2 to explain how  $\alpha$  is calculated. It reads:

“ $\alpha$  was determined by calculating the autocorrelation of the error between the HRDPS and the observations with a lag of 1.”

Line 157: How did you select additive vs. multiplicative for a given meteorological variable?

The choice of multiplicative perturbations for wind speed and shortwave radiation and additive perturbations for the other variables was motivated by Charrois et al. (2016) and Larue et al. (2018). These two studies present similarities with our manuscript since they focused on the assimilation of synthetic observations in a snowpack model using a particle filter. The text was modified to cite these two studies:

“Following Larue et al. (2018b) and Charrois et al. (2016), the additive perturbations were drawn from a normal distribution and were applied to the air temperature and incoming longwave radiation forcing while the multiplicative perturbations were drawn from a log-normal distribution and were used for the precipitation, wind speed, and shortwave radiation forcing.”

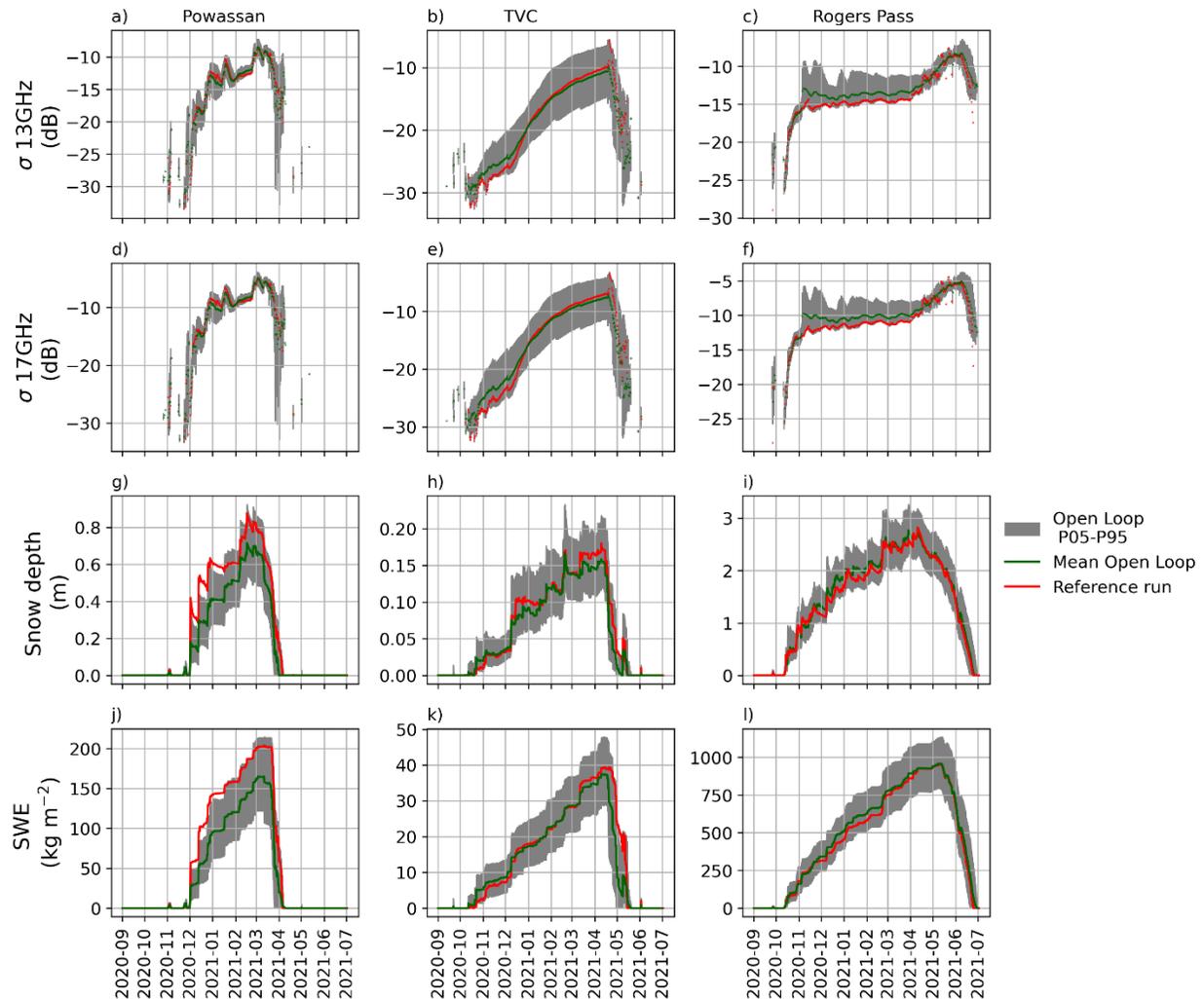
Line 169-170: Okay, this info helps answer my previous question. I suggest moving these lines (and perhaps Table 2) closer to Equation 1 in the text. Especially the detail about calculating  $\alpha$  as the autocorrelation of HRDPS residuals – that part is not clear from the initial presentation of the equations.

We agree. We moved the sentence concerning the calculation of  $\alpha$  just below equation 2. A sentence before equation 1 was modified and now reads:

“Separate perturbations are applied to different meteorological forcing in order to represent errors in the HRDPS forcing data compared to local observations.”

Figure 2: I suggest making the linewidths of the red reference runs a little narrower so it's easier to compare them to the gray shading of the ensemble spread.

We have modified Figure 2 to only include one reference run for clarity. The new Figure 2 is as follows:



**Figure 2.** Spread of the open loop (OL) ensemble composed of a 100 members (between 5th and 95th percentiles) and one randomly chosen reference run at Powassan (a, d, g, j), TVC (b, e, h, k), and Rogers Pass (c, f, i, l) for the backscatter at 13.5 GHz (a, b, c), for the backscatter at 17.25 GHz (d, e, f), for snow height (g, h, i), and for SWE (j, k, l) for the winter 2020-2021. Note the different y-axis for each sub-figure. Similar figures but with all 10 reference runs and for the three winter seasons can be found in the supplementary material (Fig. S1, S2, and S3).

The figures with all the reference runs are in the supplementary material and we modified the size of the red lines (reference runs) to better see the open loop ensemble spread.

Line 275: “despite this site having the highest number of weekly observations throughout the winter” – I thought all sites had the same weekly data (12 UTC on Mondays pulled from the hourly data to run the model). Do the assimilation timesteps vary between sites? I also notice a dashed line for “Obs. Times” in the Figure 3 legend but I don’t see any of those lines in the panels.

Observations were initially selected at weekly intervals; however, they were only assimilated if they met specific criteria: (i) presence of a dry snowpack and (ii) backscatter values within plausible ranges. Additionally, because the snow season at Rogers Pass is longer than at the other sites, more weeks with snow cover were available for generating observations. We have revised the manuscript to clarify this selection process. The relevant section now reads:

“The assimilation of backscatter at Rogers Pass barely improved the SWE estimate compared to the open loop, despite this site having the highest number of weekly observations throughout the winter that passed the selection criteria specified in Section 2.3.2 and due to longer snow seasons at the Alpine site.”

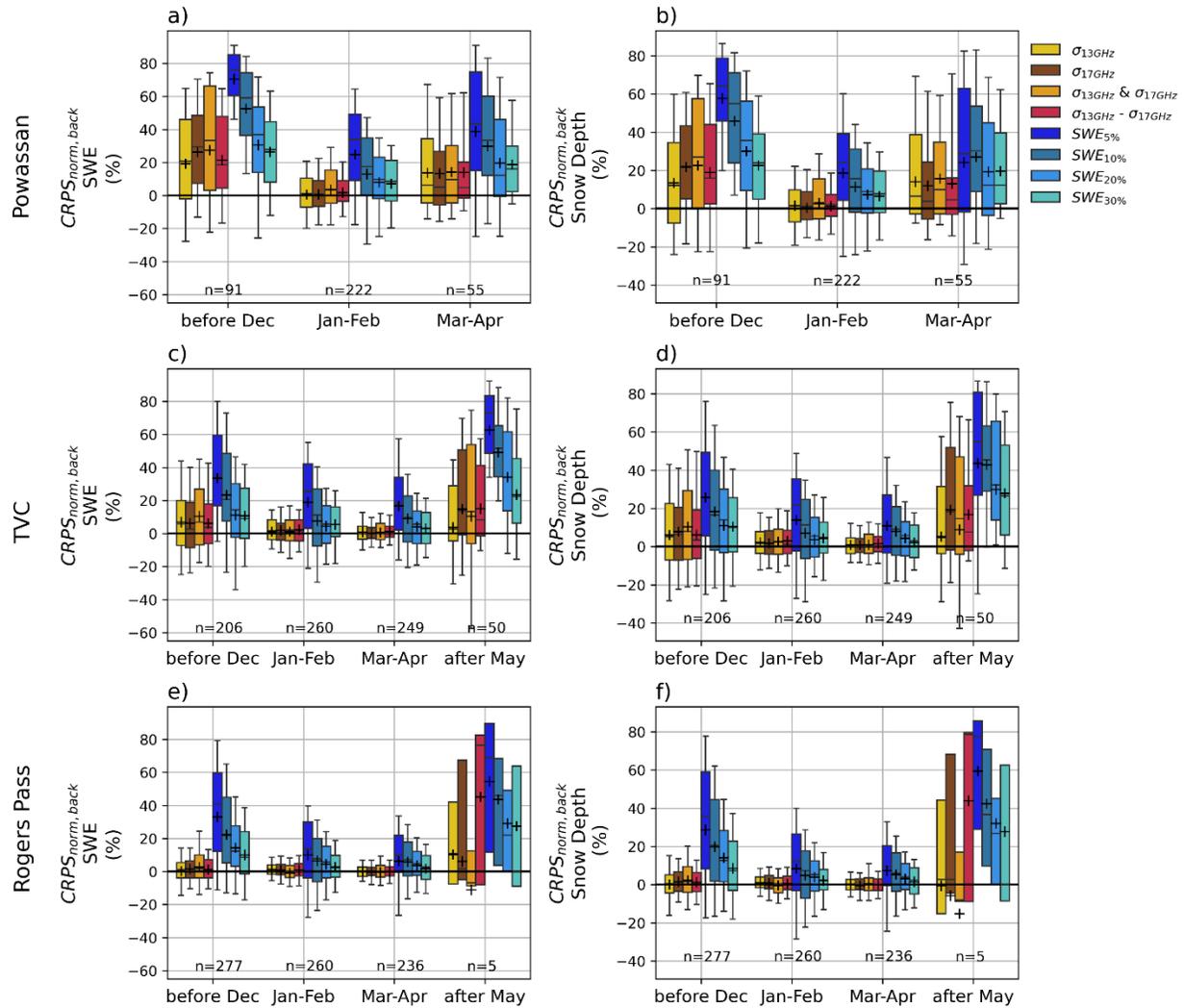
“Obs. Times” in Figure 3 was removed.

Line 312-314: Is it possible that the improvement seen during the melt season is influenced by the sample size? I imagine the modeled LWC reaches 1% some days/weeks before the SWE actually starts decreasing. In that case, how many backscatter simulations actually get assimilated during the melt season at these sites, especially in comparison to the earlier periods? I see in Figure 5 that the latest period boxplots at all sites have higher mean/median CRPS than the earlier periods, but they also have much wider spread. So are they statistically significantly higher in the latest periods? It would be helpful to run quantitative significance tests on these results, and also include sample sizes in the x tick labels, e.g. “before Dec (n=XXX)”.

We have revised Figure 5 to include the number of observations assimilated during each period. This reveals that considerably fewer observations were available during the melt period. Consequently, results from the melt period should be interpreted with caution, as they lack the statistical robustness of the accumulation and peak periods due to limited sample size. This information is now clearly displayed in the updated Figure 5 (shown below in the reply to the next comment).

Figure 5: It would be helpful to make this figure larger to see more detail. Consider a 2x3 subpanel layout (snow depth and SWE columns with sites as different rows) instead of 3x2.

Figure 5 in the revised manuscript was modified as follows:



**Figure 5.** Normalized CRPS against the background particles for all the different runs and the three winter seasons for SWE prediction (a,c,e) and snow depth prediction (b,d,f) for Powassan (a,b), TVC (c,d), and Rogers Pass (e,f) based on the month of the observations. Boxplots show median (center line), interquartile range (box), 10th–90th percentiles (whiskers), and mean (+). No outliers are shown for clarity.

Figure 5: It is probably expected that assimilating SWE gives better results for both SWE and snow depth when the assimilated SWE errors are smaller. But it is interesting that the spread of the SWE assimilations seems to decrease with increasing error (i.e. taller boxplots for SWE with 5% error than 30% error). This looks like the opposite trend compared to Figure 4. Can you provide more discussion around these results?

This apparent paradox can be explained by the relationship between observation uncertainty and ensemble behavior.

When assimilating SWE observations with 5 % uncertainty (and corresponding error covariance), the particle filter strongly constrains the ensemble spread around the observation. However, if the observation itself deviates from the reference run (which serves as truth for evaluation), the tightly constrained ensemble may fail to capture the true state. This results in highly variable performance across different reference runs and winter seasons, leading to the wider spread in normalized CRPS values seen in the boxplots.

Conversely, when assimilating with 30% SWE uncertainty, the particle filter applies weaker constraints, resulting in less reduction of ensemble spread compared to the open-loop or background particles. While this produces lower average performance improvements, the wider ensemble spreads are more consistently able to capture the reference runs across different scenarios. This consistency translates to the smaller spread in normalized CRPS values observed for higher observation uncertainties.

To conclude, lower observation uncertainty can lead to either very good or very poor performance depending on observation quality, while higher observation uncertainty produces more modest but more consistent improvements. We have added a couple of sentences before Figure 5 in the revised manuscript. It reads:

“It can be noted in Fig. 5 that the spread of the SWE assimilations sometimes decreases with increasing SWE uncertainty. This highlights a trade-off in assimilation experiments: low uncertainties and therefore high observational constraints maximize potential improvements but increase sensitivity to observation errors, whereas higher uncertainties in the observations sacrifice peak performance for greater reliability across diverse conditions.”

Figure 7: It would be helpful here as well to have an indication of the sample sizes going into these boxplots, as well as some statistical interpretation of the results. Is it one vertical profile every hour for all 10 reference runs, or only profiles from December 27 1200 UTC every year? Given that the interquartile ranges of almost all boxplots in Figure 7 span both positive and negative values, are the results significantly different from a CRPS of 0? Put another way, does assimilating backscatter or SWE actually improve the modeled density or SSA?

Thank you for this comment. We went back to the data used to generate Figure 5 and realised that there was an error in Figure 7. We intended to show the seasonal CRPS values for density and SSA profiles, i.e. the CRPS values calculated at each assimilation time averaged over each winter season as shown in Figure 4 for the snow bulk properties. This is corrected in the revised manuscript and the text was revised accordingly. In addition, following the comment on the number of samples, a new table (Table 3) was added in the

revised manuscript to summarize the number of synthetic observations that were assimilated during winter season.

The new figure is as follows:

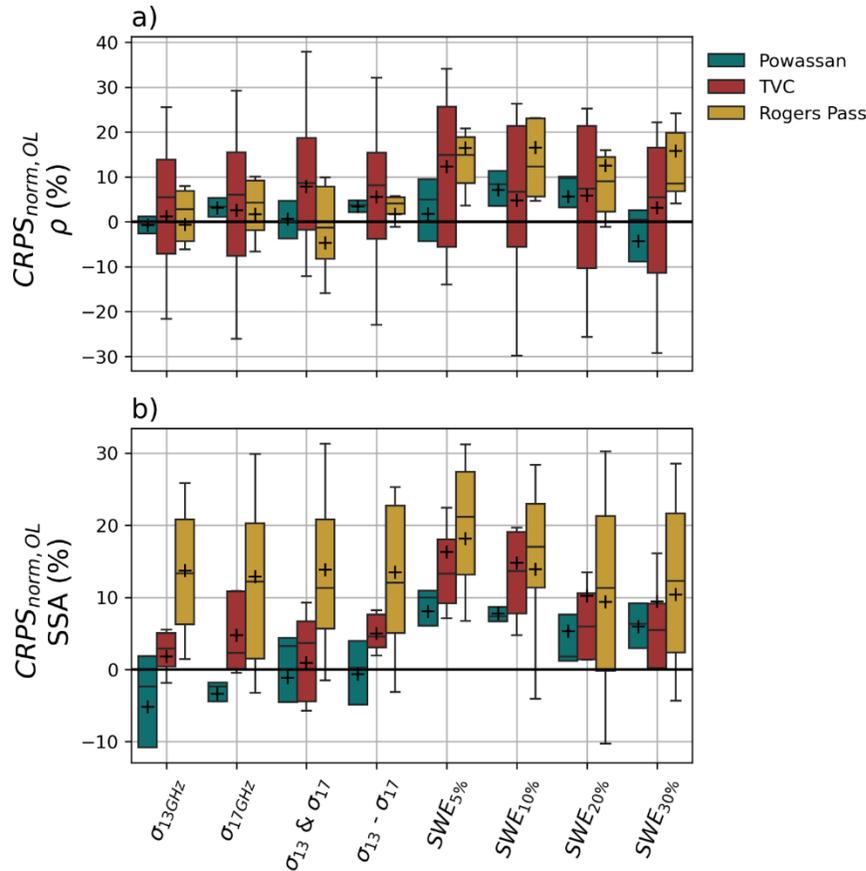


Figure 7. Normalized season CRPS (relative to open loop) for (a) vertical density profiles and (b) vertical SSA profiles, aggregated across the three sites (Powassan, TVC, and Rogers Pass) over the three winter seasons for the 10 reference runs (sample size of 30 for each boxplot). Box plots show median (center line), interquartile range (box), 10th–90th percentiles (whiskers), and mean (+). No outliers are shown for clarity.

The text was modified in Section 3.2: and now reads:

“Figure 7 shows the seasonal normalized CRPS values (relative to the open loop) for density and SSA profiles. For density profiles, improvements relative to the open loop were greatest at TVC compared to the other two sites, with the best estimates obtained when assimilating both frequencies of backscatter synthetic observations (median seasonal

normalized CRPS up to 9%). Some improvements in density were observed at Rogers Pass when assimilating backscatter at 17 GHz and the backscatter difference (median values of 5%), while at Powassan, only modest improvements were noted when assimilating the backscatter difference (median values of 3.5%). When assimilating SWE observations, density profiles were generally improved at all three sites, with the largest improvements occurring at Rogers Pass. The range of seasonal normalized CRPS values for density profile estimates at TVC exhibited large variability, indicating inconsistent assimilation performance at this site. For SSA profiles, the distributions of normalized CRPS showed improvements with positive median values at TVC (up to 17%) and at Rogers Pass, particularly when assimilating backscatter at 17 GHz and the backscatter difference (median value of 5%). No improvements over the open loop in SSA profiles were observed at Powassan. When assimilating SWE observations, SSA profiles were generally improved at all three sites, with the largest improvements overall occurring at Rogers Pass.”

The discussion concerning the new results on estimated snow profile properties was updated. It now reads:

“Backscatter assimilation demonstrated site-dependent effectiveness in improving vertical snow property estimates. For density profiles, TVC showed the most promising results, with the backscatter difference outperforming single-frequency assimilation. Assimilating the difference of backscatter frequency improved the density profiles estimates at both Powassan and Rogers Pass. For SSA profiles, Rogers Pass exhibited the strongest performance, with backscatter assimilation achieving improvements comparable to direct SWE assimilation. The improvements in density and SSA profiles at TVC are particularly encouraging, as current multi-layered snowpack models struggle with representing Arctic snowpack stratigraphy (e.g., Woolley et al., 2025; Vionnet et al., 2025). The limited improvements at Powassan, even when assimilating SWE observations, suggest challenges in constraining SSA at this site. Overall, the improvement of estimated vertical snow properties after assimilating either backscatter or SWE observations is consistent with findings from Shrestha and Barros (2025a).”

Line 344-350: This paragraph seems better suited for the conclusion of the paper.

This paragraph was removed in the revised manuscript as a similar paragraph is already written in the conclusions.

Line 361-362: “Consequently, the higher number of assimilated observations at TVC may have contributed to the improved SWE and snow depth estimates.” – this is contrary to your previous observation that Rogers Pass had the worst performance with the most assimilated data points (line 275, also 377).

In this sentence, we wanted to compare the number of observations between Powassan and TVC. This is made more clear in the revised manuscript, and it reads:

“Consequently, the higher number of assimilated observations at TVC compared to Powassan may have contributed to the improved SWE and snow depth estimates.” This is true that similar and more observations are Rogers Pass than at TVC but Rogers Pass suffered from saturation early in the winter seasons, greatly limiting the performance of the assimilation of backscatter.”

Line 393: Consider adding Bonnell et al. (2024) to the reference list here. They use an InSAR approach with much lower frequency but I think it’s an important point that dense forests are going to be tricky for any SAR-based approach for SWE retrievals. (Full disclosure – I am on this paper.)

Thank you, this citation is indeed relevant here and was added.

Line 398-401: I’m not convinced of these statements based on the results of Figure 7 (see above comment). Please revisit after some more rigorous analysis of the density and SSA results.

This paragraph was modified based on new results for Figure 7 that now shows season CRPS values and not individual CRPS values at all the assimilation times. It now reads:

“Backscatter assimilation demonstrated site-dependent effectiveness in improving vertical snow property estimates. For density profiles, TVC showed the most promising results, with the backscatter difference outperforming single-frequency assimilation. Assimilating the difference of backscatter frequency improved the density profiles estimates at both Powassan and Rogers Pass. For SSA profiles, Rogers Pass exhibited the strongest performance, with backscatter assimilation achieving improvements comparable to direct SWE assimilation. The improvements in density and SSA profiles at TVC are particularly encouraging, as current multi-layered snowpack models struggle with representing Arctic snowpack stratigraphy (e.g., Woolley et al., 2025; Vionnet et al., 2025). The limited improvements at Powassan, even when assimilating SWE observations, suggest challenges in constraining SSA at this site. Overall, the improvement of estimated vertical snow properties after assimilating either backscatter or SWE observations is consistent with findings from Shrestha and Barros (2025a).”

Line 402: Pointing to the different SVS2 setup is unsatisfying here because the whole modeling experiment is predicated on errors being from the meteorological forcing data, not the model (lines 141-142).

You are correct. This sentence was removed in the revised manuscript.

Section 4.2: This is an interesting point that could use more discussion. The assimilation resulted in the best improvement during the early season, which is exactly when we would expect ephemeral snow events that lead to the algorithm degeneracy. Figure S6 is quite noisy but it looks like the backscatter sample sizes show bigger drops in the early season compared to SWE sample sizes. Is this because snow melts and increases the soil moisture, which drops the modeled backscatter below the acceptable threshold? In this experiment it's possible to generate some new particles, but what about the approach with actual satellite data?

Thank you for this observation. During the early season, many backscatter observations fell below the thresholds specified in Section 2.3.2 (see Figures S1-S3), substantially reducing the number of observations available for assimilation.

The effective sample sizes are more reduced during the accumulation and melt periods for both SWE and backscatter assimilation due to two main factors: (i) precipitation events provide ensemble members with different precipitation amounts and phases, widening the ensemble spread and reducing the number of particles close to the observations to be resampled and (ii) melt events generate wet snowpacks, which reduces the number of synthetic observations to be assimilated and can also modify soil moisture content, which strongly affects backscatter values and adds additional constraints during backscatter assimilation. The latter factor explains why effective sample sizes for backscatter assimilation are sometimes lower than those for SWE assimilation. A sentence was added to the revised manuscript in Section 4.2:

“These periods of low Neff largely coincided with ephemeral snow events occurring before or after the main seasonal snowpack (Figs 2, S1, and S2). This is caused by two main factors: (i) precipitation events leading to ensemble members with different precipitation amounts and phases, widening the ensemble spread and reducing the number of particles close to the observations to be resampled, and (ii) melt events generate wet snowpacks, which reduces the number of synthetic observations to be assimilated and can modify soil moisture content, which strongly affects backscatter values and introduces additional variability in the ensemble, further reducing the number of particles close to the observations. The latter factor explains why effective sample sizes for backscatter assimilation are sometimes lower than those for SWE assimilation”

Regarding your question about particle generation with real satellite data, the systematic resampling method used in our particle filter framework can generate new particles regardless of whether observations are synthetic or real. This capability is inherent to the method itself.

## Technical

Line 3: capitalize Change

Modified.

Line 16: ~ appears over the 3 instead of before it. Maybe a  $\sim$  instead of a  $\tilde$ ? A few other occurrences of this throughout the manuscript.

Thank you, this is corrected in the revised manuscript.

Line 68: “RMSEs were” or “RMSE was”

Corrected.

Lines 94-97: Check for consistency between Section and Sect.

Corrected. All the “Sec.” were replaced by “Section”

Line 101: see -> sea

Corrected.

Line 105: Montpetit et al. missing a year

Corrected.

Line 116: remove “detailed”

Corrected.

Line 117: 1D -> 1-dimensional (for clarity)

Corrected.

Line 135: maybe “the version 2.0 of the MuSA” -> “MuSA version 2.0”

Corrected.

Line 142: change final comma to period.

Corrected.

Figure 2 axes and caption: snow height -> snow depth (for consistency with the rest of the paper)

Corrected.

Line 281: maybe “either backscatter” -> “backscatter at either frequency”

Corrected.

## References

Bonnell, R., Elder, K., McGrath, D., Marshall, H. P., Starr, B., Adebisi, N., Palomaki, R. T., and Hoppinen, Z.: L-band InSAR snow water equivalent retrieval uncertainty increases with forest cover fraction, *Geophysical Research Letters*, 51, e2024GL111708, <https://doi.org/10.1029/2024GL111708>, 2024.