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

Reviewer 1:

We thank Reviewer 1 for their positive comments.

The line numbers specified in the responses correspond to the line numbers of the newly submitted manuscript without track change.

Thank you for the detailed clarifications replied by the authors. I have understood the methods and results much better.

Here are some further comments for the 2nd-term review.

1. Is Figure 2 a time-series mosaic result of sequential DA? Do we need to mark the observation times (every Monday) in this figure?

Figure 2 is not a mosaic results of sequential DA but represents the spread of the open loop simulations (between 5th and 95th percentiles) comprised of 100 members (without data assimilation), with the mean of the open loop in green and one of the ten reference runs in red. Figures S1, S2, and S3 (supplementary materials) show the same figure but with the 10 reference runs in red. We debated showing the observation times; however, we finally decided to not show them to increase the clarity of the figure.

2. In sequential DA, will biases of SWE in early snow season persistently propagate into the late snow season?

Thank you for your question. We interpret the question as addressing whether biases in SWE (either from observations or model states) early in the season systematically propagate into late-season estimates.

The sequential particle filter mitigates bias propagation between SWE estimates (after assimilation) and observations, as a bias at one observation time should be corrected at the next observation time if the spread of the members after the resampling step is large enough. We have demonstrated that our particle filter approach can recover from degeneracy happening early in the season (Section 4.2), showing that the spread of the particles after assimilation is large enough.

However, we acknowledge that biases can persist in the season if (i) a systematic bias in the forcing exists, or (ii) there is a persistent bias in the observations. Concerning (i), we have shown that the open loops properly represent the truth (reference runs) in Section 2.3.2 and we therefore do not think that there is systematic biases in the forcing. As for (ii), in our study, we have neglected systematic biases in the observations.

3. In Figure R1, the values for l_mw-avg in mm are very small. Is it correct?

You are indeed correct. l_mw-avg in Figures R1 and R3 were in meters and not in mm. However, this does not change the conclusions of the analysis carried during the previous round of reviews.

The new Figures R1 and R3 are as follows:

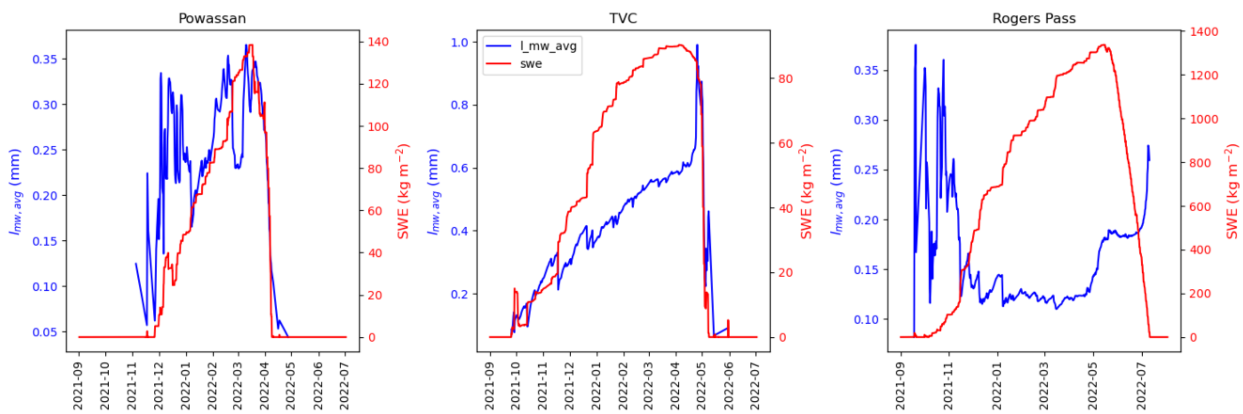

Figure R1: Temporal evolution of the average microwave grain size and bulk SWE for the 2021-2022 winter for one randomly chosen reference run, at the three study sites.

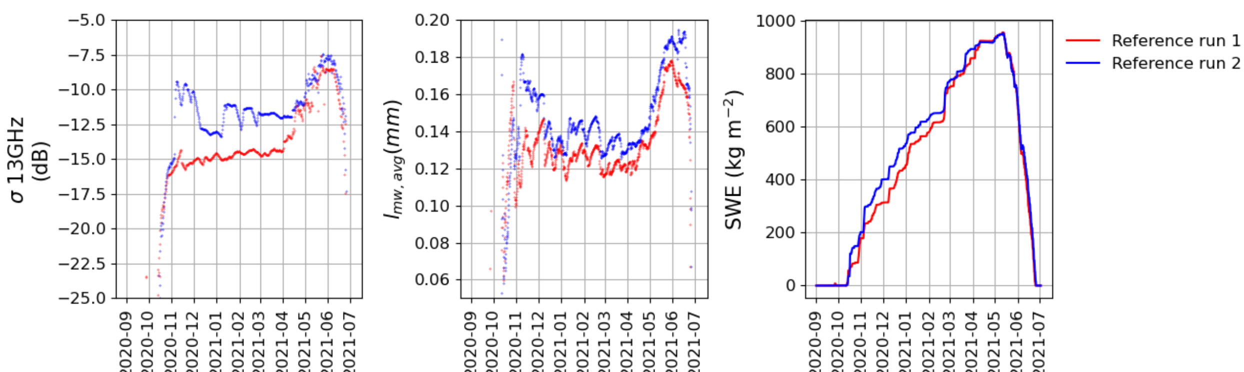

Figure R3: Temporal evolution of the backscatter at 13 GHz (left column), average microwave grain size (middle), and SWE (right) for reference runs 1 and 2.

4. Thank you for your clarification to my previous question 10. I agree that as reflected in Figure R4, the melt-freeze crust formed by high-temperature precipitation in the early winter can result in a dramatic increase in backscatter. I agree that both the simulated snow profile and backscatter are reasonable, and the sensitivity of CROCUS to this event is understandable.

Thank you for your comment.

5. For my previous point 15, we misunderstood the difference between SWE error (bias compared to the reference SWE) and the SWE uncertainty (the degree that DA trusts the observation). However, your replies to the 2-nd reviewer (Lines 312-314) have provided some clarifications.

Thank you for your comment.

6. Line 232 in the change-track manuscript: spread of truth, or truth?

This sentence was made more clear, it now reads :

L. 222-224 “The spread-skill (ensemble spread divided by RMSE) of the ensembles and the climatological variance condition (mean of the member variances divided by the variance of the synthetic truths at the observation times) were calculated to assess ensemble reliability”

7. At the end of section 2.3.3, line 262, what will the algorithm do if degeneracy happens. It reads incomplete here. In addition, 7% is a very low threshold. According to references, thresholds of 30-50% are common, and at least 20% is reasonable. Same at line 496.

In this study, nothing is done when degeneracy occurs. Degeneracy is a common issue with particle filters. However, we showed that our particle filter is robust and can recover from degeneracy, in parts by creating an ensemble with a large spread through the resampling method used and by attributing new perturbations in the forcing after an assimilation time step. Some techniques exist to avoid degeneracy, such as using an inflation technique of the observation error covariance matrix (e.g. Larue et al., 2018; Cluzet et al., 2021). No such technique was used in our study, but we showed that degeneracy was not a common occurrence and mainly happened with ephemeral snow events.

There was indeed a misinterpretation of the studies by Larue et al. (2018) and Cluzet et al. (2021), who used threshold between 14 % and 20 %. We corrected this in the manuscript and modified Figure S8 to use a threshold of 20 %. The new Figure S8 shows that even with a threshold of 20 %, degeneracy would happen only at the beginning or end of the snow seasons during ephemeral snow events, and the current conclusion of the manuscript would not change.

The following sentences were modified in the revised manuscript:

L. 251-252: "In our study, we considered that degeneracy happened when N_{eff}/Ne is below 20 %, similarly to Larue et al. (2018a) and Cluzet et al. (2021), which considered thresholds between 14 % and 20 %"

L 461-462: "In contrast, we observed that the algorithm typically degenerated either early in the snow season or towards the end of the melt season, with Nef f values dropping below the 20 % threshold (Section 2.3.3)"

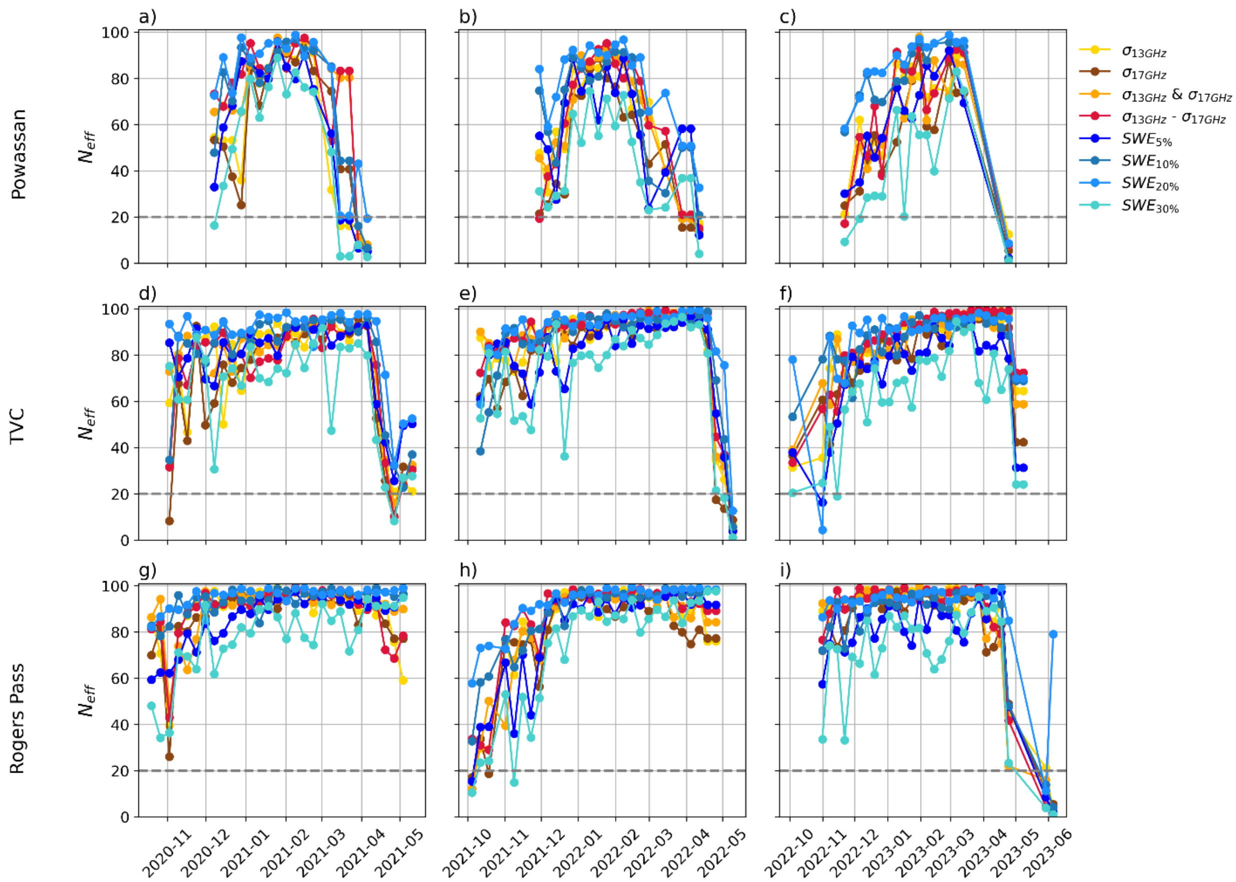

Figure S8. Evolution over time of the effective sample size (Nef f , Eq. 3 of the main manuscript) averaged over the 10 iterations for eachwinter season and for each site. The horizontal dashed line represents the 20 % threshold below which degeneracy is assumed to occur.

8. Lines 325-327 in the change-track manuscript: two 'but's.

Thank you. The new sentence reads:
L. 316-319: "The assimilation of backscatter at either frequency provided similar results in terms of mean normalized CRPS, but the assimilation of the backscatter at 17.25 GHz

slightly improved the estimates at Powassan (+1.5 % compared to 13.5 GHz for SWE estimates) while the assimilation of backscatter at 13.5 GHz worked better at Rogers Pass (+3 % compared to 17.25 GHz for SWE estimates) and at TVC (+1.5 % compared to 17.25 GHz for SWE estimates)."

9. It is very interesting to see that backscatter difference works better than dual frequencies for the deep snowpack at Rogers Pass. Could you add the difference of 13-17 GHz in Figure 2 and S1?

Figures 2 and S1 do not show any assimilation results, therefore, the results from assimilating the difference of backscatter cannot be shown in this figure. However, a figure similar to Figure 3 was created to show the assimilation results of assimilating both backscatter frequencies and the difference of backscatter frequencies. This new figure was added to the supplementary materials as Figure S4:

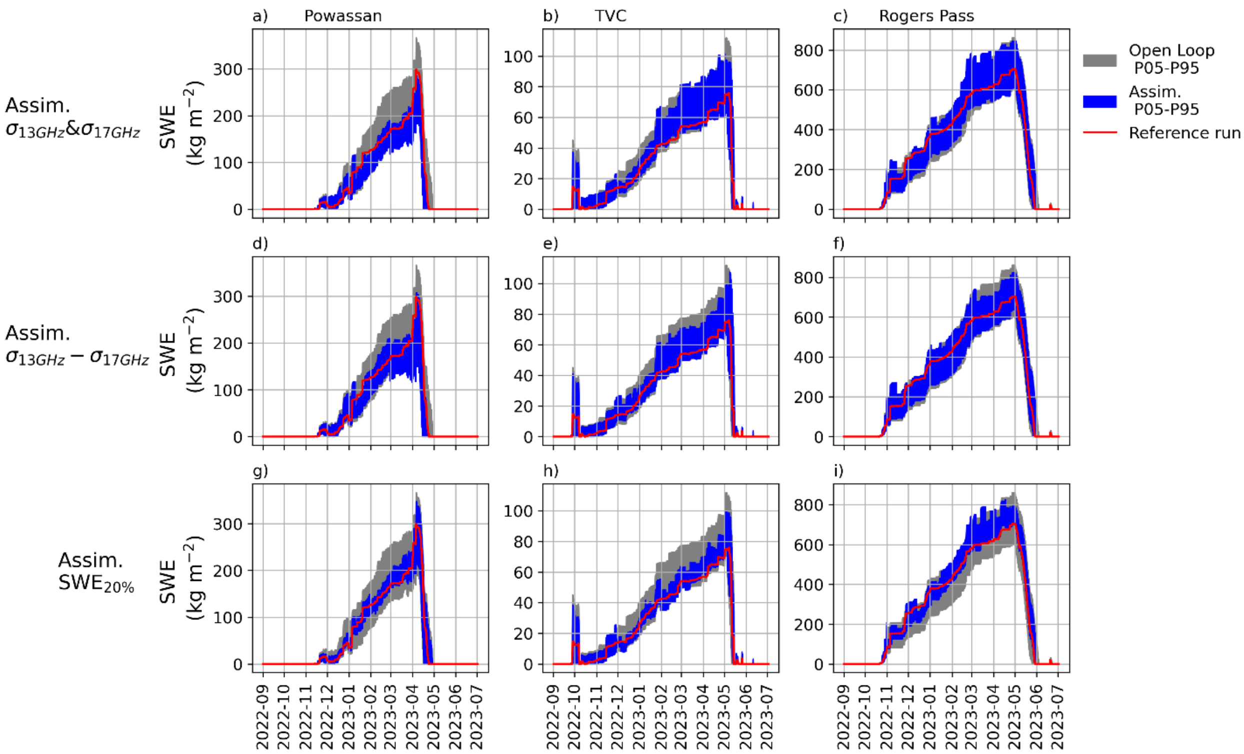

Figure S4. Results of the assimilation of the reference run #1 at (a,d,g) Powassan for the 2022-2023 snow season, (b,e,h) TVC, and (c,f,i) Rogers Pass. (a,b,c) are the results of assimilating both backscatter synthetic observations at 13.5 GHz and 17.25 GHz simultaneously, (d,e,g) for the assimilation of the difference if backscatter synthetic observations at 13.5 GHz and 17.25 GHz, and (g,h,i) for the assimilation of SWE synthetic observations with an error of 20 %. The gray envelope shows the spread between the 5th and 95th percentiles of the open loop (OL) without assimilation, while the blue envelopes show the results of the assimilation (5th to 95th percentiles) and the corresponding reference run is in red.

These sentences were added to the manuscript:

L. 303-310 “Figure S4 (supplementary material) presents the assimilation results for the same winter season and reference run, but considers additional assimilation configurations: the simultaneous assimilation of both backscatter frequencies, the assimilation of the backscatter frequency difference, and the assimilation of SWE with am observation error of 20 %. For this case, the assimilation results obtained using the various backscatter frequency combinations were largely consistent with those derived from single-frequency assimilation, except for improved SWE estimates at Powassan when both frequencies were assimilated simultaneously compared to assimilating the 13.5 GHz backscatter alone. Furthermore, the assimilation of SWE with a 20 % observation error resulted in a larger ensemble spread in the SWE estimates relative to the case in which a 10 % error was prescribed.”

10. Do results in Figure 5 actually imply an increased accuracy of weather forecast in mid-winter?

Figure 5 implies that the SWE estimates from the background particles and those of the analysis are similar in mid-winter in terms of CRPS., that is, the assimilation leaves the background particles almost unchanged at those times. This was stated in L. 344-3346: “At all sites, the assimilation algorithm did not improve on the background particles in the middle of the winter when the SWE at each site was the highest”

On the other hand, Figure S5 shows that the assimilation improves SWE and snow depth estimates across the whole season as stated in L. 360-361:

“The normalized CRPS against the open loop (Fig. S6 of the supplementary material) did not show this dependency on seasonality, meaning that the SWE or snow depth analysis consistently improved upon the open loop across the season.”

Reviewer 2:

The revised manuscript is a nice improvement over the initial submission. I especially appreciate the work that went into the new Table 4 showing the spread skills. I am convinced by that argument for the reliability of the ensembles and I found the discussions in the two Dirkson and Buehner (2025) papers very interesting. I am also glad the authors caught the error in Figure 7, which looks much more significant/convincing in the revised version. Finally, thanks for adding the paragraph in lines 346-350 about the observation uncertainty vs. ensemble behavior. I found that to be an interesting result in the initial submission and I appreciate the more detailed explanation. I recommend publication as-is.

We thank Dr. Ross Palomaki for his positive feedback on our revised manuscript.