Reviewer 1:

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

We thank reviewer 1 for their comments that helped improve the manuscript.

In this paper, the authors conducted an important research to explore the data assimilation potential of the TSMM mission of Canada, which provides dual-Ku backscatter measurements in weekly intervals.

However, firstly, the authors didn't clearly present whether all weekly backscatter coefficients were input together to constrain the entire snow season, or the snow process was updated at weekly steps.

If it is the first case, previous studies would not recommend perturbing the seasonal pattern of snowfalls. Instead, an adjustable constant multiplication factor will be applied to the entire snow season.

The current assumption cannot enumerate all possibilities in the meteorological forcing errors for the entire snow season. Therefore, instead, it adds great noise into the SWE and backscatter ensembles (e.g., Fig.2h), which have made DA really difficult.

Due to the reasons mentioned above, the presented ensembles alone in Fig.2 fail to convince me that dual-Ku backscatter can work well to constrain SWE uncertainty, although it indeed could.

Thank you for your comments. We would like to clarify how the assimilation was implemented in this study. Given its compatibility with non-linear models such as multilayered snow models, the particle filter method was adopted in this study. This sequential data assimilation approach updates the snow state estimates at each observation time by weighing and resampling particles according to their likelihood given the observations. This is made more clear in the revised manuscript. It now reads: "Given its compatibility with non-linear models such as multilayered snow models, the particle filter method was adopted in this study. This sequential data assimilation approach updates the snow state estimates at each observation time by weighing and resampling particles according to their likelihood given the observations"

Because this is a sequential approach, we did not perturb the seasonal pattern of snowfall. As indicated in Table 2 in the original manuscript, we perturbed air temperature, wind speed, longwave and shortwave radiation, and precipitation over time to generate the ensemble. Following the approach of Charrois et al. (2016), relative humidity was the only meteorological input that remained unperturbed. This perturbation strategy for creating snow ensembles prior to snow data assimilation is well-established in the literature (e.g., Charrois et al., 2016; Cluzet et al., 2022; Larue et al., 2018; Revuelto et al., 2018; Magnusson et al., 2017).

To avoid confusion between the reference runs and the open loop ensembles, only one reference run is now shown in Figure 2. The figures that include all 10 reference runs are shown in the supplementary materials.

Here is the new Figure 2 in the revised manuscript:

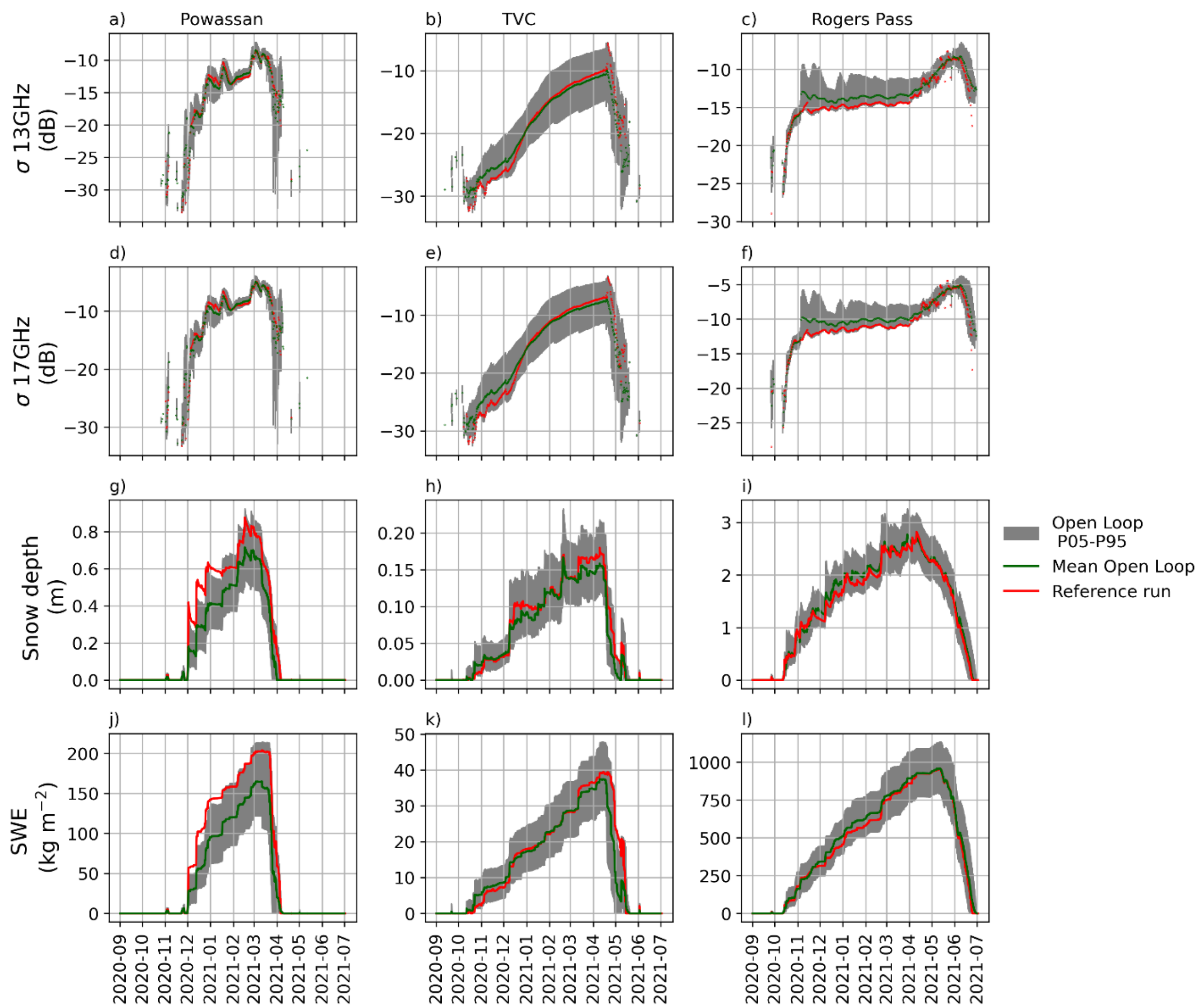


Figure 2. Spread of the open loop (OL) ensemble composed of a 100 members (between 5th and 95th percentiles) and one randomly chosen reference run at Powassan (a, d, g, j), TVC (b, e, h, k), and Rogers Pass (c, f, i, l) for the backscatter at 13.5 GHz (a, b, c), for the backscatter at 17.25 GHz (d, e, 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 subfigure. 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).

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, 2025a) 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) sur 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 member 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."

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.

The simulations also have other problems: usually, if the snowpack is deeper, the snow will tend to melt more slowly under the same energy input (air temperature+radiation). It is unrealistic that the spread of snow-off dates is so narrow in Rogers Pass, which is even narrower than the very shallow snow at TVC in Fig.2(k). It should be noted that the largest and the smallest peak SWEs for Rogers Pass differ by over 300 mm, whereas that for TVC is only 20 mm. Therefore, I guess there is also some problem in the snow process modeling. If the snow-off dates are correctly simulated, even fractional snow cover from optical sensors can be used for DA; backscatter should be more powerful.

Thank you for your comment. We would like to provide context regarding the snow model performance and the spread in snow-off dates. The snow model within SVS2, Crocus, has been extensively validated since the late 1980s (Brun et al., 1989; Vionnet et al., 2012; Lafaysse et al., 2017; Vionnet et al., 2025; Lafaysse et al., 2026), and we are confident in its representation of bulk snow processes, including snowmelt.

You correctly note that in spring 2021, the spread of snow-off dates is narrower at TVC (22.5 days) than at Rogers Pass (15.9 days). This difference can be attributed to a large snowfall event that affected some members of the open loop in May 2021 at TVC. Not all the members were as affected by this event as the amount and phase of precipitation varied between members, which substantially increased the spread of snow-off dates at that site. However, this pattern is not consistent across years. In spring 2022 (Fig. S1 of the original submission), the spread at TVC is 10.2 days compared to 21.7 days at Rogers Pass. Similarly, in spring 2023 (Fig. S2 of the original submission), the spreads are 7.2 days and 15.8 days at TVC and Rogers Pass, respectively.

It is also important to note that ensemble members experience different energy inputs due to the perturbed meteorological forcings. Consequently, a large spread in peak SWE values does not necessarily translate to a proportionally large spread in snow-off dates, as the melt dynamics are influenced by multiple meteorological factors that vary across ensemble members.

More points are listed as follows:

Major points:

1. Abs: In "Results indicate that assimilating backscatter observations reduced the mean continuous ranked probability score (CRPS) of SWE estimates by up to 32 % at the Arctic and humid continental climate sites compared to the open-loop ensemble, performing similarly to the assimilation of SWE with an observation error larger than 20 %", what is the source of the SWE directly assimilated?

We agree that this point needed clarification. The SWE observations that were assimilated were derived from the same reference runs used to extract and assimilate the backscatter data. We have revised the abstract to make this relationship more explicit. It now reads:

"Results indicate that assimilating synthetic observations of backscatter improved SWE estimates at the Arctic and humid continental sites, reducing the mean continuous ranked probability score (CRPS) by up to 32 % compared to the open-loop ensemble. This performance was comparable to assimilating synthetic SWE observations with observation errors larger than 20 %."

2. Line 150: I feel, the temporal evolution equation (1) for meteorological forcing error is needed only if the observations are inputted into DA sequentially.

Indeed, the observations are inputted into DA sequentially. Following a previous comment, we made it more clear in the revised manuscript that the assimilation method is sequential:

"Given its compatibility with non-linear models such as multilayered snow models, the particle filter method was adopted in this study. This sequential data assimilation approach updates the snow state estimates at each observation time by weighing and resampling particles according to their likelihood given the observations."

3. Follow point 2, and for lines 179-180: sequential assimilation had one problem in assimilating backscatter when new snowfall happens to occur at the observation time. It is related to the limited response of backscatter to small fresh snow particles, until the snow grains start to grow. How to avoid this problem?

This is indeed an issue. This is now stated in the discussion of the revised manuscript, and it reads:

"Also, backscatter assimilation results may have been affected when assimilation occurred during or shortly after snowfall, since backscatter has little sensitivity to fresh snow particles. Indeed, fresh snow consists of particles with small optical diameters (Mätzler, 2002), poorly scattering microwaves (Chang et al., 1976)."

References:

Chang TC, Gloersen P, Schmugge T, Wilheit TT, Zwally HJ. Microwave Emission From Snow and Glacier Ice. *Journal of Glaciology*. 1976;16(74):23-39. doi:10.3189/S0022143000031415

Mätzler C. Relation between grain-size and correlation length of snow. *Journal of Glaciology*. 2002;48(162):461-466. doi:10.3189/172756502781831287

4. Line 160: Although it may not be important, however, I feel additive perturbations may be more suitable for wind speed and shortwave radiation.

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."

5. Line 170: Do you mean the autocorrelation of the error of HRDPS compared to the observations?

You are correct, this is the autocorrelation of the error. The new sentence reads:

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

6. Line 192, Is Derksen et al.(2021) an indirect citation? Please clarify if this citation refers to more backscatter experiments in Canada or measurement experience.

This was an indirect citation to data published in other manuscripts. This is corrected in the revised manuscript. The measurements mentioned are radar measurements from a tower installed in Sodankylä, Finland for several years. The new sentence reads:

“which correspond to the lowest snow backscatter observations at these frequencies from a tower-based radar system deployed in Sodankylä, Finland (Lemmetyinen et al.,2018; Pan et al., 2024).”

References:

Lemmetyinen, J., Derksen, C., Rott, H., Macelloni, G., King, J., Schneebeli, M., Wiesmann, A., Leppänen, L., Kontu, A., and Pulliainen,J.: Retrieval of Effective Correlation Length and Snow Water Equivalent from Radar and Passive Microwave Measurements, Remote Sensing, 10, 170, https://doi.org/10.3390/rs10020170, 2018

Pan, J., Durand, M., Lemmetyinen, J., Liu, D., and Shi, J.: Snow Water Equivalent Retrieved from X- and Dual Ku-band Scatterometer Measurements at Sodankylä Using the Markov Chain Monte Carlo Method, The Cryosphere, 18, 1561–1578, https://doi.org/10.5194/tc-18-1561-2024, 2024

7. Fig.2, the authors could try to enlarge the backscatter time series for Rogers Pass and Powassan. The separation between different ensemble runs is too small.

The new Figure 2 in the revised manuscript has now different y axes for the different subplots to enlarge the backscatter time series. In addition, to avoid confusion, only one reference run is now shown in Figure 2. The figures will all the reference runs are shown in the supplementary materials. The new Figure 2 is presented in the response to the first comment of the reviewer.

Lines 240-262, This is a new term. The authors are suggested to explain the cumulative distribution function clearly. In addition, what does a higher CRPS stand for?

The text was modified in the revised manuscript to improve the clarity of what the CRPS is. This now reads:
“The Continuous Ranked Probability Score (CRPS) is a verification metric that measures

the difference between a predicted probability distribution and an observed value, with lower values indicating better forecast skill (Hersbach, 2000). The CRPS is calculated as:

$$\mathbf{CRPS}_t = \int (F_t(x) - O_t(x))^2 dx$$

where $CRPS_t$ is the score at time t, $F_{t(x)}$ is the cumulative distribution function (CDF, i.e. the probability that a variable is less than or equal to a given value) of the ensemble forecast at time t, and $O_{t(x)}$ is the CDF of the reference run (the truth) at the time of the observation (a step function that equals 0 for x < observed value and 1 for x ≥ observed value). The CRPS is the integrated squared difference between the predicted and observed CDFs across all possible values. The $CRPS_t$ values were then averaged over time for each assimilation experiment to obtain a mean CRPS, which represents the overall forecast accuracy."

8. Line 456, usually assimilating microwave observations directly should perform better. This is because the snow process model adds another group of constraints between snow depth and snow microstructure in addition to that from microwave observations, driven by the temperature gradient within the snow profile. Could you please check the correlation between snow depth (or SWE) and the snow microstructure parameter at some representative temporal points?

Here, we understand that the reviewer comment concerns L.455-456 of the original submission: “Assimilating the difference of the frequencies at the same time performed better at Rogers Pass with a mean normalized CRPS of SWE estimates over the open loop of 5 %.”

As this study represents the first attempt to assimilate backscatter observations for improving SWE estimates, we currently lack comparable studies to contextualize our results. Therefore, it is premature to conclude whether direct assimilation of microwave observations should inherently outperform the assimilation of microwave signal differences. We note that Larue et al. (2018a,b), as mentioned in L. 373 of the original manuscript, found improvements in SWE estimates when assimilating brightness temperature differences compared to assimilating single brightness temperature values.

Following your recommendation, we investigated potential correlations between snow microstructure and bulk SWE. To represent snow microstructure as a bulk variable, we calculated the microwave grain size of each layer (l_{mw}) (Picard et al., 2022) and averaged it over the snowpack using layer thicknesses as weights. The microwave grain size directly influences snow backscatter and is calculated as

$$l_{mw} = K d_{opt} \frac{2(\rho_{ice} - \rho_{snow})}{3\rho_{ice}}$$

With K refers to the snow polydispersity. It depends on the snow grain type (Picard et al., 2022) but is fixed at 1 in our study for all grain type as a preliminary step. Future applications will use the snow grain type derived from Crocus to estimate the polydispersity.

Figure R1 shows the temporal evolution of average snow microwave grain size and SWE for one winter season from a randomly selected reference run used to generate the synthetic observations. The results reveal a weak correlation between these two bulk variables at Rogers Pass, while some correlation appears to exist at the other two sites. At TVC, the snowpack experiences strong thermal gradients during the winter, resulting in a continuous increase in optical grain size (and associated microwave grain size) over the winter. At Rogers Pass, isothermal metamorphism dominates the evolution of snow microstructure, and frequent snowfall events made of snow with low optical diameter maintains the vertically-averaged microwave grain size close to a constant value in mid-winter. Finally, at Powassan, the snowpack experiences more melt-freeze cycles followed by snowfall events, resulting in successive increases and decreases in microwave grain size over the winter.

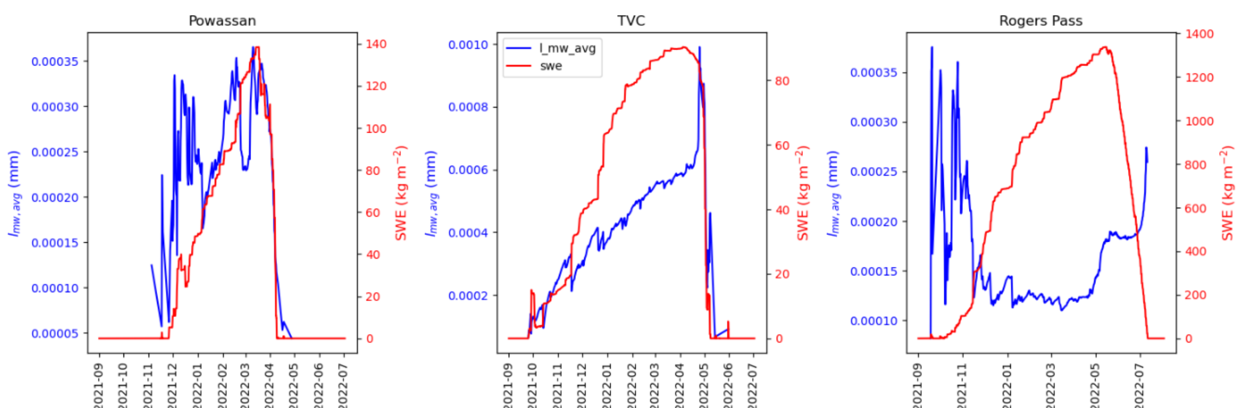


Figure R1: Temporal evolution of the average microwave grain size and bulk SWE for the 2021-2022 winter for one randomly chosen reference run, at the three study sites.

The boxplot below (Figure R2) presents correlation coefficients between vertically-averaged microwave grain size and bulk SWE across all winter seasons and all reference runs (10 reference runs per season). This analysis reveals distinct site-specific patterns: a strong correlation exists at TVC, moderate correlation at Powassan, and negligible correlation at Rogers Pass.

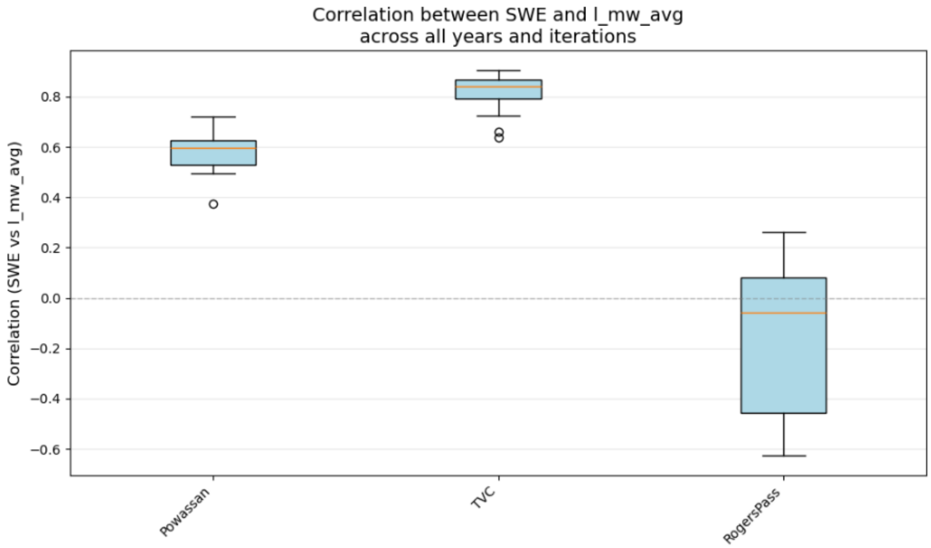


Figure R2: Boxplot of the correlation coefficients between the average microwave grain size and bulk SWE for all the reference runs over the three winter seasons at the three sites.

Despite this analysis shading new light on the correlation between SWE and the snow microstructure, it is not conclusive enough to explain how the backscatter assimilation performed in our study. Further work would be required and is outside of the intended purpose of our manuscript.

References:

Picard, G., Löwe, H., Domine, F., Arnaud, L., Larue, F., Favier, V., et al. (2022). The microwave snow grain size: A new concept to predict satellite observations over snow-covered regions. AGU Advances, 3, e2021AV000630. **https://doi.org/10.1029/2021AV000630**

9. Line 465: for the saturation of backscatter for SWE>300 kg/m3, while your statement in general is true, it was not supported by the comparison between (a) and (c) in Fig. 2. The spread of backscattering from Jan to March is larger at Rogers Pass than Powassan.

You are correct that the spread of backscatter is larger at Rogers Pass than at Powassan, which is partially attributable to snowpacks with different snow microstructures at Rogers Pass (c.f. response to comment #10). However, we would like to clarify that backscatter saturation refers to a different phenomenon than the spread of backscatter values. Saturation occurs when backscatter ceases to increase beyond a certain SWE threshold, even as SWE continues to accumulate. This saturation behavior is observed at Rogers

Pass, whereas at Powassan, backscatter continues to increase throughout the winter season in response to increasing SWE. We have revised the manuscript to clarify this distinction. The relevant section now reads:

"A major limitation found in this study is the saturation of synthetic observations of backscatter for SWE values above ~300 kg m^{-2}, resulting in quasi-constant backscatter values for SWE values above this threshold."

10. Also for Fig. 2(c)-(f), the simulated backscatter is questionable in mid-Jan and mid-Feb. The jumps of backscatter for these 3 to 4 ensembles are difficult to explain.

We conducted a detailed analysis to better understand the backscatter behaviour at Rogers Pass during January and February 2021 for several reference runs. For this analysis, we computed the microwave grain size for each snow layer in the simulated snowpack and averaged it over the entire profile using layer thicknesses as weights. The equation for the microwave grain size (Picard et al., 2012) is given in our answer to major point 9 of Reviewer 1.
Figure R3 shows the evolution of backscatter at 13 GHz, vertically averaged l_{mw}, and bulk SWE at Rogers Pass for the 2020-2021 winter for two reference runs. Reference run 1 (red) shows a moderate but continuous increase in simulated backscatter throughout the winter. On the other hand, reference run 2 (blue) shows a simulated backscatter that starts diverging from run 1 at the beginning of November 2020 and remains larger than the backscatter from run 1 over the course of the winter with several changes around an average value of -12 dB.

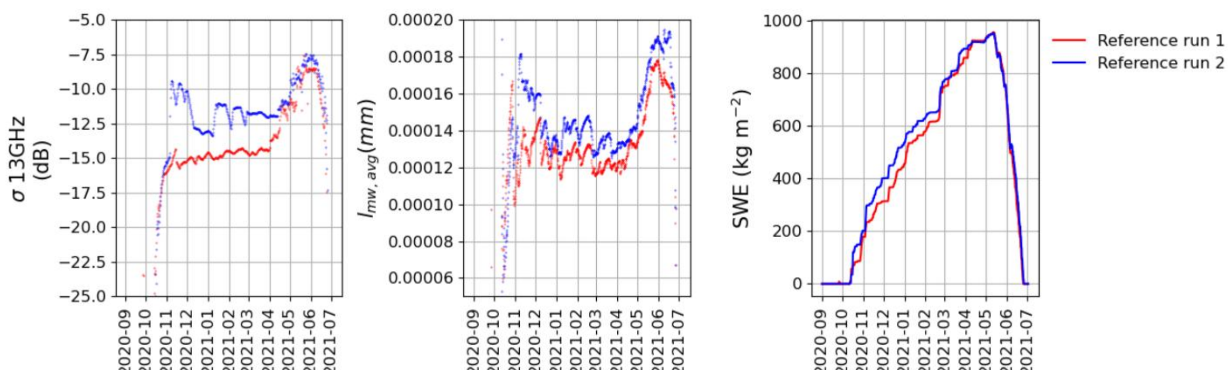


Figure R3: Temporal evolution of the backscatter at 13 GHz (left column), average microwave grain size (middle), and SWE (right) for the reference runs 1 and 2.

Figure R3 illustrates that the backscatter evolution for reference run 2 is driven by changes in vertically-averaged microwave grain size, which are less pronounced in reference run 1. Focusing on the January-February 2021 period (Figure R4), we found that the changes in microwave grain size for reference run 2 are primarily explained by changes in the microwave grain size of a snow layer located between 60 and 80 cm above the ground.

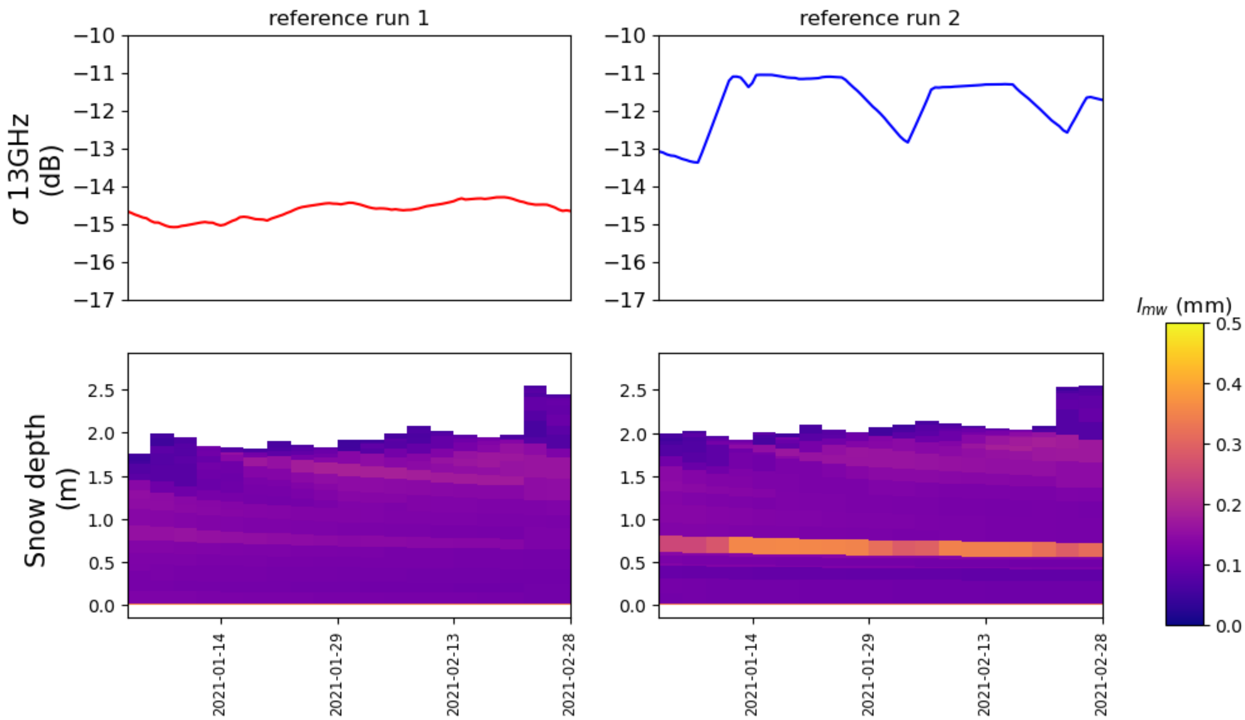


Figure R4: Temporal evolution of backscatter at 13 GHz (top row) and temporal evolution of the microwave grain size in the snow layers (bottom row) for the reference run 1 (left column) and reference run 2 (right column).

This layer of large microwave grain size in reference run 2 (Figure R4) originated from a precipitation event with mixed precipitation (rain and snow) in early November 2021 that led to the formation a layer of wet snow that then refroze (Figure R5). Due to wet snow metamorphism, the optical diameter of that layer quickly increased and remained larger than the surrounding snow layers over the course of the winter (Fig. R5). During the January-February period presenting changes in microwave grain size in reference run 2, we can observe that the snow grain type changed from melt-form (MF) to melt-form with faceted crystals grains (MF-FC) (Figure R5), associated with temperature-gradient metamorphism (TG). It led to a decrease in optical grain size (and associated microwave grain size). This behaviour is associated with limitations in the metamorphism scheme in Crocus to handle the evolution of snow optical grain size of layer made a frozen melt forms (Brun et al., 1992).

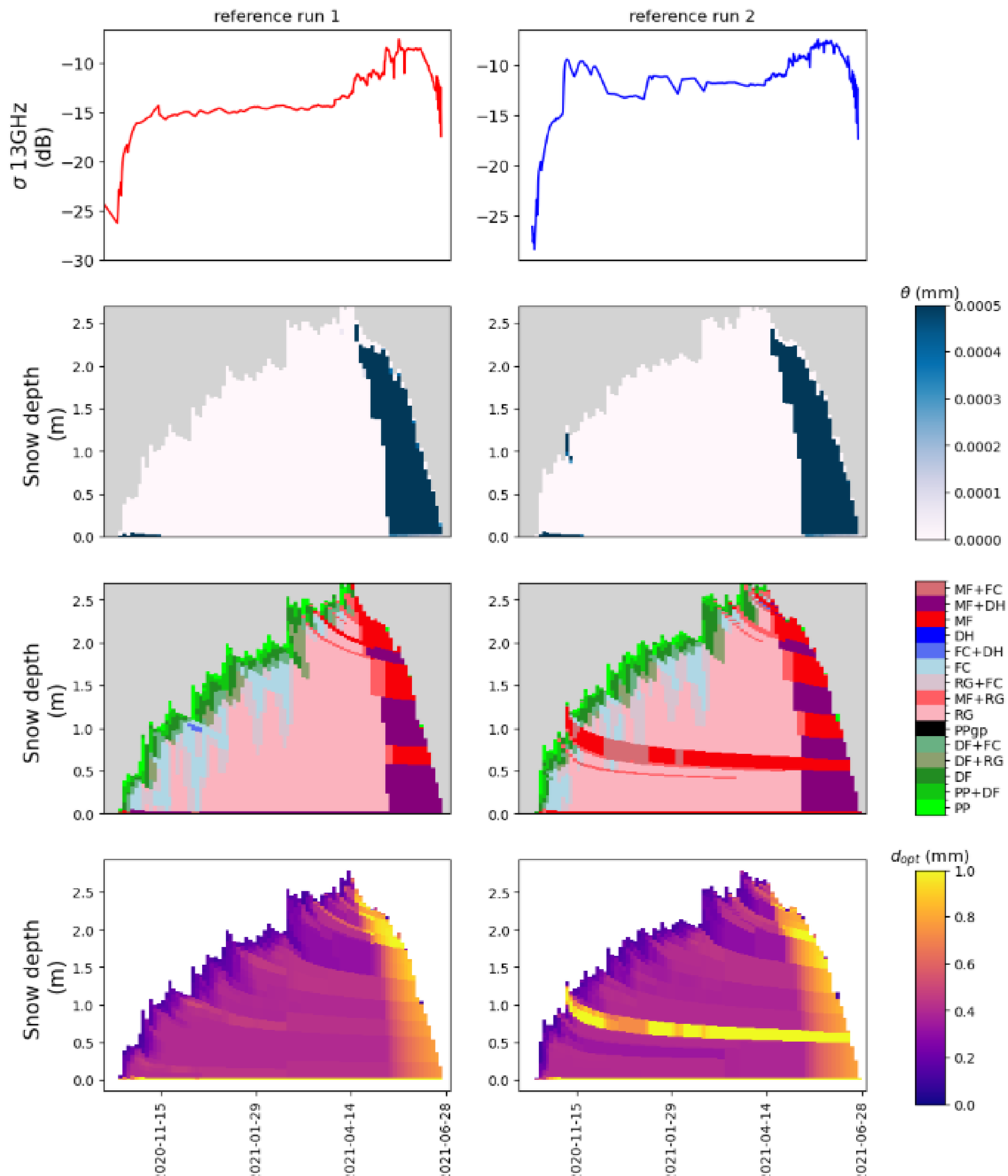


Figure R5: Rows from top to bottom: (1) temporal evolution of backscatter at 13 GHz, (2) temporal evolution of the liquid water content in the snowpack layers, (3) temporal evolution of the snow grain type, (4) temporal evolution of optical snow diameter for the reference run 1 (left column) and reference run 2 (right column).

This analysis demonstrates that the backscatter behaviour observed across different reference runs at Rogers Pass during the 2020-2021 winter can be explained by the simulated snow properties. Specifically, this behavior reflects the radiative transfer model's response to different snowpack microstructures resulting from varying meteorological forcings between reference runs. We acknowledge that these changes in microwave grain size are linked to the metamorphism scheme used in Crocus for this study and to limitations to handle the evolution of the snow optical grain size of a layer made a frozen melt forms.

References:

Brun E, David P, Sudul M, Brunot G. A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of Glaciology*. 1992;38(128):13-22. doi:10.3189/S0022143000009552

11. Fig.4, it is the systematic bias of SWE, instead of uncertainties that will greatly influence the SWE DA result. Therefore, I think the current comparison between the backscatter and the SWE assimilations is unfair.

In our synthetic experiments, we did not account for potential systematic bias in the synthetic SWE observations, as noted in lines 203-205 of the original submission. While we recognize that biases in radar retrievals exist, the particle filter assimilation method is inherently bias-blind and does not explicitly correct for systematic biases. Incorporating bias correction would be a valuable direction for future research. A sentence was added in the Section 4.3 (“Current limitations”) in the revised manuscript:

“A systematic bias in the observations was not considered, but it may affect the assimilation outcomes. When assimilating actual data, it may be necessary to implement a bias correction procedure for the observations.”

12. Line 224, what is the criterion to select THE particleS with the highest weights? How many samples were selected, and how were they combined to produce the final estimate?

As described in the text, we employed a systematic resampling method to select particles. Systematic resampling selects particles at evenly spaced intervals along the cumulative weight distribution and is a commonly used approach in particle filter applications (Kuptametee and Aunsri, 2022). We selected 100 samples to maintain a consistent ensemble size across assimilation cycles. These samples were not combined; rather, the final estimate corresponds to an ensemble comprised of the 100 selected particles.

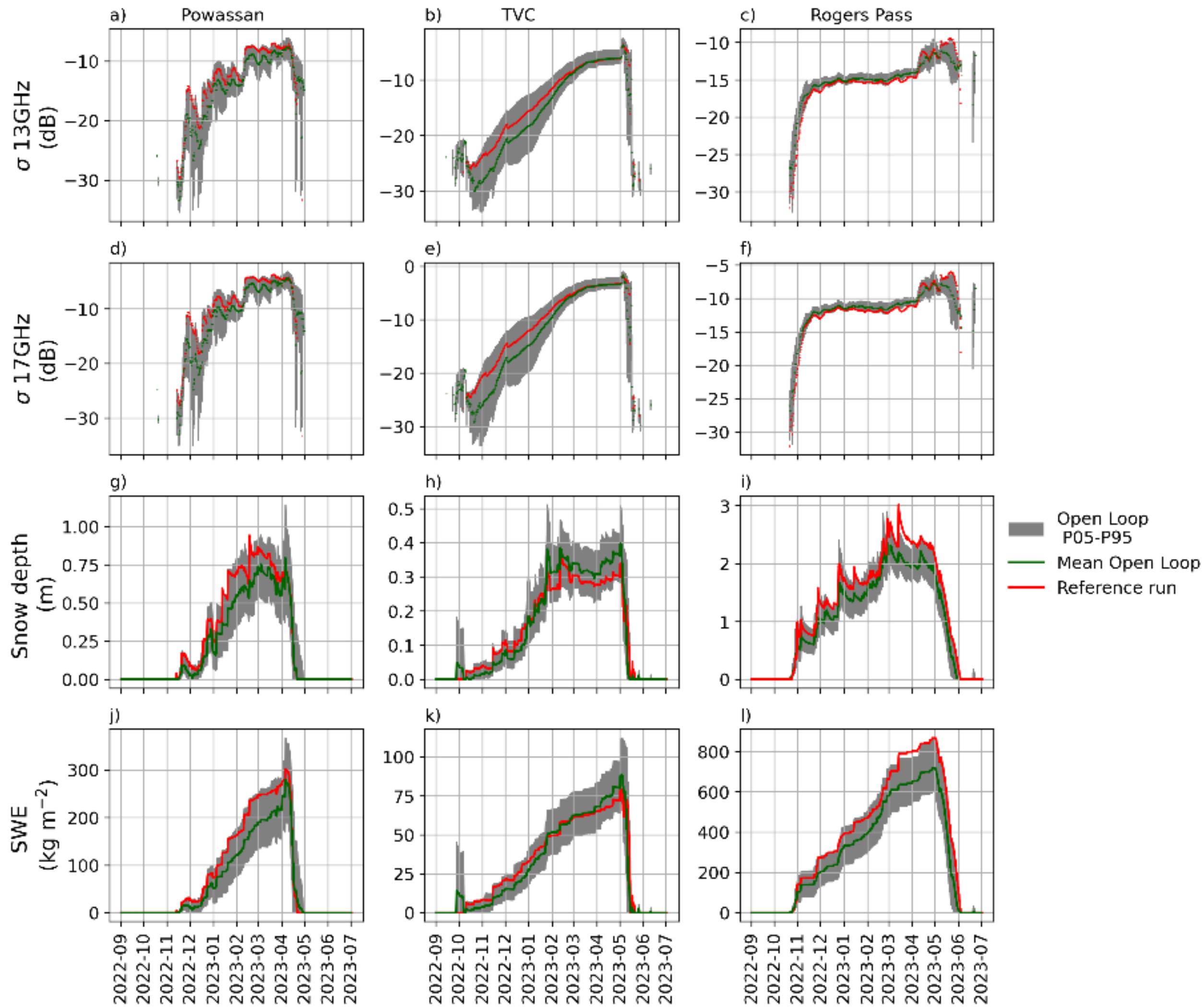
A sentence was added to the revised manuscript, it reads:
"Systematic resampling selects particles at evenly spaced intervals along the cumulative weight distribution and is a commonly used approach in particle filter applications (Kuptametee and Aunsri, 2022). This generates a new ensemble of N_e members (100 members in this study), with selection proportional to particle weights."

Reference:

Kuptametee, C. and Aunsri, N.: A review of resampling techniques in particle filtering framework, Measurement, 193, 110836, https://doi.org/10.1016/j.measurement.2022.110836, 2022.

13. In Fig.3b&3e, the spread of SWE was not narrowed down greatly using backscatter for this very shallow and backscatter-sensitive snowpack, because the ensembles are too noisy for this site (Fig.2h).

We want to highlight that the values showed in Figs. 2 and 3 do not correspond to the same winter season (2020-21 and 2022-23, respectively) and should not be compared to one another. Below is the equivalent of the Figure 2 from the revised manuscript (introduced in the first comment of the reviewer) but for the 2022-23 winter season:



To determine if the ensembles are too noisy at TVC, we calculated the signal-to-noise ratio of the SWE ensemble (ratio of the mean/coefficient of variation of the ensemble) for the 2022-2023 winter season. The mean signal-to-noise ratio for the SWE ensemble at TVC, is 4.12. Because this value is greater than 1, this indicates that the signal strength (SWE ensemble mean) exceeds the noise level.

14. Fig.2: Should the reference run be a single curve for each case?

This is indeed the case. This figure shows the 10 reference runs used in the study for each winter season. For improved clarity, only one reference run is shown in Figure 2 of the revised manuscript (see response to a previous comment) and Figures S1, S2, and S3 show all the 10 reference runs for each winter season.

15. Line 390-395: It means if the SWE error retrieved from radar drops below 30%, assimilating backscatter directly will be unnecessary. However, in my experience, this is not true.

We agree that these results suggest this conclusion. However, we would like to emphasize that this is a synthetic experiment and represents the first attempt to directly assimilate backscatter observations for SWE estimation. While our results indicate limited advantages when SWE retrieval uncertainty is below 30%, there may be benefits to assimilating backscatter that are not fully shown in our idealized experiment. A more comprehensive assessment will require testing with real backscatter observations, which we are planning for future studies

16. Lines 413-414, it is true that at a single time step, the sensitivity of MODIS-reflectance is limited to shallow snow, because only in this case, the light can penetrate through the snow medium and be influenced by the low-reflectance background beneath the snow. However, it is the time series of MODIS reflecting snow-off date (or snowmelt process) that makes it workable for deeper snowpacks.

We would like to clarify that in this sentence, we are citing the work of Revuelto et al. (2022), who assimilated MODIS surface reflectance observations, not snow cover information (snow-on/snow-off) from MODIS.

17. Line 426, for the higher sensitivity to meteorological forcing (MF) uncertainties in the early snow season, is it because the snow is an accumulated result of snowfalls, and the noises of MF have not been "smoothed" before the mid snow season?

You are correct. Precipitation amount has a greater impact during the accumulation period, as do air temperature and radiation, which determine whether the initial snow accumulations persist or melt. This explains the larger ensemble spread observed during this phase of the snow season

18. Line 614, the information for Madore et al.(2023) is not complete.

This citation is for a pre-print manuscript. The reference was updated.

Minor points:

1. Lines 207-209: Fig. 2 shows …

   Modified. It now reads: “Figure 2 shows an example of a randomly chosen reference run was used to generate the synthetic observations for assimilation. In total, 10 such reference runs were generated at each site for each winter season.”

2. Line 212: reached a plateau - reads not academic.

This is corrected. The new sentence reads: “While the backscatter at Powassan and TVC increased with SWE, it saturated early in the season at Rogers Pass, reaching a nearly constant value that barely changed with SWE”

3. Line 448: 10? or, 100?

   This is indeed 10 reference runs as specified in the manuscript.

4. Line 457: on par?

   This is modified. This sentence now reads: “The SWE estimates obtained with backscatter assimilation were comparable with estimates from assimilating synthetic SWE observations with high uncertainties (> 20 %)”