<u>Comment on egusphere-2025-978'</u>, Anonymous Referee #1, 12 Apr 2025

This paper investigates data assimilation (DA) of Sentinel-1 snow depth (S1) and snow disappearance date (SDD) over the East River Basin in Colorado. Overall, the paper is well written and of interest to the science community. The authors find that assimilating S1 data provides limited benefit, whereas SDD assimilation significantly outperforms S1 DA in improving model performance.

We thank the reviewer for taking the time to review our manuscript and for the positive feedback on the writing and relevance of the study. Our responses to the specific comments are provided below.

<u>Main comments:</u>

I disagree with the authors broad statement that S1 snow depth (SD) assimilation has "limited utility." Instead, I encourage the authors to clearly explain why the S1 DA does not work well in THIS TEST CASE. These are some main points to support this comment:

1. We know that S1 has limitations in wet snow conditions (known issue), but it can still be valuable for accumulation phases or early-season estimates. The authors say that “basins like the ERB receive significant snowfall after January, which reduces the early-season window's ability to predict SWE reliably later in the year (e.g., April-onward).” So, does this suggest that the East River Basin might not be the best place to test the utility of S1 in general?

Author’s Reply:

We appreciate the reviewer’s feedback and acknowledge that our analysis is limited to a single basin, as previously noted in the discussion section. While Lievens et al. (2019) suggested that Sentinel-1 products perform well across the Northern Hemisphere, there is a growing body of independent studies that are showing lower performance, particularly in the western U.S. (e.g., Hoppinens et al., 2022; Broxton et al., 2024; Ying et al., 2025). Although our study is limited to a single basin, it is consistent with the broader emerging result. In particular, the East River Basin in Colorado serves as a representative and neutral test site for evaluating the performance of Sentinel-1 in snow-covered mountainous regions. This basin presents notable challenges for remote sensing due to its complex topography and occurrences of late-season snowfall. However, it also possesses attributes conducive to Sentinel-1 performance, including moderate forest canopy cover and a cold, dry winter climate that reduces signal attenuation and moisture-related interference. In the revised manuscript, we will clarify that Sentinel-1 performance is regionally dependent and emphasize why the East River Basin represents a valuable case study. While it may be valuable in other basins not assessed here, it is not reliable in much of the Western U.S., including our study area. Therefore, we emphasize the limited utility of this product in much of the Western U.S. and recommend its use with caution when generalizing to the broader Northern Hemisphere.

2. Related to the above point, the current evaluation (comparing S1 DA to ASO lidar data in melt season) is not fair to the “good” (early season) S1 observations, which instead were taken earlier

in the season. Consider: 1) Separating evaluation into accumulation and melt periods; 2) Reporting S1 DA performance specifically during accumulation when S1 is most reliable.

Author's Reply:

We are individually reporting errors of each flight date (near peak accumulation and in the melt season); however, we can refer to it clearly in text to make it a fair comparison. Note that we lack "early season" (i.e., before late March) measurements from ASO. Our SNOTEL analysis reveals that a fundamental challenge with assimilating Sentinel-1 snow depth from the early season – since that is the period when snow is more likely to be dry (higher quality SAR measurement) but may be less informative for later season conditions (see Lundquist et al., 2023).

3. Some of the paper conclusions can be associated with the assimilation scheme and not entirely to S1. I think the error analysis (re their question #1) informs that the errors vary over time, is this analysis informing the choice of what measurement error is chosen in the DA scheme? Is the measurement error dynamic (varying with snowpack conditions as suggested by Fig. 4) or constant? The description and choice of the measurement error is of critical importance to indeed evaluate the utility of S1 data in a DA system.

Author's Reply:

We appreciate this important observation. The current Particle Batch Smoother implementation uses a constant measurement uncertainty, which does not reflect the time-varying nature of S1 error observed in our analysis. To address this, we tested several uncertainty values ranging from 10cm to 90 cm and used the most representative ones. However, to account for dynamic errors, in the revised manuscript we will test an alternative assimilation approach (e.g., Particle Filter), spatially and temporally, that can incorporate dynamic measurement error. If results show a significant difference, we will revise our approach accordingly. Otherwise, we will include these tests in the supplementary material to clarify whether our conclusions are independent of the choice of assimilation method.

4. In a DA experiment there are always three players: the DA scheme, the model, and the S1. It is not clear how each of them contributes to the results found by the authors. Much responsibility is given to the observations, which could be, but the contribution of DA scheme is not discussed (see for example previous point), nor the models errors/performances are reported. In all tables, authors should - at the bare minimum - inform the readers about the performances of the model. I recommend to always report model (prior to the assimilation) statistics alongside assimilation results. The readers should be able to see how much improvement/degradation is from DA vs. the model skill. I recommend adding model (prior) estimates in all relevant figures (especially Figs 5, 6, 7, 10) and tables (3, 4, 5).

Author's Reply:

We appreciate the valuable suggestion and agree that a complete evaluation involves the DA scheme, the model, and the observations. As noted in our response to Comment 3, we plan to test an alternative DA approach to assess its influence on the results. Regarding the model, we chose not to use it as the primary baseline in this study in the original submission. Instead, we use the DA results with observed Snow Disappearance Date as the reference for performance evaluation. We conducted internal tests comparing the assimilation results to the model alone and found that assimilation performs slightly better. However, to maintain a focused evaluation of Sentinel-1 utility relative to high-accuracy observations, we chose not to include model results in the main text in the original submission. In response to this comment, we will add model (prior) performance in the supplementary material and reference it in the captions and discussion of Figures 5, 6, 7, 10, and Tables 3, 4, and 5.

A few more comments:

- Line 63: Also add the more recent paper by Lievens et al., 2022

Author's Reply: We will make this change.

- There is a contradicting reporting of errors in line 90 associated with S1 vs lines 162-165

Author's Reply: We thank the reviewer for pointing out this apparent inconsistency. The values reported in Line 90 refer to RMSE from Hoppinen et al. (2024) (~0.92 m) and Lievens et al. (2022) (~0.25 m), while the values in Lines 161–165 refer to mean absolute error (MAE) across varying snow depth ranges, as reported by Lievens et al. (2022). We will revise the text to clarify the difference in error metrics (RMSE vs. MAE) and ensure consistent and non-contradictory reporting to avoid confusion.

- Line 237: "Lower and upper bounds values for precipitation are selected to ensure realistic" how are these bounds implemented? i.e., if a particle ends up being sampled higher than a bound is it set to the upper limit, resampled, or other? Please add explanation

We did not apply any resampling. Instead, we tested multiple sets of bounds for the ensemble generation (i.e., perturbations) and selected values that were wide enough to ensure no observations fell outside the range during assimilation. Since we did not perform any downscaling of precipitation, we chose conservative (higher-than-needed) bounds to account for potential variability and uncertainty. We will revise the manuscript and add number of times the observations are within the bounds.

- Fig. 4: how is the range of errors defined? From the daily values? Please add

Author's Reply: The range of errors in Figure 4 represents the monthly mean ± one standard deviation, calculated from daily error values. We will clarify this in the figure caption.

- Figure 5: I assume the gray shading is the spread of the prior particles? What about the posterior? Is it ZERO? Can you add also the posterior spread? My fear is that the chosen means. error (which is critical to know in this paper yet struggle to find what value was used) is likely just too small.

Author's Reply: The gray shading in Figure 5 represents the prior particle spread (one standard deviation). The posterior spread is not currently shown but is not zero. To address this, we will revise the figure to include the posterior spread as a shaded region (with a different color) around the posterior mean lines (Hs-F and Hs-E), allowing readers to visualize the reduction in uncertainty after assimilation. We will also clarify the measurement error used in the data assimilation in the methods section and figure caption to ensure transparency.

- Also in Fig 5, 8 legend, "particle" should not be a gray line rather a gray box, and it should also be called "prior particle spread" or something like this.

Author's Reply: Thank you for pointing this out. We will make this change.

- Table 4: How relevant is the correlation metric in this context? Isn't this primarily driven by model and meteo forcings rather than the DA of snowpack early in the season? Similarly, for table 5, why would one observation only (SDD) lead to such a higher value with respect to the values reported in Table 3?? Wouldn't the correlation values be just an artifact of the model temporal variability?

Author's Reply: Correlation is used specifically to evaluate spatial patterns in snow depth across the domain on particular aligned dates. We recognize that correlation-based metrics can be influenced by model structure and meteorological forcing hence, to provide a more comprehensive assessment, we also report RMSE, MAE, and other metrics, which directly capture magnitude and bias errors. Regarding the higher R^2 values observed with SDD assimilation (Table 5), we interpret this as a result of the strong spatial constraint provided by SDD observations on snow disappearance timing. We will clarify the purpose and interpretation of R^2 in the manuscript to avoid confusion