Response to Reviewer # 1

General Comments:

Comment # 1.1

The manuscript is excellent in the areas of scientific significance, quality, reproducibility, and presentation quality. It makes a significant contribution to the literature on the development of cryospheric data assimilation methodologies. The substantive comments listed below are suggestions for improving the manuscript in terms of clarity or elaboration on results. They are enumerated in order of appearance in the manuscript. The minor text editing suggestions are to help remove any typographical or other errors in the final manuscript..

Reply:

We thank the reviewer, Steven Margulis, for his feedback and constructive comments on our manuscript. Please find below a point-by-point response to all issues raised by the reviewer, including changes made when necessary. Our manuscript has improved substantially thanks to these helpful suggestions.

Substantive comments:

Comment # 1.1

Figure 1 does not appear to be referenced anywhere in the manuscript text. It is worth adding a sentence or two that walks the reader through the schematic figure..

Reply:

We thank the reviewer for spotting this omission, we have now added the following paragraph just above Figure 1 at the end of Section 2.6.2.

Changes:

The overall workflow of the AdaPBS algorithm is visualized in Figure 1. As with the standard PBS, AdaPBS is initialized by sampling a particle ensemble from the prior, running this ensemble through the forward model, and computing self-normalized importance weights via (36). Ensemble diversity is then diagnosed using the effective sample size N_{eff} in (22). If N_{eff} exceeds a predefined threshold, the algorithm stops successfully, outputting this diverse ensemble as the posterior approximation. Otherwise, it adapts a new sampling proposal using clipping (23) and ensemble statistics via (24) and (25). A new particle ensemble is drawn from this adapted sampling proposal, run through the forward model, and weights are obtained using (36), now incorporating the entire particle history via a mixture proposal (17). Adaptation continues until the ESS-based diversity threshold is met or a predefined maximum number of iterations is reached.

COMMENT # 1.2

Section headers 3.2 and 3.3 respectively: Consider adding "Experiment 1: ..." and "Experiment 2: ..." to clarify to the reader there are two main experiments performed in this study

Reply:

We thank the reviewer for this suggestion and have modified these section headers accordingly.

COMMENT # 1.3

Sections 3.2 and 3.3: In both sets of experiments, you should identify what the prescribed snow depth measurement errors were. This provides important context when looking at the results in terms of how well the snow depth estimates "should" fit the observations.

Reply:

We thank the reviewer for pointing out this omission. We have added the following to clarify for Experiment 1

Changes:

These drone-acquired data and forcing datasets have previously been used in data assimilation studies [Alonso-González et al., 2022, Alonso-González et al., 2023] ~~.~~ ~~The~~ Following these studies and the findings of Revuelto et al. [2021], we assume an observation error standard deviation of $\sigma_y = 0.2$ m for the drone-based snow depth retrievals. The corresponding observation error variance σ_y^2 is used to construct the diagonal observation error covariance matrix $\mathbf{R}$ in the likelihood (4).

and Experiment 2:

Changes:

In the second set of experiments, we conducted several site-level simulations using a physics-based snow model of intermediate complexity, the Flexible Snow Model ~~[Essery et al., 2025, FMSM2;]~~[Essery et al., 2025, FSM2;]. In these experiments, we assimilated hourly observations of snow depth at six different geographical locations using both ~~the~~ AdaPBS and ES-MDA for comparison. For simplicity, we assume the same observation error standard deviation of $\sigma_y = 0.2$ m as in Experiment 1. While this is likely to be an overestimate of individual measurement errors at these sites, inflating independent observation errors in our model helps implicitly account for the fact that the actual errors in these data are correlated in time, which reduces their information content [Evensen et al., 2022].

COMMENT # 1.4

Sections 3.2 and 3.3: For the ES cases, it seems from the results you are updating/estimating parameters and not just states. If so, you should explain how this is done, e.g., via state augmentation or another approach.

Reply:

We thank the reviewer for pointing out this lack of clarity regarding the ES implementation and the forcing formulation more generally. It is mentioned briefly later on page 25 (Section 4.1) that the ES relies on a forcing formulation of the data assimilation problem [Evensen et al., 2022]. In fact, all the algorithms we test here rely on this same formulation, which we have now clarified in the revised manuscript. As Alonso-González et al. [2022] points out, the forcing formulation has implicitly become the de facto standard for snow data assimilation. In this approach, when combined with a strong constraint assumption, only the internal model parameters and the forcing (rather than the state directly) are considered uncertain and jointly encoded in a lumped parameter vector θ. As such, the states obtained by running the dynamical model $\mathcal{M}$ are deterministic given a particular setting of this parameter vector. This was briefly discussed in the relevant Section 2.2 on cryospheric inverse problems, but we have now clarified this further by adding the following to this section as follows:

Changes:

~~Thereby, using the~~ In general, these parameters can include internal model parameters, initial conditions, and/or forcing terms. This corresponds to a widely used strong constraint version of the forcing formulation of cryospheric DA, wherein model states are completely determined by this parameter vector and the forward model [Alonso-González et al., 2022]. We adopt this approach throughout without loss of generality, as the forcing formulation can accommodate a weak constraint with model error [Evensen et al., 2022].

Reply:

We have also added additional information in terms of the formulation of the ensemble Kalman methods (both ES and ES-MDA) by modifying the dedicated section 2.4 as follows:

Changes:

1. Run the forward model to obtain the predicted observations $\widehat{\mathbf{y}}_i^{(\ell)} = \mathcal{G}\left(\theta_i^{(\ell)}\right)$ from the state $\mathbf{x}$ (see Section 2.2).

2. If $\ell < N_a$, perform a tempered ensemble Kalman update step

$$\theta_i^{(\ell+1)} = \theta_i^{(\ell)} + \mathbf{K}^{(\ell)}\left(\mathbf{y} - \widehat{\mathbf{y}}_i^{(\ell)}\right), \tag{1}$$

where $\mathbf{K}^{(\ell)}$ is the tempered ensemble Kalman gain for iteration ℓ obtained from (ensemble) covariance matrices [Evensen et al., 2022].

The steps above are implicitly carried out for all $i = 1, \ldots, N_e$ ensemble members, with $N_a + 1$ iterations of the prediction step and N_a iterations of the update step. The update step itself is essentially free, so the total cost of the ES-MDA is $(N_a + 1) \times N_e$ forward model simulations where it is possible to parallelize across the ensemble dimension in each iteration ℓ. ~~Recalling~~ Recall from Section 2.2 that several steps (observation, dynamics, transformation) are baked into the forward model $\mathcal{G}(\cdot)$. Therein, it is ~~nonetheless~~ the cost of running the dynamical model $\mathcal{M}(\cdot)$ to obtain the ensemble of hidden model states $\mathbf{x}$ of interest that completely dominates the computational burden of $\mathcal{G}(\cdot)$ ~~.~~ in step 1.

Comment # 1.5

Section 3.2, p. 22 when discussing the prior parameters: Presumably the perturbed parameters are uncorrelated. If so, it may be worth mentioning that explicitly.

Reply:

We thank the reviewer for pointing out this need for clarification. We have added the following to make this assumption explicit.

Changes:

For simplicity, and without loss of generality for the DA, we assume *a priori* parameter independence, such that the joint prior factorizes into the product of its marginal prior distributions. This assumption can be relaxed using additional background knowledge on parameter dependence [e.g., Pirk et al., 2022, Alonso-González et al., 2023]. Yet, even with prior independence, correlation between parameters can be inferred in the posterior via the likelihood [Cao et al., 2025].

COMMENT # 1.6

Section 3.3: Was the "gold-standard" MCMC approach used as a reference for this experiment? If so, perhaps make that clear, but if not, explain why not. Perhaps related to this, is there a reason that the KLD metric is not used in these experiments? Is that because the MCMC is not applied?

Reply:

The reasoning is exactly as the reviewer suggests: MCMC is very time consuming to run even with intermediate complexity snow models like FSM2 in higher dimensional parameter spaces. To clarify this we have added the following at the end of Section 3.3.

Changes:

Unlike Experiment 1 with the temperature index model, we did not run MCMC in this more complex experiment due to the high computational cost and marginal additional benefits to the analysis. As such, the KLD metric could not be used without an MCMC-based reference distribution. Instead, this experiment focused on comparing the predictive performance in state space of AdaPBS to that of ES-MDA, an advanced and competitive baseline for DA with more complex snow models [Alonso-González et al., 2022, Alonso-González et al., 2026]

COMMENT # 1.7

Section 3.3: For the model parameters that were perturbed, can you provide more context for why they were chosen?

Reply:

We thank the reviewer for raising this point. These parameters were selected as they are the main tunable internal snow parameters in FSM2 that are not related with vegetation. In response to Reviewer Comment #1.19 we now clarify in Section 3.3 of the revised manuscript that these parameters are nuisance parameters, where our primary interest is on the resulting model states, particularly snow depth. We have also added the following explanation to Section 3.3 in direct response to this comment:

Changes:

~~on top of which we also have included~~ . On top of this, we have also included the 12 main tunable internal model parameters ~~in~~ related to snow (not vegetation) in FSM2 [Essery et al., 2025] as listed in Table 1.

COMMENT # 1.8

Figures 2-5, MCMC samples in the Forcing Correction Parameters subplots: The color mapping is hard to see and not explained. I presume it is related to the density of the points at a given location in the parameter space. I don't think that is so important (and hard to see) and would suggest just making those points a single color.

Reply:

The reviewer is correct that the color map is related to the density of points in the MCMC samples. We agree that this does not add much information and have thus set all MCMC samples to blue in the revised versions of these Figures.

COMMENT # 1.9

Figures 2-5, red dots in legend box: The use of red dots in the legend (as "observations") could be confused with the red dots used in the parameter space. I suggest using different colors.

Reply:

We thank the reviewer for identifying this ambiguity, we have now changed the observations to yellow dots with a black edge to avoid confusion with the parameters and to be consistent with Figure 7 in Experiment 2.

COMMENT # 1.10

Figures 2-5, model state space subplots: Consider whether an MCMC-derived trajectory should be included as the "gold-standard" reference.

Reply:

We thank the reviewer for this great idea and have now added the corresponding MCMC posterior snow depth trajectory as a 'gold-standard' reference in state space to Figure 2-5. Consequently, we have improved the style of the plots to make them clearer, given the increased complexity of the figures. In the revised manuscript we do not compute the KLD with this reference trajectory, as it would no longer be a marginal KLD. We do, however, discuss the qualitative agreement between this reference trajectory and the 4 batch smoothing DA algorithms.

Changes:

It should also be noted that the estimated posterior uncertainty differs between ES-MDA and AdaPBS. The posterior standard deviation in ES-MDA is slightly higher ~~when~~ compared to that of ~~the~~ AdaPBS. At least in the present example, ~~the~~ these differences are nonetheless ~~quite minimal at least~~ relatively minor, especially in the model ~~space, so we do not consider it relevant for the majority of cryospheric data assimilation initiatives developed to date that often overlook an in depth posterior uncertainty analysis for simplicity and instead focus~~ state space visualized here in terms of snow depth. While quantitative probabilistic evaluation using diagnostics such as KLD is helpful, qualitative visual comparisons to an MCMC reference posterior ensemble trajectory in state space can be equally instructive. Inspecting the snow depth trajectories in Figures 2-5 shows that the patterns identified in parameter space carry over to snow depth in state space. Using the RAM-based MCMC posterior trajectory as a reference: PBS is biased and highly degenerate, ES is biased and highly under-confident, ES-MDA is unbiased yet slightly under-confident, whereas AdaPBS overlaps almost perfectly with the reference ensemble. To date, most snow DA studies focus primarily on the predictive performance of ~~the~~ point estimates such as the posterior mean~~.~~, median, or mode [e.g. Margulis et al., 2016, Aalstad et al., 2018, Alonso-González et al., 2021, Oberrauch et al., 2024]. Nonetheless, as snow DA continues to transition towards fully Bayesian methods, an emphasis on rigorous uncertainty quantification will become increasingly important for probabilistic prediction [Reich and Cotter, 2015] and decision-making [Robert, 2007]. As we have shown, benchmarking via MCMC reference posteriors [Law and Stuart, 2012] allows for direct qualitative and quantitative evaluation of more approximate posterior estimates obtained from tractable DA algorithms such as AdaPBS.

COMMENT # 1.11

Figures 2-5, model state space subplots: There seems to be an error in the date labeling convention being used on the time series plots. It is unclear and wraps so that "01/09" appears at both ends of the x (time)-axis.

Reply:

We thank the reviewer for pointing out this lack of clarity in the date labeling and have added the year as well to change from dd/mm format to dd/mm/yyyy format.

COMMENT # 1.12

Figure 3: Since ES is not technically an iterative method, I'd suggest using "Prior" and "Posterior" labels for clarity rather than "Iteration 0" and "Iteration 1"

Reply:

We thank the reviewer for this astute comment and have modified the labels accordingly.

COMMENT # 1.13

Figure 6: Perhaps explain or reference how you estimate the "kernel density estimates".

Reply:

We thank the reviewer for pointing out this lack of detail. We have now revised the caption to the following:

Changes:

Comparison of Gaussian kernel density estimates (KDEs) obtained via `scipy.stats.gaussian_kde` [Scott's rule bandwidth; Virtanen et al., 2020] of the marginal prior and marginal approximate posterior distributions of the parameters from the respective algorithms in the transformed (unbounded) space. The MCMC posterior (orange) from RAM is considered the gold-standard.

COMMENT # 1.14

p. 31, "... thousands of observations ...": This comment is specific to the batch/smoothing you are doing in this paper. Can you comment here or elsewhere on how/why this would be different (more tractable) if using a filtering approach?

Reply:

We thank the reviewer for pointing this out. We have now provided more background on how assimilating a large number of observations can be made more tractable.

Changes:

The large number of assimilated observations in these experiments with thousands of observations per annual DA window represents a challenge for any batch smoothing assimilation algorithm. On average, AdaPBS required 8 iterations per season to address this problem. In principle, such a high number of iterations in AdaPBS would be expected to result in a considerable increase in computational cost compared to the typical number of iterations used by ES-MDA in most cryospheric applications (which in our experience is around four) since a higher number of iterations involve a large number of sequential model realizations. However in this case, the computational cost of the linear algebra in ES-MDA scaled with the number of observations, resulting in a much higher execution time compared with AdaPBS. The site with the largest run times due to the longest simulation, namely WSJ with 6 water years, took 16 minutes on a single core with AdaPBS compared with 5 h using two cores for ES-MDA. Nonetheless, this conclusion should be interpreted with some caution, as both linear algebra operations and ensemble generation can be parallelized. There may be solutions to circumvent these issues depending on the problem at a hand ~~, the~~ related to the the parallelization scheme of the computing infrastructure~~and~~ , the hyperparamters (i.e. number of iterations for ES-MDA and ~~effective sample size~~ diversity threshold for AdaPBS), and further optimizing linear algebra operations [Hennig et al., 2022]. Moreover, previous experience suggests that a larger number of iterations of ES-MDA, potentially exceeding $N_a = 10$, may also be necessary to ensure convergence based on earlier work applying the ES-MDA to higher-dimensional problems [Pirk et al., 2024]. In a similar vein, the validation metrics of AdaPBS may improve if a larger ~~effective sample size~~ N_{eff} threshold is selected, at the cost of increasing the computational ~~cost.~~ expense. More generally, the large number of observations can be assimilated in a more tractable manner via Bayesian filtering [Särkkä and Svensson, 2023] and tempering [Chopin and Papaspiliopoulos, 2020] techniques. Combining adaptive particle methods with these techniques is likely to be a promising future research direction in cryospheric DA.

Comment # 1.15

p. 31, "On average, AdaPBS required 8 iterations ...": Given that the iterative/adaptive aspect of this method is novel and discussed as a benefit, is there any useful insight to be gained by presenting the interannual and inter-site number of iterations that were seen across site-years and discuss why those differences may have occurred? Perhaps a broader short discussion (not necessarily here) of when/why/where the number of iterations is expected to be small or large would benefit the reader

Reply:

We thank the reviewer for this suggestion. While the idea of a more detailed presentation of the variability in the number of required iterations is intriguing, explaining this variability in a non qualitative manner is not straightforward. As such, we have instead opted for the second suggestion by the reviewer to add a broader discussion of recommendations using the AdaPBS to a new Section 4.3 as follows:

Changes:

Broadening the discussion, we provide recommendations informed by the above snow DA experiments and recent applications of AdaPBS to permafrost [Willmes et al., 2025] and glaciers [Yang et al., 2025]. In addition to ensemble size N_e, which is foundational for all ensemble-based DA, two key algorithm-specific hyperparameters control AdaPBS: the diversity threshold $0 < \tau < 1$ and the maximum number of iterations $\ell_{max} \geqslant 1$ (Section 2.6). On the one hand, τ serves as a lower bound for the desired quality of the posterior approximation: an ensemble run that converges successfully will achieve an N_{eff} of *at least* τN_e. On the other hand, ℓ_{max} serves as an upper bound on the computational cost: AdaPBS will run for *at most* ℓ_{max} iterations with a total of $\ell_{max} N_e$ model runs, but it may converge and terminate much earlier. This upper bound is only reached if AdaPBS converges in the final possible iteration or, in the worst case, if it fails to converge within the prescribed ℓ_{max} iterations. There is thus an inherent tradeoff in these hyperparameters: a higher threshold $\tau \simeq 1$ helps improve the particle approximation of the posterior while often requiring more iterations for convergence, demanding a larger ℓ_{max} with increased computational cost. At the opposite extreme, one could simply set $\ell_{max} = 1$ to minimize computational cost, but then AdaPBS is identical to the PBS and has no way to adapt and evolve beyond degeneracy when it occurs. A useful default is $\ell_{max} = 5$, mirroring the typical cost of the ES-MDA that uses $N_a + 1$ ensemble iterations with $N_a = 4$ assimilation cycles [Alonso-González et al., 2026], but where AdaPBS imposes fewer assumptions and can stop early upon convergence.

The optimal balance between these two hyperparameters depends on the problem at hand and the available computational budget. Setting the threshold in the range $0.1 < \tau < 0.5$ is generally enough to obtain a good posterior approximation, while the number of iterations depends on the problem. This number may vary considerably with the complexity of the problem and even within a given distributed multi-year simulation. Ensuring convergence across the board may require setting ℓ_{max} to 10 or more. In practice, it is often entirely acceptable to end up with a low percentage (say $\leqslant 10\%$) of cases in which the algorithm failed to converge. We emphasize that failure to converge does not imply a complete failure of the algorithm, but rather that the particle approximation is not as diverse as desired for uncertainty quantification via (e.g.) the posterior variance. The ensemble may still provide a reasonable point estimate for the posterior mean and avoid complete degeneracy ($N_{eff} = 1$). Since the AdaPBS is designed as a direct extension of the PBS, instances of failed convergence will correspond to cases when the PBS would be at least as degenerate.

Experiments 1 and 2 demonstrate the role of problem complexity. For the simpler temperature index model with two parameters, AdaPBS exceeded the desired threshold of $\tau = 0.3$ (i.e., $N_{eff} = 30$ with $N_e = 100$) with $N_{eff} = 84$ in 5 iterations, providing a strong approximation that nearly converged already at 4 iterations (Figure 5). In Experiment 2, the intermediate complexity snow model FSM2, a higher dimensional parameter space, and numerous observations conspired to make inference challenging [Snyder et al., 2008, Morzfeld et al., 2017]. Following the 'needle in a haystack' analogy for inferring the posterior: increased model complexity may result in an oddly shaped needle, more informative data makes the needle smaller, and higher dimensionality makes the haystack bigger. Thus, AdaPBS required, on average, nearly twice as many iterations (8) to converge with the same $\tau = 0.3$ as Experiment 1. For both experiments, the number of iterations could have been reduced by lowering the bar on diversity to say $\tau = 0.1$, which can suffice in operational settings focused on prediction using point estimates rather than uncertainty quantification [e.g. Oberrauch et al., 2024] .

Valuable lessons can also be learned from the two other contemporaneous studies using AdaPBS for cryospheric DA. Willmes et al. [2025] used AdaPBS to assimilate Sentinel-2 and in-situ FSCA data into the detailed multiphysics cryospheric model CryoGrid [Westermann et al., 2023], inferring two parameters to constrain ground surface temperatures around Bayelva near Ny-Ålesund, Svalbard. To configure AdaPBS, the maximum number of iterations was set to $\ell_{max} = 8$, and the diversity threshold to $\tau = 0.3$. In this two-dimensional parameter space, across all grid cells, convergence was achieved within an average of $\ell = 3$ iterations, and only rarely did AdaPBS fail to converge. To contend with the relatively high computational cost of this CryoGrid configuration, parallel simulations were run with the ensemble size $N_e = 20$ equal to the number of locally available cores. As such, AdaPBS could optimally leverage parallelization to avoid waiting for new particles to complete while benefiting from adaptation across iterations to avoid degeneracy. For example, the total computational cost of 5 iterations with this configuration is identical to running a non-adaptive PBS with a larger ensemble of $N_e = 100$, as both would require 5 sequential batches of 20 parallel simulations to finish. This approach can be extended to spatially distributed simulations across large domains, where AdaPBS is, like all ensemble-based methods, highly parallelizable in the ensemble dimension. Furthermore, later adaptive iterations will be cheaper to run as some grid cells will have converged and can be skipped. It may also be possible to fold in new cells to replace converged cells, helping to maintain an even computational load. For very large scale applications, it is advantageous to jointly chunk simulations in both the ensemble and spatial dimensions, vectorizing within chunks and parallelizing across. Generally, multiplying the ℓ_{max} hyperparameter with typical model run times provides a useful worst-case estimate for wall-clock time for allocating computing resources.

Further lessons can be learned from Yang et al. [2025], who used AdaPBS to jointly calibrate 4 parameters controlling surface mass balance and frontal ablation in the coupled PyGEM-OGGM glacier models [Rounce et al., 2023, Maussion et al., 2019]. Therein, the evolution of 71 marine-terminating glaciers on Svalbard was inferred for the period 2000-2019 by jointly assimilating glacier-wide mass balance and terminus-positions. Using a configuration with $N_e = 200$, $\tau = 0.1$, and $\ell_{max} = 10$, AdaPBS converged successfully within 10 iterations for 87% of the glaciers considered. In terms of performance, the posterior mean from AdaPBS considerably reduced root mean squared errors compared to the prior mean for both the converged and non-converged glaciers. By extending the maximum number of iterations to $\ell_{max} = 100$, AdaPBS also managed to converge for most of the remaining glaciers. This showcases both the need for selecting an appropriate AdaPBS configuration and that infrequent non-convergence need not adversely affect overall performance. Moreover, as suggested in Yang et al. [2025], methods such as Bayesian optimization [Hennig et al., 2022] can be used to automatically tune these hyperparameters.

COMMENT # 1.16

p. 31, "The site with the largest run times ...": It would benefit the reader (here or later) to have some discussion on the scalability to spatially-distributed cases.

Reply:

We thank the reviewer for this helpful suggestion. Please see the changes immediately above in response to Comment 1.15 where we discuss scalability in spatially-distributed simulations including examples from other recent applications of AdaPBS.

COMMENT # 1.17

Figure 7: I would suggest either omitting or changing the naming of the "reference" curve. From what I gather it is a deterministic version of the open-loop and is therefore somewhat redundant to the open-loop curves, but "reference" could be misinterpreted as the target for which the estimates should be compared.

Reply:

We thank the reviewer for highlighting the potential for misinterpretation with regards to this use of 'reference' and the redundancy of this curve. As suggested, we have removed this curve from Figure 7.

COMMENT # 1.18

p. 33, "... more challenging high-dimensional setting": Here, when you say "highdimensional" are you mainly referring to the number of measurements, parameters, or both?

Reply:

We thank the reviewer for spotting this ambiguity, it is both. We propose the following changes.

Changes:

The results of AdaPBS in this more challenging ~~high-dimensional setting~~ setting with higher-dimensional parameter and observation spaces are similar to those obtained by ES-MDA~~as outlined~~ , as shown in Table 4.

COMMENT # 1.19

p. 33, results section: Given that this experiment involved a significantly larger number of parameters, I think it would be worth it to show results on how the parameter estimates differ (or are similar) between the ES-MDA and AdaPBS cases (e.g., perhaps along the lines of Figure 6).

Reply:

We thank the reviewer for this suggestion. While visualizing the parameter posteriors was instructive for the two-dimensional case in Experiment 1 (Figures 2-6), extending this approach to the 19 parameters (7 forcing, 12 internal) in Experiment 2 presents both visualization and interpretability challenges. A full visual comparison, such as a 19x19 pairwise scatter plot matrix, would be far too cluttered to be readable.

Even if we were to isolate a subset of the 19 parameters, such an analysis would not add much beyond what was found in Experiment 1. In Experiment 2, most of the parameters are so-called nuisance parameters [Jaynes, 2003] that are uncertain but not of primary interest in themselves. It is the forcing parameters, especially those related to precipitation and air temperature, that are the most uncertain (have the broadest prior range) and to which the model response is likely to be most sensitive [Günther et al., 2019]. The remaining parameters are mostly perturbed to increase the difficulty of the inference problem (the size of the haystack), to see if the AdaPBS (and ES-MDA) can still infer reasonable posterior approximations in a higher-dimensional parameter space. Our interest is primarily in the resulting posterior in state space in terms of snow depth estimates, which can be compared to observations and evaluated via the CRPS. The results show that even in this higher-dimensional setting with a more complex model (FSM2), the performance of AdaPBS can match that of ES-MDA.

To briefly explain why we did not conduct an analysis of posterior parameters in Experiment 2, we have added the following to the manuscript at the end of Section 3.3:

Changes:

In this higher-dimensional experiment, the entire 19-dimensional parameter space is too large for comprehensive and instructive visualization. An alternative could be to investigate the importance of each parameter by estimating the information gain [van Hove et al., 2026], but such parameter analysis strays beyond the focus of this experiment. Except for a handful of forcing parameters partly analyzed in Experiment 1, the majority of the 19 parameters are so-called nuisance parameters [Jaynes, 2003]. They are not of primary interest beyond their effect on the state and their role in making the DA problem more challenging. We thus focus on the performance of AdaPBS and ES-MDA for estimating the snow depth state observed at the selected ESM-SnowMIP sites.

COMMENT # 1.20

Depending on how some of the comments above are handled, I could envision a paragraph in the conclusions discussing some of the implications around: scaling up to spatially distributed applications, tradeoffs with filtering approaches, expected benefits of the adaptive nature of the AdaPBS method, etc. There are some instances of text throughout that touch on these issues, but having them in one place might help orient the reader when thinking about applications and extensions to this work.

Reply:

We thank the reviewer for this idea which fits nicely with the newly added Section 4.3 in response to Comment # 1.15 and Comment # 1.16. In response to these past comments, we have already added some discussions on spatially distributed applications to this new section. We have also added a final paragraph to this section mentioning possible extensions of the AdaPBS, some of which were touched on sporadically elsewhere in the text, as an outlook.

Changes:

The AdaPBS algorithm that we have presented relies on adaptive importance sampling [Bugallo et al., 2017], specifically AMIS [Cornuet et al., 2012], to extend the PBS [Margulis et al., 2015] by allowing it to evolve beyond collapse when degeneracy strikes. At the same time, adaptation is but one direction for improvement to make particle methods less prone to degeneracy. Likelihood tempering [Neal, 2001, Murphy, 2023], which also powers the ES-MDA [Emerick and Reynolds, 2013, Stordal and Elsheikh, 2015], inflates the observation error covariance and iteratively ensures a smoother transition from prior to posterior. Data tempering [Chopin, 2002, Murphy, 2023] randomly divides the observation vector into mini-batches and assimilates these sequentially, allowing particle methods to assimilate more data without degeneracy [e.g. van Hove et al., 2025, 2026] . Particle rejuvenation methods such as jitter [Farchi and Bocquet, 2018] or the resample-move approach using MCMC [Gilks and Berzuini, 2001] also help avoid degeneracy and are frequently used together with tempering for advanced particle methods [Chopin and Papaspiliopou . Hybrid methods are also widely used [van Leeuwen et al., 2019], for example, employing the ES-MDA to design the proposal distribution [Pirk et al., 2022]. Crucially, tempering, rejuvenation, and hybridization can all be used together with adaptation to design more robust particle methods for cryospheric DA. Furthermore, as shown theoretically in Elvira et al. [2019], particle methods relying on AMIS, such as AdaPBS, can be readily applied to the auxiliary particle filter, which may further improve operational snow hydrological forecasting systems [Oberrauch et al., 2024]. Using adaptive particle methods for spatio-temporal DA [e.g. Mazzolini et al., 2025, Alonso-González et al., 2 , will require further progress on particle localization techniques [Farchi and Bocquet, 2018] for cryospheric DA by building on the efforts of Cluzet et al. [2021]. As an extension of the PBS, AdaPBS has the potential to be directly applied to large-scale snow reanalysis efforts [Margulis et al., 2016, Alonso-González et al., 2021, Liu et al., 2021] , where $\ell_{\max}$ provides an upper bound on computational cost.

Response to Reviewer # 2 Minor text editing suggestions:

Comment # 2.1

On p. 2: "... remotely sensed products ...". Perhaps for clarity mention you are referring to passive microwave products.

Reply:

Thanks, we changed this to:

Changes:

Global passive microwave remote sensing products are able to directly retrieve SWE-related information exhibit a pixel resolution on the order of 10 km that is too coarse for many applications [Zschenderlein et al., 2023].

COMMENT # 2.2

On p. 3: "Ensembe" misspelled

Reply:

We thank the reviewer for spotting this unfortunate typo that is now fixed.

COMMENT # 2.3

On p. 5: "Unfortunately . . . case , . . . " has extra space after "case"

Reply:

Thanks, fixed.

COMMENT # 2.4

On p. 5: "ill-possednes" misspelled

Reply:

Thanks, fixed.

COMMENT # 2.5

On p. 8: Two instances of "long enough" could be replaced by "enough iterations" to indicate the meaning is related to iteration rather than time

Reply:

We thank the reviewer for pointing out this ambiguity and providing suggestions for clarification. We have modified the section as follows to increase clarity.

Changes:

To sample from the posterior with MCMC~~one merely~~ , one needs to run the Markov chain for ~~long enough~~ enough iterations to ensure that it mixes properly ~~to 'converge'~~ and converges to sampling from the posterior. In practice, although there are some ~~heuristic pointers and metrics~~ diagnostics that can be used as a guide [Gelman et al., 2013], ~~it is hard to know what is long enough~~ what constitutes enough iterations is uncertain *a priori* and can remain challenging to verify *post hoc*.

COMMENT # 2.6

On p. 23: "FMSM2;" has a typographical error.

Reply:

Fixed, thanks.

COMMENT # 2.7

Figure 2: "aditive" on y-axis is misspelled.

Reply:

Thanks for spotting this, we've now removed both "additive" and "multiplicative" in Figure 2 for consistency with Figures 3-5.

COMMENT # 2.8

On p. 28: "In our experience . . . ". This sentence seems a bit out of place here. It sounds more like a conclusion statement.

Reply:

We thank the reviewer for pointing out this inconsistency and have modified the tone accordingly to reflect that this has not been demonstrated yet at this stage in the manuscript.

Changes:

Here we introduce a new algorithm, namely the AdaPBS, as an effort to combine the advantages of the different algorithms presented above into a single tool. ~~In our experience, AdaPBS has proven~~ The AdaPBS has the potential to be a powerful ~~method, sampling~~ iterative DA method, aspiring to sample from the posterior with ~~a~~ performance similar to that of ES-MDA.

COMMENT # 2.9

On p. 32: "In a similar vein, . . . ". The word "cost" appears twice in sentence. Perhaps replace last instance with "expense".

Reply:

We thank the reviewer for this suggestion and have modified the text accordingly.

Changes:

In a similar vein, the validation metrics of AdaPBS may improve if a larger ~~effective sample size~~ N_{eff} threshold is selected, at the cost of increasing the computational ~~cost.~~ expense. More generally, the large number of observations can be assimilated in a more tractable manner via Bayesian filtering [Särkkä and Svensson, 2023] and tempering [Chopin and Papaspiliopoulos, 2020] techniques. Combining adaptive particle methods with these techniques is likely to be a promising future research direction in cryospheric DA.

COMMENT # 2.10

Table 3 caption: Replace "prior (Pri)" with just "Prior".

Reply:

We thank the reviewer for this legacy inconsistency and have changed this to just 'Prior'.

We would like to thank Steven Margulis for his constructive and detailed review. We are especially grateful to receive positive feedback from the architect of the seminal PBS algorithm [Margulis et al., 2015] that served as the foundation for AdaPBS.

On behalf of all coauthors,
Kristoffer Aalstad and Esteban Alonso-González

REFERENCES

K. Aalstad, S. Westermann, T. V. Schuler, J. Boike, and L. Bertino. Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *The Cryosphere*, 12:247–270, 2018. doi: 10.5194/tc-12-247-2018.

E. Alonso-González, E. Gutmann, K. Aalstad, A. Fayad, M. Bouchet, and S. Gascoin. Snowpack dynamics in the Lebanese mountains from quasi-dynamically downscaled ERA5 reanalysis updated by assimilating remotely sensed fractional snow-covered area. *Hydrol. Earth Syst. Sci.*, 25:4455–4471, 2021. doi: 10.5194/hess-25-4455-2021.

E. Alonso-González, K. Aalstad, M. W. Baba, J. Revuelto, J. I. López-Moreno, J. Fiddes, R. Essery, and S. Gascoin. The Multiple Snow Data Assimilation System (MuSA v1.0). *Geosci. Model Dev.*, 2022. doi: 10.5194/gmd-15-9127-2022.

E. Alonso-González, K. Aalstad, N. Pirk, M. Mazzolini, D. Treichler, P. Leclercq, S. Westermann, J. I. López-Moreno, and S. Gascoin. Spatio-temporal information propagation using sparse observations in hyper-resolution ensemble-based snow data assimilation. *Hydrol. Earth Syst. Sci.*, 27:4637–4659, 2023. doi: 10.5194/hess-27-4637-2023.

E. Alonso-González, A. Harpold, J. D. Lundquist, C. Piske, L. Sourp, K. Aalstad, and S. Gascoin. Ensemble-based data assimilation improves hyperresolution snowpack simulations in forests. *The Cryosphere*, 20:209–225, 2026. doi: 10.5194/tc-20-209-2026.

Monica F. Bugallo, Victor Elvira, Luca Martino, David Luengo, Joaquin Miguez, and Petar M. Djuric. Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Processing Magazine*, 34(4):60–79, 2017. doi: 10.1109/MSP.2017.2699226.

W. Cao, K. Aalstad, L.S. Schmidt, S. Westermann, and T.V. Schuler. Glacier data assimilation on an Arctic glacier: Learning from large ensemble twin experiments. *Journal of Glaciology*, 2025. doi: 10.48550/arXiv.2502.09314. Preprint in review.

N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89:539–552, 2002. doi: 10.1093/biomet/89.3.539.

N. Chopin and O. Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer, 2020. doi: 10.1007/978-3-030-47845-2.

B. Cluzet, M. Lafaysse, E. Cosme, C. Albergel, L.-F. Meunier, and M. Dumont. CrocO_v1. *Geosci. Model Dev.*, 14:1595–1614, 2021. doi: 10.5194/gmd-14-1595-2021.

J.-M. Cornuet, J.-M. Marin, M. Antonietta, and C.P. Robert. Adaptive Multiple Importance Sampling. *Scandinavian Journal of Statistics*, 39:798–812, 2012. doi: 10.1111/j.1467-9469.2011.00756.x.

Victor Elvira, Luca Martino, Monica F. Bugallo, and Petar M. Djuric. Elucidating the auxiliary particle filter via multiple importance sampling [lecture notes]. *IEEE Signal Processing Magazine*, 36(6):145–152, 2019. doi: 10.1109/MSP.2019.2938026.

A. A. Emerick and A. C. Reynolds. Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55:3–15, 2013. doi: 10.1016/j.cageo.2012.03.011.

R. Essery, G. Mazzotti, S. Barr, T. Jonas, T. Quaife, and N. Rutter. A flexible snow model (fsm 2.1.1) including a forest canopy. *GMD*, 18:3583–3605, 2025. doi: 10.5194/gmd-18-3583-2025.

G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen. *Data Assimilation Fundamentals*. Springer, 2022. doi: 10.1007/978-3-030-96709-3.

A. Farchi and M. Bocquet. Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, 25:765–807, 2018. doi: 10.5194/npg-25-765-2018.

A. Gelman, J.B. Carlin, H. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. CRC Press, 3 edition, 2013.

W. R. Gilks and C. Berzuini. Following a moving target-Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:127–146, 2001. doi: 10.1111/1467-9868.00280.

D. Günther, T. Marke, R. Essery, and U. Strasser. Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance. *Water Resources Research*, 55:2779–2800, 2019. doi: 10.1029/2018WR023403.

P. Hennig, M.A. Osborne, and H.P. Kersting. *Probabilistic Numerics*. Cambridge, 2022. doi: 10.1017/9781316681411.

E.T. Jaynes. *Probability Theory*. Cambridge, 2003. doi: 10.1017/CBO9780511790423.

K.J.H. Law and A.M. Stuart. Evaluating Data Assimilation Algorithms. *Monthly Weather Review*, 140(11):3757–3782, 2012. doi: 10.1175/MWR-D-11-00257.1.

Y. Liu, Y. Fang, and S. A. Margulis. Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year Landsat–MODIS era snow reanalysis dataset. *The Cryosphere*, 15:5261–5280, 2021. doi: 10.5194/tc-15-5261-2021.

S. A. Margulis, M. Girotto, G. Cortés, and M. Durand. A particle batch smoother approach to snow water equivalent estimation. *Journal of Hydrometeorology*, 16:1752–1772, 2015. doi: 10.1175/JHM-D-14-0177.1.

S. A. Margulis, G. Cortés, M. Girotto, and M. Durand. A landsat-era Sierra Nevada snow reanalysis (1985-2015. *Journal of Hydrometeorology*, 17:1203–1221, 2016. doi: 10.1175/JHM-D-15-0177.1.

F. Maussion, A. Butenko, N. Champollion, M. Dusch, J. Eis, K. Fourteau, P. Gregor, A. H. Jarosch, J. Landmann, F. Oesterle, B. Recinos, T. Rothenpieler, A. Vlug, C. T. Wild, and B. Marzeion. The open global glacier model (oggm) v1.1. *Geoscientific Model Development*, 12(3):909–931, 2019. doi: 10.5194/gmd-12-909-2019. URL https://gmd.copernicus.org/articles/12/909/2019/.

M. Mazzolini, K. Aalstad, E. Alonso-González, S. Westermann, and D. Treichler. Spatio-temporal snow data assimilation with the icesat-2 laser altimeter. *The Cryosphere*, 19:3831–3848, 2025. doi: 10.5194/tc-19-3831-2025.

M. Morzfeld, D. Hodyss, and C. Snyder. What the collapse of the ensemble Kalman filter tells us about particle filters. *Tellus A: Dynamic Meteorology and Oceanography*, 69:1–14, 2017. doi: 10.1080/16000870.2017.1283809.

K.P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT, 2023. URL https://probml.github.io/book2. (last access: 18.12.2023).

R.M. Neal. Annealed importance sampling. *Statistics and Computing*, 11:125–139, 2001. doi: 10.1023/A:1008923215028.

Moritz Oberrauch, Bertrand Cluzet, Jan Magnusson, and Tobias Jonas. Improving fully distributed snowpack simulations by mapping perturbations of meteorological forcings inferred from particle filter assimilation of snow monitoring data. *Water Resources Research*, 60(12):e2023WR036994, 2024. doi: 10.1029/2023WR036994.

N. Pirk, K. Aalstad, S. Westermann, A. Vatne, A. van Hove, L.M. Tallaksen, M. Cassiani, and G. Katul. Inferring surface energy fluxes using drone data assimilation in large eddy simulations. *Atmospheric Measurement Techniques*, 15:7293–7314, 2022. doi: 10.5194/amt-15-7293-2022.

Norbert Pirk, Kristoffer Aalstad, Erik Schytt Mannerfelt, François Clayer, Heleen de Wit, Casper T. Christiansen, Inge Althuizen, Hanna Lee, and Sebastian Westermann. Disaggregating the carbon exchange of degrading permafrost peatlands using bayesian deep learning. *Geophysical Research Letters*, 51(10):e2024GL109283, 2024. doi: https://doi.org/10.1029/2024GL109283. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL109283. e2024GL109283 2024GL109283.

S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge, 2015. doi: 10.1017/CBO9781107706804.

J. Revuelto, E. Alonso-Gonzalez, I. Vidaller-Gayan, E. Lacroix, E. Izagirre, G. Rodríguez-López, and J.I. López-Moreno. Intercomparison of UAV platforms for mapping snow depth distribution in complex alpine terrain. *Cold Regions Science and Technology*, 190:103344, 2021. doi: 10.1016/j.coldregions.2021.103344.

C. Robert. *The Bayesian Choice*. Springer, 2 edition, 2007. doi: 10.1007/0-387-71599-1.

David R. Rounce, Regine Hock, Fabien Maussion, Romain Hugonnet, William Kochtitzky, Matthias Huss, Etienne Berthier, Douglas Brinkerhoff, Loris Compagno, Luke Copland, Daniel Farinotti, Brian Menounos, and Robert W. McNabb. Global glacier change in the 21st century: Every increase in temperature matters. *Science*, 379(6627):78–83, 2023. doi: 10.1126/science.abo1324.

S. Särkkä and L. Svensson. *Bayesian Filtering and Smoothing*. Cambridge, 2 edition, 2023. doi: 10.1017/9781108917407.

Chris Snyder, Thomas Bengtsson, Peter Bickel, and Jeff Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629 – 4640, 2008. doi: 10.1175/2008MWR2529.1.

A. S. Stordal and A. H. Elsheikh. Iterative ensemble smoothers in the annealed importance sampling framework. *Advances in Water Resources*, 86:231–239, 2015. doi: 10.1016/j.advwatres.2015.09.030.

A. van Hove, K. Aalstad, V. Lind, C. Arndt, V. Odongo, R. Ceriani, F. Fava, J. Hulth, and N. Pirk. Inferring methane emissions from african livestock by fusing drone, tower, and satellite data. *Biogeosciences*, 22(16):4163–4186, 2025. doi: 10.5194/bg-22-4163-2025.

Alouette van Hove, Kristoffer Aalstad, and Norbert Pirk. Actively inferring methane sources with drones. *Environmental Data Science*, 5:e2, 2026. doi: 10.1017/eds.2026.10029.

P. J. van Leeuwen, H. R. Künsch, L. Nerger, R. Potthast, and S. Reich. Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, 145:2335–2365, 2019. doi: 10.1002/qj.3551.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, Ilhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm,

G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, February 2020. ISSN 1548-7105. doi: 10.1038/s41592-019-0686-2. URL http://dx.doi.org/10.1038/s41592-019-0686-2.

S. Westermann, T. Ingeman-Nielsen, J. Scheer, K. Aalstad, J. Aga, N. Chaudhary, B. Etzelmüller, S. Filhol, A. Kääb, C. Renette, L. S. Schmidt, T. V. Schuler, R. B. Zweigel, L. Martin, S. Morard, M. Ben-Asher, M. Angelopoulos, J. Boike, B. Groenke, F. Miesner, J. Nitzbon, P. Overduin, S. M. Stuenzi, and M. Langer. The cryogrid community model (version 1.0) – a multi-physics toolbox for climate-driven simulations in the terrestrial cryosphere. *Geosci. Model Dev.*, 16(9):2607–2647, 2023. doi: 10.5194/gmd-16-2607-2023. URL https://gmd.copernicus.org/articles/16/2607/2023/.

C. Willmes, K. Aalstad, and S. Westermann. Assimilating high-resolution satellite snow cover data in a permafrost model. *EGUsphere*, 2025. doi: 10.5194/egusphere-2025-3142.

R. Yang, L. Ultee, K. Aalstad, M. Debolsky, R. Hock, P. Schmitt, and T. Li. Calibrating Coupled Frontal Ablation and Surface Mass Balance in Global Glacier Models Using an Adaptive Particle Smoother. *In prep*, 2025.

Lina Zschenderlein, Kari Luojus, Matias Takala, Pinja Venäläinen, and Jouni Pulliainen. Evaluation of passive microwave dry snow detection algorithms and application to swe retrieval during seasonal snow accumulation. *Remote Sensing of Environment*, 288:113476, 2023. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2023.113476. URL https://www.sciencedirect.com/science/article/pii/S0034425723000275.