

Evolving beyond collapse: An adaptive particle batch smoother for cryospheric data assimilation

Kristoffer Aalstad^{1*}, Esteban Alonso-González^{2*}, Norbert Pirk¹, Sebastian Westermann¹,
Clarissa Willmes¹, and Ruitang Yang¹

¹Department of Geosciences, University of Oslo (UiO), Oslo, Norway

²Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Jaca, Spain

*These authors contributed equally to this work.

Correspondence: Esteban Alonso-González (alonsoe@ipe.csic.es)

Abstract. We present a new adaptive particle-based data assimilation scheme for cryospheric applications that leverages promising developments in importance sampling. The proposed approach seeks to combine some of the advantages of two widely used classes of schemes: particle methods and iterative ensemble Kalman methods. Specifically, it extends the Particle Batch Smoother (PBS) that is commonly used in cryospheric data assimilation, with the Adaptive Multiple Importance Sampling algorithm. This adaptive formulation transforms the PBS into an iterative scheme with improved resilience against ensemble collapse and the ability to implement early-stopping strategies. As such, computational cost is automatically adapted to the complexity of the problem at hand, even down to the grid-cell and water year level in distributed multiyear simulations.

In homage to the schemes that it builds on, we coin this new algorithm the Adaptive Particle Batch Smoother (AdaPBS) and we test it across a range of scenarios. First, we conducted an intercomparison of some of the most commonly used cryospheric data assimilation algorithms using Markov Chain Monte Carlo (MCMC) simulation as a costly gold-standard benchmark in a simplified temperature index model assimilating snow depth observations. We further evaluated AdaPBS by assimilating snow depth observations from the ESMSnowMIP project at 6 different sites spanning 3 continents, using an ensemble of simulations generated with the more complex Flexible Snow Model (FSM2). Our results demonstrate that AdaPBS is a robust and reliable tool, outperforming or at least matching the performance of other commonly used algorithms and successfully handling complex cases with dense observational datasets. All experiments were carried out using the open-source Multiple Snow Data Assimilation System (MuSA) toolbox, which now includes AdaPBS and MCMC among the growing list of available cryospheric data assimilation methods. Beyond our cryospheric focus, the scheme has the potential to be applied directly to the closely related fields of land surface and hydrological data assimilation as well as more general geoscientific Bayesian inference problems.

1 Introduction

Billions of people dwell downstream of cryosphere-dominated basins that provide seasonal snowmelt and glacier meltwater as vital freshwater resources (Barnett et al., 2005; Immerzeel et al., 2020). Seasonal snow and glaciers in these cold regions, i.e. mountains and/or high latitudes, constitute a key natural water storage system (Gascoin, 2024), providing vital water resources

during spring and summer for drinking water, agriculture, hydropower, and ecosystems. The cryosphere provides additional
25 climate services as our planet's air conditioner (Euskirchen et al., 2013), for example by modulating the global energy cycle
(Riihelä et al., 2021), preserving considerable amounts of organic carbon in permafrost as opposed to the atmosphere (Pirk
et al., 2024), and storing frozen water in glaciers rather than raising sea levels (Rounce et al., 2023). These services are threat-
ened by ongoing anthropogenic global warming (Gottlieb and Mankin, 2024), which is amplified in cold regions leading to
strong perturbations of the terrestrial cryosphere according to the vast literature summarized in reports by the Intergovernmental
30 Panel on Climate Change (e.g. Hock et al., 2019; Meredith et al., 2019). Thus, the development and implementation of im-
proved cryospheric monitoring systems constitute a priority for humanity with numerous downstream scientific and operational
applications.

Due to the harsh conditions and heterogeneity of the cold regions where the cryosphere manifests, it is usually challenging
to deploy representative ground-based monitoring networks. In this context, remotely sensed retrievals of surface properties
35 have emerged as a partial solution to help monitor the state of the cryosphere at regional to global scales (Gascoin et al., 2024).
Unfortunately, the quest for direct, accurate, and gap-free estimates of key cryospheric states, such as snow mass (Dozier
et al., 2016), is a recalcitrant problem. Remotely sensed information that is retrievable from space is often limited to surface
processes that are only indirectly related to the full internal state of the cryosphere, further complicating its study as a dynamic
fresh water reservoir. As mentioned, space-borne estimates of snow mass, also known as snow water equivalent (SWE), in
40 complex terrain remain elusive leading to a big gap between what satellites observe and what hydrologists need (Dozier et al.,
2016). 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). This gap obfuscates the
important task of connecting snowpack storage and fluxes in complex terrain with downstream hydrology and land surface
models (De Lannoy et al., 2024). At the same time, mechanistic numerical modeling has become a powerful tool to simulate
45 the evolution of the snowpack (Essery et al., 2025) and the other main components of the terrestrial cryosphere in the form of
glaciers (Rounce et al., 2023) and permafrost (Westermann et al., 2023). Nonetheless, numerical models rely on the availability
of accurate high resolution meteorological forcings (Günther et al., 2019). This is currently not available on the global scale
where state-of-the-art products (Hersbach et al., 2020) do not resolve key processes or even major terrain features. In addition,
all physically-based numerical models rely at least in part on empirical parameterizations with parameters that are generally
50 uncertain and not transferable (Krinner et al., 2018).

The shortcomings of observations and models can be greatly minimized by leveraging data assimilation (DA) techniques,
with satellite-based cryospheric DA emerging as a particularly promising approach (Largerion et al., 2020; Giroto et al., 2020;
Alonso-González et al., 2022; Willmes et al., 2025; Yang et al., 2025). Using cryospheric DA techniques it is possible to per-
form uncertainty-aware monitoring of ungauged areas by simultaneously constraining the uncertainty in satellite observations
55 and simulations, to get the best of both worlds. Data assimilation is a promising way to infer uncertain parameters and update
model states (Reich and Cotter, 2015; Evensen et al., 2022), with many applications in the development of reanalysis (Margulis
et al., 2016; Hersbach et al., 2020; Sun et al., 2025) and the implementation of operational forecasting systems (Magnusson
et al., 2017; Carrassi et al., 2018; van Leeuwen et al., 2019). Moreover, data assimilation allows us to leverage global satellite

data provided by space agencies (Aalstad et al., 2018; Alonso-González et al., 2018), making the most of the long-term information from the existing climate data record while also offering the potential to ingest emerging satellite retrievals (Cluzet et al., 2024; Mazzolini et al., 2025) in a physically consistent way.

Various ensemble-based (also known as Monte Carlo) cryospheric DA algorithms have been proposed for this purpose (see e.g. Alonso-González et al., 2022). Two main families of ensemble-based schemes have been deployed to date in cryospheric DA: ensemble Kalman methods and particle methods. Of late, the latter particle-based methods have become especially popular in cryospheric applications (Leisenring and Moradkhani, 2011; Margulis et al., 2015; Charrois et al., 2016; Navari et al., 2016; Magnusson et al., 2017; Cortés and Margulis, 2017; Piazzzi et al., 2018; Fiddes et al., 2019; Smyth et al., 2019; Liu et al., 2021; Alonso-González et al., 2021; Cluzet et al., 2021; Landmann et al., 2021; Giroto et al., 2024; Oberrauch et al., 2024; Sun et al., 2025; Cao et al., 2025). This is due to their relatively simple implementation when used in their most basic ‘bootstrap’ form (Gordon et al., 1993; Särkkä and Svensson, 2023), their ease of interpretation, and the few assumptions they require (Largerón et al., 2020). However, issues with their practical implementation can make these methods problematic in certain settings. It is well known that if the proposal distribution, typically the prior, differs significantly from the target posterior distribution (MacKay, 2003; Rainforth et al., 2020), then all probability mass risks collapsing to a few or even a single particle, a problem known as particle degeneracy or ensemble collapse (Snyder et al., 2008; Morzfeld et al., 2017; Murphy, 2023). Degeneracy results in suboptimal approximate Bayesian inference by greatly degrading uncertainty quantification in the posterior simulations. This issue is aggravated by incorporating highly informative (i.e., numerous and/or very precise) observations, as well as when working in higher dimensional state and parameter spaces (van Leeuwen et al., 2019). Drawing on MacKay (2003), a useful analogy is looking for a needle (target posterior) in a haystack (prior proposal): a larger haystack (broader, higher-dimensionality) or a smaller needle (bigger data, more informative observations) exponentially reduces the chance of hitting the target needle when probing the haystack. While this needle in a haystack effect is a general problem, particle methods are especially vulnerable to it due to high Monte Carlo variance induced by the mismatch between the proposal and target (MacKay, 2003; Rainforth et al., 2020), challenges with localization (Farchi and Bocquet, 2018), and generally poor scaling with high effective dimensions (Snyder et al., 2008; Morzfeld et al., 2017).

Ensemble Kalman methods have demonstrated high resistance to collapse, even in high-dimensional systems. However, this is achieved by relying on the strong underlying assumption (although not a strict requirement) that all distributions involved in the analysis are Gaussian and that both the dynamical and observational models are linear for optimal inference (Evensen et al., 2022). These conditions are typically not met in cryospheric DA problems in particular (Largerón et al., 2020) or more generally in geosciences (Carrassi et al., 2018), resulting in these methods often being (arguably prematurely) discarded out of hand in many settings. To mitigate the linear assumption, the use of multiple data assimilation (MDA) iterations that allows a more progressive transition from prior to posterior has emerged as a promising enhancement of ensemble Kalman methods (Emerick and Reynolds, 2013; Aalstad et al., 2018; Alonso-González et al., 2022; Groenke et al., 2023). Moreover, this relaxation of the linear assumption via iteration also collaterally permits workarounds to soften the Gaussian assumption. In particular, transformation techniques (Gelman et al., 2013), such as Gaussian anamorphosis (Bertino et al., 2003; Carrassi

et al., 2018; Aalstad et al., 2018), allow the ensemble Kalman update to occur in a transformed Gaussian space rather than in the possibly bounded model space at the cost of an additional non-linear transform.

95 Previous work has demonstrated the potential of MDA-based iterative ensemble Kalman methods by showing that they can outperform or at least match other computationally tractable Monte Carlo algorithms in various comparisons (Aalstad et al., 2018; Alonso-González et al., 2022; Pirk et al., 2022; Keetz et al., 2025). The efficacy of this iterative ensemble Kalman approach has also been demonstrated in higher dimensional spatio-temporal cryospheric DA problems (Alonso-González et al., 2023; Mazzolini et al., 2025; Alonso-González et al., 2026) and in other complex non-linear and high dimensional inference
100 problems such as gradient-free training of deep neural networks (Pirk et al., 2024). Despite the clear benefits of iterative ensemble Kalman methods, there are some issues that need to be considered. Although the underlying linear and Gaussian assumptions can be strongly relaxed particularly in the limit of a large number of MDA iterations, this does not mean that they do not hold the potential of affecting the results without incurring a considerable computational cost. This makes the number of iterations an important hyperparameter, which has to be chosen with caution depending on the complexity of the problem.
105 In addition, in the seminal MDA approach with fixed observation error inflation (Emerick and Reynolds, 2013), the number of MDA iterations must be chosen a priori and it is necessary to perform them all to avoid violating the consistency of Bayesian inference (Stordal and Elsheikh, 2015; Alonso-González et al., 2022; Murphy, 2023), which complicates the implementation of early stopping strategies. Recent iterative ensemble Kalman schemes can circumvent the need to fix the number of iterations (Garbuno-Inigo et al., 2020; Groenke et al., 2023), although the adaptive pseudo-timestep may still require a relatively large
110 number of iterations to converge. Furthermore, with a large number of observations (in the order of thousands), owing to high spatio-temporal density and/or large DA windows, ensemble Kalman-based methods can be more costly than particle-based methods due to the large linear algebra operations involved in the computation of the Kalman gain in the analysis (Evensen et al., 2022).

In this paper, we explore the potential to overcome the shortcomings of particle-based methods through the use of iterations.
115 In doing so, we have been inspired by combining ideas from several established algorithms, namely the aforementioned PBS (Margulis et al., 2015) and iterative ensemble Kalman methods (Emerick and Reynolds, 2013; Stordal and Elsheikh, 2015), to build a new cryospheric data assimilation method based on developments in adaptive importance sampling (Cornuet et al., 2012; Bugallo et al., 2017). This new iterative and adaptive particle-based method that we coin the adaptive PBS (AdaPBS) has the potential to evolve beyond collapse unlike traditional particle-based methods, while making use of fewer assumptions
120 than ensemble Kalman-based methods and allowing the implementation of early stopping strategies that save substantial computational cost. In concurrent studies, we also demonstrate successful applications of AdaPBS to challenging cryospheric data assimilation problems related to glacier (Yang et al., 2025) and permafrost (Willmes et al., 2025) modeling. Our contribution here is devoted to describing this new scheme in detail and benchmarking it against existing schemes in several snow data assimilation experiments. By performing these experiments in the open source MuSA snow data assimilation toolbox (Alonso-González et al., 2022), a working AdaPBS Python code implementation is made available to the cryospheric community to
125 freely use and remix (Alonso-González and Aalstad, 2025). In the following, we will outline the relevant theory of Bayesian

cryospheric data assimilation in the context of this new adaptive particle method. Subsequently, we perform algorithm benchmarks in different scenarios of varying difficulty with different models to demonstrate the potential of the algorithm.

2 Cryospheric data assimilation

130 2.1 Bayesian inference

Data assimilation, loosely the fusion of data and models, can be formalized as the application of Bayesian inference (Wikle and Berliner, 2007). As such, we begin the section by briefly reviewing the key ideas behind Bayesian inference that is at the core of most modern DA schemes, including those explored herein. We refer the reader to the comprehensive texts of Gelman et al. (2013), Särkkä and Svensson (2023), Murphy (2023) and Evensen et al. (2022) for a thorough Bayesian treatment from the
135 perspectives of statistics, applied mathematics, machine learning and geophysical DA, respectively. A more thorough analysis of the connection between formal Bayesian inference and practical cryospheric DA is provided in Alonso-González et al. (2022).

In essence, Bayesian inference can be viewed as updating beliefs about some uncertain (also known as random) variables θ given some data $\mathbf{y} \in \mathbb{R}^{N_o}$. The uncertain variables in the vector θ can generally be of any form. Here, without loss of generality,
140 we will restrict our attention to continuous model parameters $\theta \in \mathbb{R}^{N_p}$. Belief updating can then be achieved by combining the basic sum and product rules of probability (Jaynes, 2003; MacKay, 2003) to obtain Bayes' rule

$$p(\theta | \mathbf{y}) = \frac{p(\mathbf{y} | \theta)p(\theta)}{p(\mathbf{y})}, \quad (1)$$

which states that the posterior belief after conditioning on the data, $p(\theta | \mathbf{y})$, is proportional to the product of the likelihood, $p(\mathbf{y} | \theta)$, which is loosely speaking what the data tell us, and the prior, $p(\theta)$, which encodes what we believed about θ before
145 considering the data. For the model evidence term $p(\mathbf{y})$ in the denominator of Eq. (1) we again combine the product and sum rules to see that

$$p(\mathbf{y}) = \int p(\theta, \mathbf{y}) d\theta = \int p(\mathbf{y} | \theta)p(\theta) d\theta, \quad (2)$$

where here and throughout this study the limits of integration are implicitly over the entire support of the integrand. From (2) the evidence is independent of the parameters θ and simply the integral of the numerator of Eq. (1) which ensures that the
150 posterior integrates to one over its support. Thus, the evidence can be seen as just a normalizing constant, although it plays an important role as a marginal likelihood for higher levels of inference, since it is implicitly conditioned on the model (MacKay, 2003; Murphy, 2023). In this Bayesian framework, the probability densities $p(\cdot)$ encode beliefs (i.e. epistemic uncertainties) about their arguments. If the subjective nature of beliefs is unpalatable, it may help to imagine that these are the beliefs held by an abstract numerical agent with a probabilistic model of the world (Hennig et al., 2022). This probabilistic numerics
155 perspective is instructive for the adaptive methods that we will present here.

In theory, by inspection of Eq. (1), performing Bayesian inference is simply (up to a normalizing constant) multiplying the prior and the likelihood. Naively then, we could just enumerate the posterior through an exhaustive grid approximation

provided that we select a sufficiently fine discretization of the variables θ (MacKay, 2003). However, as soon as we have more than a couple of uncertain variables in θ and/or very informative data in $\mathbf{y}$, this grid approach becomes intractable due to the so-called curse of dimensionality (Murphy, 2023). Recalling and extending the earlier analogy, the task of inferring a possibly multi-modal posterior distribution $p(\theta | \mathbf{y})$ is akin to looking for an unknown number of needles in a multi-dimensional haystack: the larger the support (e.g. dimensionality) of the prior, the larger the haystack, and the more informative (numerous and/or accurate) the observations, the smaller the needles. It is this computational challenge that has spurred the development of sophisticated Bayesian inference algorithms, including ensemble-based (also known as Monte Carlo) algorithms that emerged from statistical physics (MacKay, 2003; Robert and Casella, 2004) and have, along with variational methods, enjoyed widespread adoption in geophysical DA (Reich and Cotter, 2015; Evensen et al., 2022). In this study, we restrict our attention to ensemble-based methods since these gradient-free methods are the most widely used in cryospheric DA due to their relative ease of use and comparatively robust uncertainty quantification (see Alonso-González et al., 2022, and references therein). At the same time, more research is warranted to explore the ever-growing plethora of inference algorithms (Evensen et al., 2022; Hennig et al., 2022; Murphy, 2023; Särkkä and Svensson, 2023), many of which remain relatively untested in cryospheric science and geoscience more generally.

2.2 Cryospheric inverse problems

To help concretize the above formalization of cryospheric DA, it is instructive to consider the perspective of inverse modeling (Sanz-Alonso et al., 2023). Let $\mathcal{G}(\cdot)$ denote our forward (or data generating) model that maps from the uncertain snow model parameters to predicted (modeled) snow observations $\hat{\mathbf{y}} \in \mathbb{R}^{N_o}$, i.e., $\hat{\mathbf{y}} = \mathcal{G}(\theta)$. In practice, the forward model combines an observation model for $\hat{\mathbf{y}} = \mathcal{H}(\mathbf{x})$ with a dynamical model for the full model state trajectory $\mathbf{x} = \mathcal{M}(\phi)$ and a (bounding) parameter transformation step $\phi = \mathcal{T}(\theta)$ (Gelman et al., 2013; Alonso-González et al., 2022; Pirk et al., 2022). Next, consider the typical case where we are given a set of noisy snow observations $\mathbf{y}$ which we assume to be related to some true snow observable $\mathbf{y}^*$ through $\mathbf{y} = \mathbf{y}^* + \epsilon$, where ϵ is the observation error. By making the usual strong constraint (or perfect model) assumption (Evensen et al., 2022) that the forward model maps perfectly (without error) from the parameters to the observed variables, we can define some true parameter set θ^* to exist such that $\mathbf{y}^* = \mathcal{G}(\theta^*)$. 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).

Using the definition of the observation error, we can establish the following forward relationship between the noisy observations we are given and some presumed true parameters of interest

$$\mathbf{y} = \mathcal{G}(\theta^*) + \epsilon. \quad (3)$$

At least conceptually, the task in cryospheric data assimilation can now be cast as somehow *inverting* $\mathcal{G}(\cdot)$ in (3) to solve for θ^* . Unfortunately, even in an idealized linear and noise-free ($\epsilon = 0$) case, this is typically an ill-posed problem in that an

exact solution θ^* may not exist or be unique (Sanz-Alonso et al., 2023). In the more challenging noisy and possibly non-linear practical settings the problem is always ill-posed since we invariably assimilate noisy observations where the observation error is *uncertain*.

Due to ill-posedness, the quest for a universally optimal (let alone exact) solution θ^* is nonsensical since infinitely many solutions θ are typically admissible. A way forwards is to instead seek a probabilistic (i.e. Bayesian) solution to the inverse problem in (3) where we construct an observation error model by treating ϵ as an uncertain variable. For convenience, as is common practice in DA (Carrassi et al., 2018), we assume independent additive zero-mean Gaussian observation errors $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ where $\mathbf{R}$ is an $N_o \times N_o$ diagonal observation error covariance matrix. This assumption helps formulate the likelihood $p(\mathbf{y} | \theta)$, the probability of obtaining the fixed observations $\mathbf{y}$ given that a parameter set θ is true, since if $\theta = \theta^*$ (due to the conditional $| \theta$) then from (3) $\epsilon = \mathbf{y} - \mathcal{G}(\theta)$ is the residual so we apply the assumed observation error model to obtain a Gaussian likelihood (Evensen et al., 2022)

$$p(\mathbf{y} | \theta) = c_y \exp \left(-\frac{1}{2} [\mathbf{y} - \hat{\mathbf{y}}]^T \mathbf{R}^{-1} [\mathbf{y} - \hat{\mathbf{y}}] \right), \quad (4)$$

where $c_y = \det(2\pi\mathbf{R})^{-1/2}$ and $\hat{\mathbf{y}} = \mathcal{G}(\theta)$ denotes the predicted (i.e. modeled) observations given a particular parameter set θ whereby diagnosing the likelihood in (4) requires point-wise evaluations of the forward model to evaluate the residual. In accordance with the likelihood principle, $p(\mathbf{y} | \theta)$ should be viewed as a function of the uncertain parameters θ rather than the fixed (albeit noisy) observations $\mathbf{y}$ that we are assimilating (MacKay, 2003). If we now combine this likelihood with a regularizing prior $p(\theta)$ that encodes initial beliefs concerning the parameters, then in principle the full probabilistic solution to the inverse problem (3) is obtained by inferring the posterior through Bayes' rule in (1). As noted by Evensen et al. (2022) this process of Bayesian inference is just point-wise multiplication that does not in itself involve any explicit (matrix or function) inversion, but it nonetheless offers a probabilistic framework for solving general geophysical inverse problems (Sanz-Alonso et al., 2023). Although distinctions are sometimes made between DA and inverse modeling, the two fields are highly complementary and unified under the umbrella of Bayesian inference (Reich and Cotter, 2015; Evensen et al., 2022; Sanz-Alonso et al., 2023). Thereby, the process of cryospheric DA can be viewed as the solution to dynamical (time-varying) cryospheric inverse problems. The Bayesian dynamics then determines if one is solving a filtering or more general smoothing problem (Alonso-González et al., 2022; Särkkä and Svensson, 2023), either way the solutions are typically obtained via numerical approximation of Bayesian inference which is our focus.

2.3 Markov chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods, which are asymptotically exact, are widely considered the reference computational tool for approximating Bayesian posteriors in practice (Gelman et al., 2013; Murphy, 2023). Here we provide a brief overview of MCMC since we will use it as a gold-standard computational benchmark against which to gauge the performance of other cryospheric DA methods (Law and Stuart, 2012). As expounded in Robert and Casella (2004), MCMC originated with the seminal physics paper of Metropolis et al. (1953), whose results were later generalized to statistics by Hastings (1970), introducing the Random Walk Metropolis (RWM) algorithm that is the ancestor of modern MCMC methods, among which

gradient-based methods such as Hamiltonian Monte Carlo (MacKay, 2003; Neal, 2011) are arguably the state-of-the-art. All
 225 MCMC schemes construct Markov chains so as to asymptotically sample from a target distribution of interest, where the natural
 choice for Bayesian DA is the posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$. More specifically, MCMC algorithms take sequential Markovian (memory-
 less) steps through the parameter space and probabilistically either accept or reject the newly proposed step by comparing its
 posterior density to that of the current step, eventually obtaining samples from the posterior distribution in (1).

We base our implementation here on the aforementioned RWM algorithm which remains arguably the most archetypal
 230 MCMC method. In this approach, a new step $\boldsymbol{\theta}_{i+1}^*$ in the chain is proposed by randomly drawing from a proposal distribution
 $q(\boldsymbol{\theta}_{i+1}^* \mid \boldsymbol{\theta}_i)$ that is conditioned (usually by centering) on the current step $\boldsymbol{\theta}_i$. This proposed step is probabilistically accepted
 based on the value of the acceptance ratio

$$\alpha_{i+1} = \frac{p(\mathbf{y} \mid \boldsymbol{\theta}_{i+1}^*)p(\boldsymbol{\theta}_{i+1}^*)}{p(\mathbf{y} \mid \boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)}, \quad (5)$$

on the condition that the acceptance rule $u_{i+1} \leq \min(\alpha_{i+1}, 1)$ holds, where $u_{i+1} \sim \text{U}(0, 1)$ is a realization of a random variable
 235 that is uniformly distributed between 0 and 1, otherwise it is rejected. The evidence $p(\mathbf{y})$, which appears as a constant in the
 posterior density (1) for both the current and the proposed step, cancels out in the acceptance ratio, so we do not need to
 estimate this intractable quantity. This, together with asymptotic guarantees, helps explain the historical popularity of MCMC
 sampling. The more general form of Eq. (5) introduced by Hastings (1970) also involves proposal densities, but these cancel
 out in RWM and were hence omitted here. Upon acceptance, the chain moves to the proposed point in parameter space such
 240 that $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_{i+1}^*$. Upon rejection, the chain stays at the current point $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$. Note that a proposed point will always be
 accepted if it has a higher posterior density since then $\alpha_{i+1} > 1$ so the acceptance rule always holds. Moreover, the proposed
 point $\boldsymbol{\theta}_{i+1}^*$ will always have a non-zero probability of being accepted. This probabilistic acceptance rule helps ensure that the
 chain will asymptotically ($i \rightarrow \infty$) sample from the posterior. To sample from the posterior with MCMC, one needs to run the
 Markov chain for enough iterations to ensure that it mixes properly and converges to sampling from the posterior. In practice,
 245 although there are some diagnostics that can be used as a guide (Gelman et al., 2013), what constitutes enough iterations is
 uncertain *a priori* and can remain challenging to verify *post hoc*. As such, MCMC is typically run for a very large number (i.e.,
 tens of thousands) of iterations to ensure convergence (Cleary et al., 2021). An initial part of the chain is discarded as a burn-in
 period to avoid the biasing effects of the chain initialization. Due to the Markov property, where the next step only depends on
 the current step, it is also clear that the subsequent iterations in the chain will be auto-correlated rather than independent. As
 250 such, the remaining part of the chain can be subsampled uniformly at random to obtain more independent samples from the
 posterior distribution.

The most standard RWM approach employs a multivariate normal (Gaussian) proposal of the form $q(\boldsymbol{\phi}_{j+1}^* \mid \boldsymbol{\phi}_j) = \text{N}(\boldsymbol{\phi}_{j+1}^* \mid$
 $\boldsymbol{\phi}_j, \boldsymbol{\Sigma}_q)$ where the mean is the current point $\boldsymbol{\phi}_j$ and $\boldsymbol{\Sigma}_q$ is the proposal covariance matrix. To reduce the number of tuning
 parameters in the proposal, the latter can be made isotropic by using a scalar (constant diagonal) matrix of the form $\boldsymbol{\Sigma}_q = \sigma_q^2 \mathbf{I}$
 255 where σ_q^2 is the proposal variance and $\mathbf{I}$ is an identity matrix. In practice, good mixing of the RWM method requires judicious
 hand-tuning of σ_q , and even then the use of an isotropic covariance matrix remains wasteful to efficiently explore higher
 dimensional parameter spaces given that the posterior is often anisotropic (MacKay, 2003). A relatively simple workaround is

to employ an adaptive RWM method that automatically modifies a generally anisotropic proposal covariance on the fly using the history of the evolving Markov chain to obtain better mixing properties. Among the existing adaptive MCMC algorithms, here we chose to employ the robust adaptive Metropolis (RAM) method proposed by Vihola (2012) using the hyperparameters suggested therein with $N_s = 20 \times 10^3$ steps and discarding the first 10% as a burn-in phase (Murphy, 2023). The RAM method was chosen in-lieu of even more sophisticated MCMC methods such as Hamiltonian Monte Carlo (Neal, 2011) that may have mixed even faster since the RAM method is considerably easier to implement. Moreover, by running RAM for tens of thousands of iterations we can be fairly confident that it converges and represents a strong benchmark. A crucial point here is that it is likely that most if not all MCMC methods are too computationally expensive to deploy at scale (i.e., for large areas) in practical cryospheric DA, although further research into applying more sophisticated MCMC schemes (Murphy, 2023) is needed. Herein, MCMC in the form of RAM is used as a reference standard against which we gauge the performance of the more tractable ensemble-based cryospheric DA schemes (Law and Stuart, 2012), especially the adaptive particle smoother that is the focus of this study.

2.4 Ensemble Kalman methods

Ensemble Kalman methods, introduced by Evensen (1994) and described in detail in Evensen et al. (2022), are ensemble-based extensions of the classic Kalman methods (Jazwinski, 1970) which provide exact Bayesian inference methods for linear Gaussian models (Särkkä and Svensson, 2023). In such models the mapping from hidden states and/or parameters to observations is linear while both the prior and the likelihood are Gaussian. Ensemble Kalman methods help to relax these assumptions in the sense that approximate yet efficient inference is still possible when they are violated, making them applicable also to non-linear geophysical problems (Evensen et al., 2022). Although other non-linear variations on classic Kalman methods also exist (Särkkä and Svensson, 2023), the ease of implementation and the robust performance of ensemble Kalman methods have made them highly applicable for geophysical DA in general (Carrasi et al., 2018) as well as cryospheric DA in particular (see Alonso-González et al., 2022, and references therein). Crucially, the more recent development of iterative ensemble Kalman methods (Emerick and Reynolds, 2013; Garbuno-Inigo et al., 2020) have helped to enhance the capability of this class of DA schemes for highly non-linear and/or complex problems (e.g. Pirk et al., 2022, 2024; Keetz et al., 2025).

In practical cryospheric DA, several numerical experiments have previously shown that iterative ensemble Kalman methods can outperform basic particle methods while maintaining more robust posterior uncertainty quantification (Aalstad et al., 2018; Alonso-González et al., 2022). It is thus instructive to also include such experiments here to provide an additional benchmark for the performance of the proposed adaptive particle method. Rather than providing a high-fidelity benchmark like the MCMC experiments, the more computationally tractable iterative ensemble Kalman experiments should be seen as a more practical operational benchmark for cryospheric DA. For the sake of completeness, the rest of this section provides a brief overview on the implementation of iterative ensemble Kalman methods. We refer to Evensen et al. (2022) and Alonso-González et al. (2022) for more details on the theory of ensemble Kalman methods and their implementation for cryospheric DA, respectively.

Here we adopt a specific iterative ensemble Kalman method known as the ensemble smoother with multiple data assimilation (ES-MDA; Emerick and Reynolds, 2013) due to its relative ease of implementation and robust performance (Aalstad et al.,

2018; Alonso-González et al., 2022; Alonso-González et al., 2023). We note in passing that other promising variations on the iterative ensemble Kalman method exist (Garbuno-Inigo et al., 2020; Evensen et al., 2022) and are worthy of further investigation in cryospheric DA, but we do not expect their performance to differ markedly from the ES-MDA. The ES-MDA scheme is initialized by sampling an initial ensemble of $i = 1, \dots, N_e$ parameter vectors from the prior $\theta_i^{(0)} \sim p(\theta)$ and it then proceeds by cycling between a prediction and update step for $\ell = 0, \dots, N_a$ iterations:

1. Run the forward model to obtain the predicted observations $\hat{\mathbf{y}}_i^{(\ell)} = \mathcal{G}(\theta_i^{(\ell)})$ from the state $\mathbf{x}$ (see Section 2.2).
2. If $\ell < N_a$, perform a tempered ensemble Kalman update step

$$\theta_i^{(\ell+1)} = \theta_i^{(\ell)} + \mathbf{K}^{(\ell)} (\mathbf{y} - \hat{\mathbf{y}}_i^{(\ell)}), \quad (6)$$

300 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 $i = 1, \dots, N_e$ 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 . Recall from Section 2.2 that several steps (observation, dynamics, transformation) are baked into the forward model $\mathcal{G}(\cdot)$. Therein, it is 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. So, while (single chain) MCMC costs tens of thousands of strictly sequential forward model runs, with a typical setting of $N_a = 4$ (Aalstad et al., 2018; Pirk et al., 2022; Alonso-González et al., 2022) ES-MDA only incurs a computational cost of $N_a + 1 = 5$ iterations of an ensemble of $N_e = 100$ parallelizable forward model runs. With $N_a = 1$ the ES-MDA reverts to the original (non-iterative) ensemble smoother (ES) scheme proposed by van Leeuwen and Evensen (1996) which we also include here in the benchmarking of the new AdaPBS method. We emphasize that even this non-iterative ES with $N_a = 1$ has a cost of $(N_a + 1) \times N_e = 200$ model runs as it requires rerunning an ensemble of model simulations with the updated parameters to obtain posterior state predictions.

2.5 Particle methods

315 Particle methods (van Leeuwen, 2009; van Leeuwen et al., 2019), also known as sequential Monte Carlo (SMC) (Chopin and Papaspiliopoulos, 2020), rose to prominence with the work of Gordon et al. (1993) and Kitagawa (1996) around the same time as ensemble Kalman methods (Evensen, 1994). Moreover, similar inference methods have arguably independently been (re)discovered in geoscience (Beven and Binley, 1992; van Leeuwen and Evensen, 1996; Margulis et al., 2015). These particle methods also have roots back to the dawn of Monte Carlo methods in physics (Hammersley and Morton, 1954; Robert and Casella, 2004). The key inference mechanism that powers particle methods is importance sampling (MacKay, 2003), which weighs the importance of an ensemble of particles (ensemble members) θ_i according to their posterior probability density and the density of the proposal. By performing this sequentially in time and resampling the particles based on their weights (i.e.,

fitness) after each importance sampling step, we recover the algorithm known as sequential importance resampling (SIR; Smith and Gelfand, 1992; Doucet et al., 2000) at the core of particle methods used for Bayesian filtering and smoothing (Kitagawa, 1996) and static inference problems (Chopin, 2002). The cycling of prediction (mutation) and updating (selection) has clear connections with metaheuristic evolutionary methods (Campelo and Aranha, 2023) such as genetic algorithms (Holland, 1992) that can be mathematically formalized as particle methods (Del Moral, 2004). This perspective also provides a connection to other evolutionary-inspired Bayesian such as Shuffled Complex Evolution Metropolis (SCEM-UA; Vrugt et al., 2003) and Differential Evolution Adaptive Metropolis (DREAM; Vrugt et al., 2008), that remain state of the art optimizers and samplers, respectively, in hydrology. However, following Sørensen (2015) we emphasize that the adaptive particle method proposed herein and our playful title use evolutionary principles as a conceptual model for inspiration and not as a justification. Justification can be found in the literature on the convergence of general SMC methods (Chopin and Papaspiliopoulos, 2020) and more specifically adaptive multiple importance sampling (Marin et al., 2019). Having introduced the notion of particle methods, the rest of this section will outline the principle of importance sampling and how this is incorporated into SIR. Together, this basic SIR theory suffices to grasp vanilla particle methods related to the seminal ‘bootstrap’ particle filter of Gordon et al. (1993). These basic particle methods have become popular approaches to DA in snow science (e.g. Leisenring and Moradkhani, 2011; Margulis et al., 2015; Charrois et al., 2016; Magnusson et al., 2017; Piazzzi et al., 2018; Fiddes et al., 2019; Smyth et al., 2019; Alonso-González et al., 2021; Cluzet et al., 2021; Oberrauch et al., 2024; Giroto et al., 2024; Sun et al., 2025) and glaciology (Navari et al., 2016; Landmann et al., 2021; Cao et al., 2025).

2.5.1 Monte Carlo integration

Importance sampling is a generalized form of indirect Monte Carlo integration that can be used to estimate expectations with respect to a complex target distribution, in our case the posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$, by sampling from a simpler proposal distribution $q(\boldsymbol{\theta})$ (MacKay, 2003). To unpack this definition, it can be helpful to recall how basic direct Monte Carlo integration works in the context of estimating expectations. The posterior expectations we wish to estimate are of the form (Särkkä and Svensson, 2023)

$$\mathbb{E}[g(\boldsymbol{\theta}) \mid \mathbf{y}] = \int g(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}, \quad (7)$$

where we recall that the integral is implicitly over the support of the integrand. Here $g(\boldsymbol{\theta})$ is an arbitrary (possibly vector-valued) function of the parameters $\boldsymbol{\theta}$ that we may wish to take the posterior expectation of. For example, using $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$ in (7) yields the posterior mean $\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}$ while using $g(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}})^2$ in (7) yields the posterior variance $\hat{\boldsymbol{\sigma}}_{\boldsymbol{\theta}}^2$. Now suppose we could generate $i = 1, \dots, N_e$ independent samples from the posterior $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta} \mid \mathbf{y})$ then we could obtain a direct Monte Carlo estimate of (7) through the sample mean

$$\mathbb{E}[g(\boldsymbol{\theta}) \mid \mathbf{y}] \simeq \frac{1}{N_e} \sum_{i=1}^{N_e} g(\boldsymbol{\theta}_i), \quad (8)$$

which, thanks to the law of large numbers and the central limit theorem, will be an unbiased estimate that asymptotically ($N_e \rightarrow \infty$) converges to the true posterior expectation in (7) with a Monte Carlo error that decays at a rate $\mathcal{O}(N_e^{-1/2})$ (Chopin

355 and Papaspiliopoulos, 2020). Although this error reduction rate is relatively slow (Hennig et al., 2022), e.g. increasing the ensemble size by a factor 100 only reduces the error by a factor 10, it is independent of dimension which makes these Monte Carlo (also known as ensemble-based) methods a viable (and sometimes the sole) option for challenging inference problems that arise in geophysical DA (Carrassi et al., 2018). Nonetheless, following Hennig et al. (2022), it is worth remembering that Monte Carlo methods are a last resort in line with the principle of Jaynes (2003) that for every randomized method there is
 360 usually a better performing deterministic method that requires more thought.

2.5.2 Importance sampling

The obvious problem with this method is that we are not able to directly sample independently (let alone efficiently) from the exact posterior distribution. In fact, posterior sampling is often the very problem that we need to solve. Nonetheless, once we have obtained posterior samples we can use Monte Carlo integration to approximate the desired posterior expectations of
 365 interest. Slow MCMC methods (Section 2.3) are only asymptotically exact samplers that do not provide independent samples. The efficient yet approximate ensemble Kalman methods (Section 2.4) are only exact samplers for Gaussian linear models. This is where the more general and indirect Monte Carlo technique of importance sampling shines.

Defining a proposal distribution $q(\boldsymbol{\theta})$ that we can easily generate independent samples from and multiplying the integrand in (7) by $1 = q(\boldsymbol{\theta})/q(\boldsymbol{\theta})$ then clearly

$$370 \quad \mathbb{E}[g(\boldsymbol{\theta}) | \mathbf{y}] = \int g(\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta} | \mathbf{y})}{q(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (9)$$

under the sufficient condition that the support of the proposal encompasses that of the posterior, i.e. $q(\boldsymbol{\theta}) > 0$ wherever $p(\boldsymbol{\theta} | \mathbf{y}) > 0$ (Chopin and Papaspiliopoulos, 2020). If we now generate independent samples from the proposal $\boldsymbol{\theta}_i \sim q(\boldsymbol{\theta})$ we can obtain the following importance sampling estimate of (7) from (9)

$$\mathbb{E}[g(\boldsymbol{\theta}) | \mathbf{y}] \simeq \frac{1}{N_e} \sum_{i=1}^{N_e} g(\boldsymbol{\theta}_i) \frac{p(\boldsymbol{\theta}_i | \mathbf{y})}{q(\boldsymbol{\theta}_i)}, \quad (10)$$

375 where we have great freedom in the choice of the proposal $q(\boldsymbol{\theta})$ (van Leeuwen, 2009).

2.5.3 Self-normalized importance sampling

Unfortunately we still can not evaluate (10) since the posterior density $p(\boldsymbol{\theta}_i | \mathbf{y})$ term defined in Eq. (1) is only known up to an *unknown* normalizing constant, namely the evidence. The evidence term $p(\mathbf{y})$ in (2) is a typically intractable integral over the unnormalized posterior $p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})$. Nonetheless, comparing (2) and (7) the evidence can be seen as the prior (rather than
 380 posterior) expectation of the likelihood by setting $g(\boldsymbol{\theta}) = p(\mathbf{y} | \boldsymbol{\theta})$ and replacing the posterior with the prior in (7). Analogously to (9), we can then recast the evidence as an expectation involving the proposal density

$$p(\mathbf{y}) = \int p(\mathbf{y} | \boldsymbol{\theta}) \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (11)$$

so that we can use proposal samples $\theta_i \sim q(\theta)$ to obtain the estimate

$$p(\mathbf{y}) \simeq \frac{1}{N_e} \sum_{i=1}^{N_e} p(\mathbf{y} | \theta_i) \frac{p(\theta_i)}{q(\theta_i)} = \frac{1}{N_e} \sum_{i=1}^{N_e} \tilde{w}_i, \quad (12)$$

385 where we have defined the unnormalized weights $\tilde{w}_i = \tilde{w}(\theta_i) = p(\mathbf{y} | \theta_i)p(\theta_i)/q(\theta_i)$ as a useful shorthand. Now we are in a position to revisit the importance sampling estimate of the posterior expectation (10) which, using the definition of the posterior in (1), can be re-written as

$$\mathbb{E}[g(\theta) | \mathbf{y}] \simeq \frac{1}{N_e} \sum_{i=1}^{N_e} g(\theta_i) \frac{p(\mathbf{y} | \theta_i)p(\theta_i)}{p(\mathbf{y})q(\theta_i)}. \quad (13)$$

If we now use the definition of the unnormalized weights $\tilde{w}_i$ and insert for the approximation of $p(\mathbf{y})$ in (12), we obtain the
390 so-called self-normalized importance sampling (SNIS) estimate (Rainforth et al., 2020; Murphy, 2023)

$$\mathbb{E}[g(\theta) | \mathbf{y}] \simeq \frac{1}{N_e} \sum_{i=1}^{N_e} g(\theta_i) \frac{\tilde{w}_i}{\frac{1}{N_e} \sum_{k=1}^{N_e} \tilde{w}_k} = \sum_{i=1}^{N_e} g(\theta_i) w_i \quad (14)$$

where the normalized weights w_i are defined by $w_i = \tilde{w}_i / \sum_{k=1}^{N_e} \tilde{w}_k$ with the property that $\sum_{i=1}^{N_e} w_i = 1$. Although the additional approximation has downsides (Rainforth et al., 2020), this self-normalization step means that we can ignore normalizing constant terms in the proposal, prior, and likelihood since these will cancel upon normalization.

395 The SNIS estimate of the posterior expectation in (14) forms the basis of the majority of SIR-based particle methods applied to the geosciences (van Leeuwen et al., 2019). Moreover, SNIS is formally mathematically equivalent to using a particle approximation (Särkkä and Svensson, 2023) of the posterior in (7) of the form

$$p(\theta | \mathbf{y}) \simeq \sum_{i=1}^{N_e} w_i \delta(\theta - \theta_i), \quad (15)$$

with weights w_i given by (14) and $\theta_i \sim q(\theta)$. This is relatively trivial to verify by recalling the sifting property of the Dirac delta
400 function that $\int g(\theta) \delta(\theta - \theta_i) d\theta = g(\theta_i)$ (Murphy, 2023). On the one hand, this particle approximation helps conceptualize the posterior estimate in (15) as a sum of point particles whose relative importance is given by their weights. On the other hand, the SNIS formalism that we have outlined clarifies the origin of these weights.

2.5.4 Resampling

In settings such as Bayesian filtering in state space models (Gordon et al., 1993; Kitagawa, 1996; Gilks and Berzuini, 2001;
405 Särkkä and Svensson, 2023) or applying data tempering for static parameter inference with large amounts of data (Chopin, 2002; Murphy, 2023; van Hove et al., 2025) it can be helpful to embed importance sampling in Sequential Importance Resampling (SIR) that is the basis of particle filtering (Chopin and Papaspiliopoulos, 2020). The workflow in SIR is to sequentially assimilate data as it becomes available in time or through minibatches and then to perform a resampling step on the dynamic weights after each SNIS step. This exploits the computational benefits of the sequential nature of Bayesian inference, where

the posterior of the current step can become the prior for the next step (Särkkä and Svensson, 2023), while using resampling to avoid the inevitable weight degeneracy that would otherwise occur. Herein, our focus is on inferring static parameters via batch smoothing within a given water year so the sequential aspect of SIR is not as important. At the same time, the adaptive particle method presented herein could also be embedded within a sequential particle filtering framework. Moreover, the adaptation step that we use does make use of particle resampling so we also briefly outline what resampling entails.

In the resampling step, particles are resampled with replacement according to the probability mass given by their weights. As such, more fit high weight particles are copied whereas less fit low weight particles are removed. Many particle resampling methods exist (Li et al., 2015) and the particular method used is often of secondary importance as long as it is valid. After resampling, all particles are assigned an equal weight of $1/N_e$ and can be treated as independent samples from the target allowing for straightforward Monte Carlo integration to approximate expectations. The fact that resampling results in equal weights means that it avoids weight degeneracy since there will no longer be just a few particles carrying all the weight. Although resampling trivially solves the weight degeneracy problem, it does not solve the more general problem of path degeneracy wherein many identical resampled particles results in a poor posterior approximation (Murphy, 2023). A promising way to alleviate the path degeneracy problem is to rejuvenate the ensemble of particles through the resample-move algorithm by taking a few MCMC steps targeting the posterior to diversify the ensemble after resampling (Gilks and Berzuini, 2001; van Hove et al., 2025). Since we do not pursue filtering herein, we only mention this resample-move strategy in passing to motivate further work on particle filtering in cryospheric DA using the adaptive method we will present.

2.6 Adaptive particle methods

The typical particle methods used in cryospheric DA (see Alonso-González et al., 2022, and references therein) can be considered to be basic particle methods in the spirit of the seminal bootstrap particle filter (Gordon et al., 1993) in that they do not leverage recent developments (van Leeuwen et al., 2019; Chopin and Papaspiliopoulos, 2020). They are basic in the sense that they use the simplest and most convenient choice of proposal distribution (van Leeuwen, 2009), namely using the prior as the proposal. In this case, i.e. inserting $q(\theta_i) = p(\theta_i)$ in (12), the unnormalized weights simply correspond to likelihood evaluations $\tilde{w}_i = p(\mathbf{y} | \theta_i)$. Thereby, the normalized weights in the weighted mean approximation of the posterior expectation in (14) just become normalized likelihoods of the form $w_i = p(\mathbf{y} | \theta_i) / \sum_{k=1}^{N_e} p(\mathbf{y} | \theta_k)$. This choice of proposal also makes it easy to implement particle filtering via SIR since one does not need direct access to the dynamic prior which is generally no longer available in closed form. Nonetheless, in terms of the importance sampling step itself the use of the prior as the proposal is generally highly suboptimal. This has motivated more sophisticated particle methods such as those involving tempering (Chopin and Papaspiliopoulos, 2020; Murphy, 2023), hybrid methods (Pirk et al., 2022; Särkkä and Svensson, 2023), and adaptive importance sampling methods (Bugallo et al., 2017). Our focus will be on the latter, although we note in passing that it is entirely possible to combine these approaches (e.g. Koblents and Míguez, 2015).

The idea behind adaptive importance sampling is fairly simple: apply importance sampling iteratively and use the results of the previous iteration to adapt the proposal for the next. The reason that this approach tends to work is also clear: after each iteration the weighted particle ensemble tends to be a better approximation of the posterior than the proposal and so we can

use these particles to allow the proposals to evolve by adapting toward the posterior. As such, adaptive importance sampling leverages the power of iterations that iterative ensemble Kalman methods also exploit. At the same time, adaptive importance sampling adapts the proposal to the complexity of the target posterior at hand so that the cost of the algorithm is not set in stone. Thereby, these adaptive algorithms can converge rapidly via early stopping and only in the worst case will they run for a predefined maximum number of iterations. A plethora of adaptive importance sampling methods exist (Bugallo et al., 2017) and many of these are worthy of further investigation in cryospheric DA. The AMIS method that we chose to adopt was based on preliminary non-exhaustive testing of a subset of these methods and came out on top in terms of efficiency and performance.

2.6.1 Adaptive multiple importance sampling

Here we introduce the general adaptive multiple importance sampling (AMIS) algorithm proposed by Cornuet et al. (2012). Our particular adaptation of this method to cryospheric DA with the AdaPBS is described in the subsequent section. This AMIS approach is adaptive in a couple of ways. Firstly, the proposal distribution iteratively adapts to better approximate the target posterior distribution. Secondly, the algorithm has an (optional) convergence criterion that is reached if an adequately high effective sample size is obtained. This means that AMIS will revert to the behaviour of computationally cheaper non-iterative basic importance sampling as typically used in cryospheric DA in cases where the basic method already provides a good enough approximation of the posterior. Note that these two adaptive features are common among many adaptive importance sampling methods as outlined in Bugallo et al. (2017). What makes the AMIS method unique is that it employs a so-called deterministic mixture proposal (Owen and Zhou, 2000) which makes it stable, relatively quick to converge, and waste-free in that the entire history of particles is used at each iteration. Due to this latter waste-free property, the AMIS algorithm is somewhat involved and requires tracking this evolving particle history through several generations. As such, it is easiest to present AMIS through a series of 7 sequential steps that we cycle through iteratively. That is, we initialize the iteration counter $\ell = 1$ and set the current sampling proposal to the prior $q^{(1)}(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ and then proceed as follows:

1. Generate an ensemble of particles indexed by $i = 1 \dots, N_e^{(\ell)}$ from the current sampling proposal

$$\boldsymbol{\theta}_i^{(\ell)} \sim q^{(\ell)}(\boldsymbol{\theta}) \quad (16)$$

2. For the current particle history, that is for all $k = 1, \dots, \ell$ historical iterations and the ensembles of $i = 1, \dots, N_e^{(k)}$ particles in these iterations, evaluate the *deterministic mixture* (DM) proposal density $v^{(\ell)}$ for each particle $\boldsymbol{\theta}_i^{(k)}$

$$v^{(\ell)}(\boldsymbol{\theta}_i^{(k)}) = \frac{1}{N_h^{(\ell)}} \sum_{j=1}^{\ell} N_e^{(j)} q^{(j)}(\boldsymbol{\theta}_i^{(k)}), \quad (17)$$

where $N_h^{(\ell)} = \sum_{k=1}^{\ell} N_e^{(k)}$ is the total number of particles in the current particle history. The ratio $N_e^{(j)}/N_h^{(\ell)}$ corresponds to the weight of each iteration's proposal density $q^{(j)}$ in the current DM $v^{(\ell)}$. The particles $\boldsymbol{\theta}_i^{(k)}$ for each iteration k have been drawn independently from their respective sampling proposal distributions $q^{(k)}$. Nonetheless, using this DM concept from Owen and Zhou (2000) we can treat all the particles *as if* they were collectively drawn from the mixture $v(\cdot)$ where we happened to end up with exactly $N_e^{(k)}$ particles from each proposal.

475 3. Using the DM in (17) as our effective proposal for importance sampling compute unnormalized weights for the current particle history through

$$\tilde{w}_i^{(k)} = \frac{p(\mathbf{y} | \boldsymbol{\theta}_i^{(k)})p(\boldsymbol{\theta}_i^{(k)})}{v^{(\ell)}(\boldsymbol{\theta}_i^{(k)})}. \quad (18)$$

4. Using the entire particle history, compute $N_h^{(\ell)}$ self-normalized particle weights

$$w_i^{(k)} = \frac{\tilde{w}_i^{(k)}}{N_h^{(\ell)} Z_v^{(\ell)}}, \quad (19)$$

480 where the normalizing constant in the denominator

$$Z_v^{(\ell)} = \frac{1}{N_h^{(\ell)}} \sum_{k=1}^{\ell} \sum_{i=1}^{N_e^{(k)}} \tilde{w}_i^{(k)}, \quad (20)$$

also provides estimate for the model evidence using the DM proposal since

$$p(\mathbf{y}) = \int \frac{p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{v^{(\ell)}(\boldsymbol{\theta})} v^{(\ell)}(\boldsymbol{\theta}) d\boldsymbol{\theta} \simeq Z_v^{(\ell)}. \quad (21)$$

5. Compute the effective sample size (cf. Elvira et al., 2022) using *all* the $N_h^{(\ell)}$ historical particles

$$485 \quad N_{\text{eff}}^{(\ell)} = 1 / \sum_{k=1}^{\ell} \sum_{i=1}^{N_e^{(k)}} (w_i^{(k)})^2, \quad (22)$$

if the predefined maximum number of iterations has been reached ($\ell = \ell_{\text{max}}$) or a desired effective sample size is obtained $N_{\text{eff}}^{(\ell)} \geq \tau N_e$ given a predefined threshold $0 < \tau \leq 1$ then this ℓ becomes the final AMIS iteration which we denote as L .

If this stopping criterion is not satisfied, apply the clipping approach of Koblements and Míguez (2015) to improve proposal adaptation. First, identify the $\mathcal{T}$ -th largest unnormalized weight denoted $\tilde{w}_{\mathcal{T}}$ where $\mathcal{T} = \text{round}(\tau N_e)$. Subsequently, as

490 long as $\tilde{w}_{\mathcal{T}} > 0$, apply weight clipping

$$\tilde{w}_i^{(k)} \leftarrow \min(\tilde{w}_i^{(k)}, \tilde{w}_{\mathcal{T}}), \quad (23)$$

with subsequent renormalization via (19). Clipping ensures equality of the $\mathcal{T}$ largest clipped weights which guarantees a clipped $N_{\text{eff}} \geq \mathcal{T}$ that in turn leads to a more robust and less degenerate sampling proposal adaptation.

495 6. Resample an ensemble of $r = 1, \dots, N_e$ equally weighted particles $\boldsymbol{\theta}_r^{(\ell)}$ using the $N_h^{(\ell)}$ weights $w_i^{(k)}$ of the current particle history obtained from (19) via (23) if needed. Use of the index r emphasizes this is an ensemble of *resampled* particles which approximate the posterior rather than draws from the sampling proposals (16).

7. If the final AMIS iteration has been reached $\ell = L$ according to the stopping criterion in step 5, stop iterating and use the resampled particles $\{\boldsymbol{\theta}_r^{(\ell)}\}_{r=1}^{N_e}$ as a particle approximation of the posterior distribution. Otherwise, use the resampled particles to construct a new sampling proposal distribution $q^{(\ell+1)}$ for the next iteration, then update the iteration counter $\ell \leftarrow \ell + 1$ and return to step 1.

500

Generating an adaptive particle history by iterating the sequence of steps 1-7 outlined above until convergence is the general workflow for AMIS. Nonetheless, to be able to implement this algorithm for particle smoothing in data assimilation we need to specify the sampling proposal distributions $q^{(\ell)}$ and how the resampled particles can be used to update these proposals from one iteration to the next in step 7.

505 2.6.2 Adaptive particle batch smoother

Recall that we seek to enhance a popular cryospheric DA algorithm known as the PBS that was introduced by Margulis et al. (2015) and has since been widely adopted for cryospheric reanalysis (e.g. Margulis et al., 2016; Navari et al., 2016; Cortés and Margulis, 2017; Fiddes et al., 2019; Alonso-González et al., 2021; Liu et al., 2021; Giroto et al., 2024; Cao et al., 2025; Sun et al., 2025) by embedding it within the powerful and more general adaptive framework of the AMIS algorithm (Cornuet et al., 2012). We will refer to this proposed new DA method as the adaptive PBS, or AdaPBS for short. In the current AdaPBS implementation, we immediately make one simplification compared to the general AMIS algorithm laid out above which is to fix the number of particles $N_e^{(\ell)}$ sampled in each iteration to a constant ensemble size, i.e. $N_e^{(\ell)} = N_e$ for all iterations ℓ . The parsimonious choice of a constant ensemble size for all iterations reduces the number of hyperparameters that the user has to specify and facilitates comparisons with the conventional PBS with a given ensemble size. We still suspect that varying the size of the ensemble during the iterations may lead to performance improvements, which is a topic worthy of future research. In our case of a fixed ensemble size at each iteration, the ratio in the DM simplifies to $N_e^{(j)}/N_h^{(\ell)} = 1/\ell$ since $N_h^{(\ell)} = \ell N_e$ so each component of the DM proposal is equally weighted. As such, the typically suboptimal choice of the prior as the initial proposal, which is implicit in the PBS, with a mixture weight of $1/\ell$ becomes less influential as ℓ increases and the latter adaptive iterations together occupy a continuously increasing total mixture weight of $(\ell - 1)/\ell \rightarrow 1$ as $\ell \rightarrow \infty$.

For simplicity, but without loss of generality, in the current version of AdaPBS we employ multivariate normal (Gaussian) proposal distributions $q^{(\ell)}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta} | \boldsymbol{\mu}^{(\ell)}, \mathbf{C}^{(\ell)})$ that are parametrized by hyperparameters in the form of a mean vector $\boldsymbol{\mu}^{(\ell)}$ and covariance matrix $\mathbf{C}^{(\ell)}$. In principle, we could set these proposals to anything with at least the same support as the target posterior so we have great freedom in picking our proposal distribution. In practice, however, importance sampling works better if the proposal is similar to the target posterior distribution. As such, it is advantageous to update the hyperparameters of the successive proposal distributions using the current particle approximation of the posterior in the form of the ensemble $\boldsymbol{\theta}_r^{(\ell)}$ obtained in step 6 of AMIS. Recalling that this is an equally weighted ensemble, an obvious choice is to estimate these hyperparameters using ensemble statistics from the successive posterior approximations. Thus, in step 7 of AMIS the mean vector and covariance matrix for the Gaussian proposal in the next iteration $\ell + 1$ are simply set to the ensemble mean vector

$$\boldsymbol{\mu}^{(\ell+1)} = \frac{1}{N_e} \boldsymbol{\Theta}^{(\ell)} \mathbf{1}, \quad (24)$$

530 and ensemble covariance matrix

$$\mathbf{C}^{(\ell+1)} = \frac{1}{N_e} \left(\boldsymbol{\Theta}^{(\ell)} - \boldsymbol{\mu}^{(\ell+1)} \mathbf{1}^T \right) \left(\boldsymbol{\Theta}^{(\ell)} - \boldsymbol{\mu}^{(\ell+1)} \mathbf{1}^T \right)^T, \quad (25)$$

where the particle ensemble $\boldsymbol{\theta}_r^{(\ell)}$ is stored as column vectors in the $N_p \times N_e$ matrix $\boldsymbol{\Theta}^{(\ell)}$ and $\mathbf{1}$ is a $N_e \times 1$ column vector of ones.

If, using suitable transformations, we further require a prior $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \mathbf{C})$ that is Gaussian and employing the usual
535 Gaussian likelihood (4) of the PBS $p(\mathbf{y} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y} \mid \hat{\mathbf{y}}, \mathbf{R})$ where the mean $\hat{\mathbf{y}} = \hat{\mathbf{y}}(\boldsymbol{\theta})$ is the predicted observation vector and $\mathbf{R}$ is the observation error covariance then the (unnormalized) target posterior is

$$p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}) = A \exp\left(-\frac{1}{2}\phi(\boldsymbol{\theta})\right), \quad (26)$$

where $A = \det(2\pi\mathbf{R})^{-1/2} \det(2\pi\mathbf{C})^{-1/2}$ is a constant that does not depend on $\boldsymbol{\theta}$ and we have introduced the unnormalized and scaled negative log posterior ϕ (not to be confused with the uncertain bounded parameters $\boldsymbol{\varphi} = \mathcal{T}(\boldsymbol{\theta})$) defined as

$$540 \quad \phi(\boldsymbol{\theta}) = [\mathbf{y} - \hat{\mathbf{y}}]^T \mathbf{R}^{-1} [\mathbf{y} - \hat{\mathbf{y}}] + [\boldsymbol{\theta} - \boldsymbol{\mu}]^T \mathbf{C}^{-1} [\boldsymbol{\theta} - \boldsymbol{\mu}], \quad (27)$$

for economy. We may construct a similar expression for the Gaussian proposal distributions $q^{(\ell)}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{\mu}^{(\ell)}, \mathbf{C}^{(\ell)})$ of the form

$$q^{(\ell)}(\boldsymbol{\theta}) = c^{(\ell)} \exp\left(-\frac{1}{2}\psi^{(\ell)}(\boldsymbol{\theta})\right), \quad (28)$$

where we have defined the shorthand function

$$545 \quad \psi^{(\ell)}(\boldsymbol{\theta}) = [\boldsymbol{\theta} - \boldsymbol{\mu}^{(\ell)}]^T \mathbf{C}^{(\ell)-1} [\boldsymbol{\theta} - \boldsymbol{\mu}^{(\ell)}], \quad (29)$$

where $c^{(\ell)} = \det(2\pi\mathbf{C}^{(\ell)})^{-1/2}$ is a normalizing constant that only depends on the sampling proposal covariance $\mathbf{C}^{(\ell)}$. Inserting (29) in (17) with fixed N_e the Gaussian DM proposal at iteration ℓ has the following simple analytical form

$$v^{(\ell)}(\boldsymbol{\theta}) = \frac{1}{\ell} \sum_{j=1}^{\ell} c^{(j)} \exp\left(-\frac{1}{2}\psi^{(j)}(\boldsymbol{\theta})\right), \quad (30)$$

whereby it is readily verified that a numerically stable estimate of the logarithm of the unnormalized weights in (18) for the
550 particle history ($k = 1, \dots, \ell$ and $i = 1, \dots, N_e$) at iteration ℓ is given by

$$\log(\tilde{w}_i^{(k)}) = \log(A) - 0.5\phi(\boldsymbol{\theta}_i^{(k)}) - \text{LSE}_j\left(a^{(j)}(\boldsymbol{\theta}_i^{(k)})\right), \quad (31)$$

where with some abuse of notation the log-sum-exp term $\text{LSE}_j(\cdot)$ related to the DM is

$$\text{LSE}_j(a^{(j)}(\boldsymbol{\theta}_i^{(k)})) = a^{(*)}(\boldsymbol{\theta}_i^{(k)}) + \log\left(\sum_{j=1}^{\ell} \exp\left(a^{(j)}(\boldsymbol{\theta}_i^{(k)}) - a^{(*)}(\boldsymbol{\theta}_i^{(k)})\right)\right), \quad (32)$$

which is implicitly a function with $j = 1, \dots, \ell$ arguments

$$555 \quad a^{(j)}(\boldsymbol{\theta}_i^{(k)}) = \log(c^{(j)}) - \log(\ell) - 0.5\psi^{(j)}(\boldsymbol{\theta}_i^{(k)}), \quad (33)$$

whose maximum across ℓ iterations is $a^{(*)}(\boldsymbol{\theta}_i^{(k)}) = \max_j \left(a^{(j)}(\boldsymbol{\theta}_i^{(k)})\right)$. We emphasize that to obtain the logarithm of the unnormalized weights (31) for the particle history at each iteration ℓ both the negative log posterior in (27) and the LSE form

of the DM in (32) have to be evaluated for each of the ℓN_e particles. To obtain the logarithm of the self-normalized weights (19) we also need to use the logarithm of all the ℓN_e historical unnormalized weights from (31) to evaluate the logarithm of the evidence approximation (20) which is given by

$$\log(Z_v^{(\ell)}) = -\log(\ell N_e) + \text{LSE}_{k,i} \left(\log(\tilde{w}_i^{(k)}) \right), \quad (34)$$

where the new log-sum-exp term $\text{LSE}_{k,i}(\cdot)$ is implicitly a function with ℓN_e arguments

$$\text{LSE}_{k,i} \left(\log(\tilde{w}_i^{(k)}) \right) = \Omega + \log \left(\sum_{k=1}^{\ell} \sum_{i=1}^{N_e} \exp \left((\log(\tilde{w}_i^{(k)}) - \Omega) \right) \right), \quad (35)$$

where $\Omega = \max_{k,i} \left(\log(\tilde{w}_i^{(k)}) \right)$. Note that (34) only needs to be evaluated once per iteration ℓ . Finally, combining (31) and (34), the stable logarithm of the self-normalized weights (19) that we seek can now be computed as

$$\log(w_i^{(k)}) = \log(\tilde{w}_i^{(k)}) - \log(\ell N_e) - \log(Z_v^{(\ell)}), \quad (36)$$

which we have expressed in terms of the log evidence approximation (34) since this is a useful additional output from AdaPBS for hierarchical Bayesian inference (Robert, 2007; Murphy, 2023). The self-normalized weights for AdaPBS are now obtained trivially by taking the exponential of (36). These weights can then be resampled N_e times to obtain an equally weighted ensemble of particles as described in step 6 of the AMIS algorithm. Next, as described in step 7, upon convergence these N_e resampled particles are output as a particle approximation of the posterior. Otherwise we use these particles to design the Gaussian sampling proposal through (24)&(25) for the next iteration $\ell + 1$.

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. Workflow of the adaptive particle batch smoother (AdaPBS), illustrating its extension (green) from the non-iterative PBS method (yellow). The algorithm iterates by adapting the proposal until the effective sample size (N_{eff}) reaches a predefined fraction of the ensemble size (N_e), determined by the diversity threshold (τ), or until a maximum number of iterations (ℓ_{max}) is reached. AdaPBS is mathematically equivalent to PBS if the diversity threshold is met in the first iteration ($\ell = 1$) or if the algorithm is constrained to a single iteration ($\ell_{\text{max}} = 1$).

2.7 Probabilistic evaluation diagnostics

To evaluate the performance of the respective data assimilation schemes we employ two probabilistic evaluation diagnostics, namely the Continuous Ranked Probability Score (CRPS; Gneiting et al., 2005) and the Kullback–Leibler Divergence (KLD;

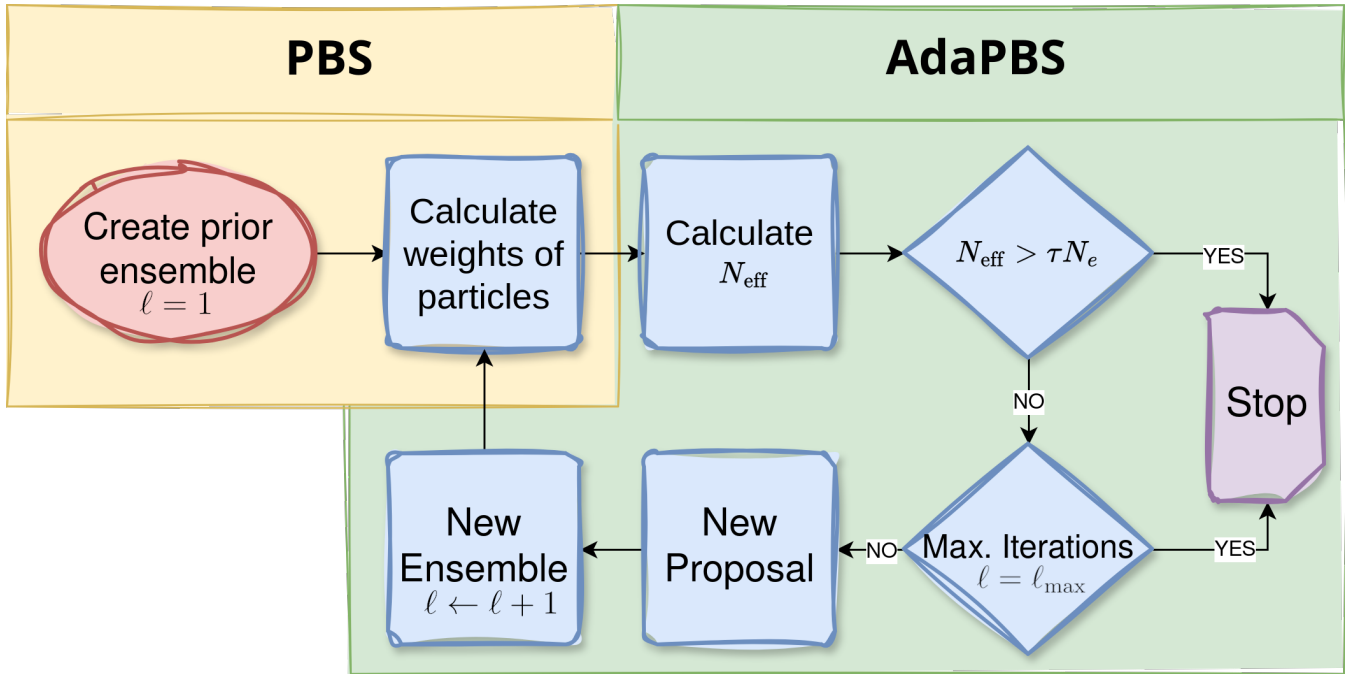


Figure 1. Flowchart showing the workflow in the adaptive particle batch smoother (AdaPBS) as an extension (green) of the non-iterative PBS method (yellow). The AdaPBS algorithm continues iterating by adapting the proposal until convergence, achieved when the effective sample size (N_{eff}) reaches at least a fraction of the ensemble size (N_e) determined by the desired diversity threshold (τ), or until a pre-defined maximum number of iterations (ℓ_{max}) is reached. Both in the special case of convergence in the first iteration ($\ell = 1$) and in the non-iterative case ($\ell_{\text{max}} = 1$), AdaPBS reduces to PBS.

Murphy, 2023). These probabilistic verification measures allow us to evaluate the performance of the entire approximate posterior distribution as represented by an ensemble of particles rather than just point estimates such as the posterior mean.

590 The CRPS is a strictly proper scoring rule defined by (Gneiting and Raftery, 2007)

$$\text{CRPS}(P, x^*) = \int [P(x) - H(x - x^*)]^2 dx, \quad (37)$$

where $P(x) = \int_{-\infty}^x p(\chi) d\chi$ is the cumulative (posterior or prior) probability distribution of the uncertain variable or parameter x of interest with density $p(x)$ that we are inferring, x^* is the reference truth value, and $H(\cdot)$ is the Heaviside function which is 1 for positive arguments (here $x > x^*$) and 0 otherwise. The CRPS is non-negative and inherits the same units as the variable x of interest. In the special case of a deterministic forecast, where $P(x)$ is also a Heaviside function, the CRPS reduces to the absolute error. As such, the CRPS is negatively oriented with the best possible value being 0 indicating perfect agreement with the reference. In the general non-degenerate probabilistic setting the CRPS measures both the accuracy (goodness of fit) and the precision (sharpness) of the distribution $p(x)$ relative to the reference x^* , making it an apt choice for evaluating ensemble-based data assimilation experiments in cryospheric science (Piazzi et al., 2018; Cao et al., 2025). In practice, in

595

line with previous studies (Alonso-González et al., 2023; Mazzolini et al., 2025; Alonso-González et al., 2026), we estimate the CRPS for probabilistic predictions obtained via ensemble-based DA schemes in MuSA by assuming that the marginal predictive prior and posterior snow depth distributions follow a normal distribution. Thereby, we only need to store the mean and standard deviation of the ensemble predictions of each scheme which can be plugged into the simple analytical expression for the Gaussian CRPS given by Gneiting et al. (2005).

The CRPS is used to evaluate the performance of the ensemble of snowpack states obtained in MuSA by comparing them to assimilated or independent observations. In a similar vein, we employ the KLD as a probabilistic evaluation diagnostic that quantifies how close the approximate posterior distributions over parameters are to the reference posterior obtained through MCMC simulation using the RAM method. In particular, we use the so-called reverse KLD defined by (Murphy, 2023)

$$D_{\text{KL}}(q(\boldsymbol{\theta} | \mathbf{y}) || p(\boldsymbol{\theta} | \mathbf{y})) = \int q(\boldsymbol{\theta} | \mathbf{y}) \log \left(\frac{q(\boldsymbol{\theta} | \mathbf{y})}{p(\boldsymbol{\theta} | \mathbf{y})} \right) d\boldsymbol{\theta}, \quad (38)$$

where $q(\boldsymbol{\theta} | \mathbf{y})$ is an approximating distribution of the target posterior distribution $p(\boldsymbol{\theta} | \mathbf{y})$. The KLD, also known as relative entropy (MacKay, 2003), is a dimensionless measure of the distance between the two distributions that it takes as arguments. As with the CRPS, it is negatively oriented and non-negative with the best possible value being 0 in the case that the input distributions used as arguments are equal. It happens to be asymmetric in these arguments, and this so-called reverse form of the KLD is widely used as an objective function in variational Bayesian inference which recasts inference as an optimization problem over tractable variational distributions $q(\boldsymbol{\theta} | \mathbf{y})$ used to approximate the intractable posterior $p(\boldsymbol{\theta} | \mathbf{y})$ (Murphy, 2023). While this use of reverse KLD as an objective function motivated our use of this metric, we use it in the slightly different context of evaluation rather than optimization. In particular, we use the reference posterior samples simulated via MCMC using RAM to define the target posterior distribution $p(\boldsymbol{\theta} | \mathbf{y})$ and we use the reverse KLD to measure how close the more tractable approximations $q(\boldsymbol{\theta} | \mathbf{y})$ from the ensemble-based DA schemes, including the AdaPBS, are to this target. To simplify the calculation of the KLD, especially given that we only have samples from the distributions p and q , we focus on the reverse KLD of the marginal distributions of the perturbation parameters θ in the vector $\boldsymbol{\theta}$ in transformed space as introduced in Section 3.2. Note that the KLD is invariant to transformations (Murphy, 2023), and by sticking to the transformed unbounded space we can better approximate these marginals as Gaussian. Assuming Gaussianity, the marginal reverse KLD has a simple analytical solution of the form (see Supplementary material 5.1.2 of Murphy, 2023)

$$D_{\text{KL}}(q(\theta | \mathbf{y}) || p(\theta | \mathbf{y})) \simeq \log \left(\frac{\sigma_p}{\sigma_q} \right) - \frac{1}{2} + \frac{(\mu_p - \mu_q)^2 + \sigma_q^2}{2\sigma_p^2}, \quad (39)$$

where μ_q and σ_q are the mean and standard deviation, respectively, of the ensemble-based approximation $\theta_i \sim q(\theta | \mathbf{y})$. In addition to computing reverse KLD of the various ensemble-based DA schemes in MuSA we also compute the reverse KLD of the marginal prior, i.e. setting $q(\theta | \mathbf{y}) = p(\theta)$, as an additional benchmark to help contextualize this distance measure. For all these reverse KLD calculations, we use the MCMC samples obtained via RAM (Section 2.3) to define the reference marginal posterior distribution $\theta_j \sim p(\theta | \mathbf{y})$ with reference sample mean μ_p and standard deviation σ_p .

3 Data and methods

3.1 Data assimilation framework and experimental design

All experiments presented in this study were developed using the Multiple Snow Data Assimilation System (MuSA) (Alonso-González et al., 2022). MuSA is a versatile, open-source data assimilation tool designed to integrate diverse snowpack ob-
635 servations with an ensemble of numerical snowpack simulations. Its modular architecture allows users to implement different algorithms and snowpack models. In addition to the set of algorithms already available in MuSA, we have implemented three new algorithms to develop the experiments proposed in this work. This includes the aforementioned AdaPBS algorithm and two MCMC algorithms, namely RWM and RAM, where RAM is used as a sta. The updated MuSA code is released as a new version of the tool (Alonso-González and Aalstad, 2025). In order to provide an intuitive understanding of the behavior of the
640 AdaPBS, we have performed several experiments that compare its performance to that of previously proposed ensemble-based cryospheric DA algorithms.

3.2 Experiment 1: Assimilating drone-based snow depth in a temperature index model

First, we assimilated snow depth observations in an ensemble of simulations generated by a simple temperature index model (Hock, 2003) implemented in MuSA. We emphasize that the primary objective of these experiments is to assess the perfor-
645 mance of the AdaPBS algorithm by exploiting a simpler yet computationally efficient model and not to achieve the most physically complex simulations possible. Moreover, in part due to their simplicity in terms of input data and few calibration parameters, temperature index models are being used effectively in operational settings (Lussana et al., 2018) as well as for hemispheric-scale snow reanalysis (Elias Chereque et al., 2024) and global glacier modeling (Rounce et al., 2023). Our tem-
perature index forward modeling implementation relies on the strong assumption of constant density, fixed at a climatological
650 value of $300 \text{ [kg m}^{-3}\text{]}$, along with a seasonally constant hourly temperature index factor a set to $0.1375 \text{ [mm h}^{-1} \text{ K}^{-1}\text{]}$. Typically a would also be treated as an uncertain parameter (e.g. Rounce et al., 2023), but to facilitate the comparison of schemes and visualizing the results we chose to fix it here. Thus, the near surface air temperature $T_n \text{ [K]}$ dependent snowmelt rate $M_n \text{ [mm h}^{-1}\text{]}$ at each hourly timestep n is computed as follows:

$$M_n = \max(a[T_n + b - T_0], 0), \quad (40)$$

655 where the melting temperature $T_0 = 273.15 \text{ [K]}$ is set to the freezing point of water. Often this is treated as a calibration parameter to account for errors in the air temperature forcing, but we do this more explicitly by perturbing the air temperature field itself with an additive bias parameter b that we seek to infer. By combining the snowmelt rate with the snowfall rate $S_n \text{ [mm h}^{-1}\text{]}$, the hourly SWE $D_n \text{ [mm]}$ is updated via

$$D_{n+1} = \max(D_n + cS_n - M_n, 0), \quad (41)$$

660 where c is the precipitation perturbation parameter that we also seek to infer. Note that the air temperature bias b also affects the snowfall rate S_n directly because the precipitation phase is diagnosed using the perturbed air temperature $T_n + b$.

The temperature index model DA experiments were conducted in the Izas experimental catchment, in the Central Spanish Pyrenees (Revuelto et al., 2017). The snow depth data were acquired through drone surveys utilizing structure-from-motion techniques (Revuelto et al., 2021). Meteorological forcing data in the form of air temperature and total precipitation (P), were generated using the Micromet downscaling tool (Liston and Elder, 2006b), driven by the ERA5 reanalysis (Hersbach et al., 2020). These experiments were conducted for a single grid cell from which we retrieve both the forcing and observations using a grid spacing of 5 meters spatial resolution. The cell was located in a concavity within the basin and therefore exhibited significant snow accumulations resulting from snow redistribution that makes it particularly challenging to simulate for the temperature index model with standard forcing. 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). 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). The ensemble of simulations was generated by perturbing the precipitation and temperature timeseries with the pseudo-randomly sampled constant in time prior perturbation parameters b and c covering the 2018/2019 snow season. The data assimilation window covered the whole snow season and, as such, our experiments were developed using batch smoothers, but we would still expect similar relative performance (i.e. ranking) of these cryospheric DA schemes when applied as filters.

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). The prior probability distribution for the additive air temperature bias is a normal distribution with mean $\mu_b = 0$ and standard deviation $\sigma_b = 1$. For the prior on the multiplicative precipitation scaling, we use a strictly positive lognormal distribution with an associated normal mean $\mu_c = 0.1$ (so the actual median perturbation in model space is $\exp(\mu_c) = 1.11$) and standard deviation $\sigma_c = 0.5$. Note that we carry out inference directly on the unbounded space of the log-transformed precipitation perturbation parameter and then apply an exponential transform to map this back to the model parameter c following common practice (Gelman et al., 2013; Alonso-González et al., 2022). The location parameters of these prior distributions were chosen in a conservative manner, i.e., near unbiased, with a mean temperature bias of 0 and a precipitation scaling centered close to 1. This corresponds to quite a general case where, without additional background knowledge or data, the orientation of the forcing bias is not known *a priori*. The relatively large prior spread parameters (σ_c and σ_b) encode considerable initial epistemic uncertainty in the two forcing perturbation parameters.

Here we have compared the results of the approximate Bayesian inference using different DA algorithms, to evaluate and benchmark the performance of the novel AdaPBS algorithm. First, we have solved the data assimilation problem using adaptive MCMC using RAM (Vihola, 2012), which we consider a gold-standard albeit computationally costly reference. Taking inspiration from Emerick and Reynolds (2011), we initialized the Markov chain used in RAM at the approximate posterior mean obtained from ES-MDA with $N_a = 4$ iterations and $N_e = 100$ ensemble members. The idea is to accelerate MCMC by starting

the chain within or close to the typical set of the target posterior distribution. The length of the Markov chain in the RAM algorithm was set to 20000 and the initial 10% of the samples were discarded as burn in. Discarding the burn in period avoids initialization artifacts whereas the relatively long chain gives the Markov Chain time to reach its target posterior distribution.

700 We then performed the same exercise using a Particle Batch Smoother (PBS), an ES, an ES-MDA and finally we ran the AdaPBS. The hyperparameters of the analysis (i.e. the $N_e = 100$ number of ensemble members and the prior distributions) were the same in all cases. In the case of the AdaPBS, the N_{eff} limit was set to 30%, which means that for our $N_e = 100$ member ensemble at least 30 particles should show non-negligible weights before stopping the iterations. The results of each algorithm were compared with the samples from MCMC-RAM which we used as a reference posterior. The agreement between
705 the approximate posteriors obtained via ensemble-based DA and the MCMC reference was quantified using the KLD as a probabilistic evaluation diagnostic as described in Section 2.7.

3.3 Experiment 2: Assimilating hourly ESM-SnowMIP snow depth data in FSM2

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 (FSM2; Essery et al., 2025). In these experiments, we assimilated hourly observations
710 of snow depth at six different geographical locations using both 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). Assimilating snow depth observations at different locations over different periods of time allowed us to obtain a
715 general overview of the behavior of the algorithms in different climates and weather regimes. Observations were obtained from a database generated under the framework of the Earth System Model-Snow Model Intercomparison Project (ESM-SnowMIP; Ménard et al., 2019). Specifically, we conducted simulations at Reynolds Mountain East, Idaho, USA (RME); Sapporo, Japan (SAP); Senator Beck, Colorado, USA (SNB); Sodankylä, Finland (SOD); Swamp Angel, Colorado, USA (SWE) and Weissfluhjoch, Switzerland (WFJ). The meteorological forcing was obtained directly from the nearest grid cell of the ERA5 reanalysis (Hersbach et al., 2020) without applying topographic corrections. While site-level meteorological forcing was available
720 from Ménard et al. (2019), we used the more uncertain, unadjusted global reanalysis data to provide a more challenging experimental setup that is generally applicable to any site with snow depth data. Specifically, each water year, we assimilated a large batch of $24 \times 365 = 8760$ hourly snow depth observations.

To explore the performance of the algorithms in the case of a higher dimensional parameter space, we have generated 7 per-
725 turbation parameters to correct the forcing (air temperature, precipitation, relative humidity, surface pressure, wind speed, and incoming long- and short-wave radiations). On top of this, we have also included the 12 main tunable internal model parameters related to snow (not vegetation) in FSM2 (Essery et al., 2025) as listed in Table 1. The prior distributions for the temperature and precipitation parameters are the same as those for the previous experiment described in Section 3.2. For the new model parameters, we used a multiplicative logit-normal distribution bounded in the physical space between 0.8 and 1.2, defined by
730 the moments of its underlying normal distribution $\mu = 0$ and $\sigma = 1$. As such, we perturb these new parameters in the $\pm 20\%$

range from their reference values in the physical space, ensuring that the combinations parameters remain in their physical bounds, maintaining the numerical stability of FSM2. We performed a dependent validation to compare the performance of AdaPBS with ES-MDA and PBS, by comparing the posterior snow simulations with the assimilated observations, using bias and root mean squared error (RMSE) as well as an uncertainty-aware metric in the form of the continuous ranked probability score (CRPS) described in Section 2.7. We removed the timesteps where *both* the simulation and observations had snow depths of zero in the computation of the error and CRPS to avoid artificially deflating the validation metrics by not considering the seasonality of the snow depth values. That is to say, we are primarily concerned with model performance in the periods in which the model or observations indicate the presence of a seasonal snowpack. It is only in these periods that the model is susceptible to snow commission or omission errors rather than the mostly trivially correct no-snow prediction outside the snow season.

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). 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. The non-iterative PBS is also tested in this experiment to directly assess the impact of the adaptations within the iterative AdaPBS.

4 Results and discussion

4.1 Inter-comparison of algorithms using a temperature index model

Despite relying on identical priors, observations, observation error model, and forward model in the form of a temperature index snow model, the performance of the DA algorithms differed substantially. This conclusion aligns with the few previous publications that developed inter-comparison experiments on ensemble-based cryospheric data assimilation algorithms (Leisenring and Moradkhani, 2011; Margulis et al., 2015; Aalstad et al., 2018; Alonso-González et al., 2022; Cao et al., 2025) and highlights the importance of selecting the algorithm with careful consideration of the problem at hand by balancing accuracy and computational cost.

Given the highly informative nature of the drone-based snow depth observations, the posterior approximation obtained through MCMC samples via RAM showed a challenging, narrow, banana-shaped distribution (Fig. 2). This demonstrates that even in this relatively simple case—where in practice we are only updating 2 parameters (b and c) by assimilating just a handful

Table 1. Tabular overview of the internal FSM2 parameters that we perturbed at ESM-SnowMIP sites. Columns from left to right: the parameter name in the FSM2 code, its default value, physical units (- is dimensionless), and a description adapted from Essery (2015).

Name	Default	Units	Description
asmn	0.5	-	Min. albedo for melting snow
asmx	0.85	-	Max. albedo for fresh snow
eta0	3.7×10^7	Pa s	Reference snow viscosity
hfsn	0.1	m	Snow cover fraction depth scale
rgr0	5×10^{-5}	m	Fresh snow grain radius
rho _w	300	kg m ⁻³	Wind-packed snow density
Salb	10	kg m ⁻²	Snowfall to refresh albedo
snda	2.8×10^{-6}	s ⁻¹	Thermal metamorphism parameter
tcld	1000	h	Cold snow albedo decay time scale
tmlt	100	h	Melting snow albedo decay time scale
Wirr	0.03	-	Irreducible liquid water content of snow
z0sn	0.001	m	Snow surface roughness length

(5) of snow depth observations—efficiently sampling from the posterior is nonetheless a challenging task. This is probably exacerbated in this case by the fact that we are using batch smoothers. On the one hand, such smoothers can provide more information than filters by assimilating the complete trajectory of observations in the water year all at once and propagating information backwards in time. On the other hand, the larger immediate information gain from batch smoothers can make posterior sampling more challenging than with sequential filters.

The original PBS algorithm of Margulis et al. (2015) has several benefits. It is relatively easy to implement, relies on few assumptions, and its computational cost is lower than that of other schemes, even when considering the same number of members in the ensemble since (unlike the ES) it does not require any reruns. These benefits do come at a cost, as PBS is prone to collapse in more challenging settings. This means that sometimes, especially with very informative (i.e. numerous and/or accurate) observations, the ensemble collapses and degeneracy ensues (Snyder et al., 2008; Morzfeld et al., 2017). Using a familiar analogy, this is like looking for smaller needles in a haystack. A similar result arises if the priors are relatively far (as measured, e.g., by a modified variant of the KLD in Section 2.7) from the posterior in the parameter space, such as in the case of overconfident and/or strongly biased priors. More generally, ‘far’ is not just the distance between the prior and posterior mean, but whenever the overlap between the prior and the posterior is small, including when a broad and/or high-dimensional prior encompasses a narrow posterior. Following a similar analogy, this is akin to having a larger haystack. These needle and haystack effects can yield sub-optimal solutions with high Monte Carlo variance (Robert and Casella, 2004) that make it difficult, if not impossible, to accurately estimate uncertainty. So even if SNIS, which powers the PBS, is asymptotically unbiased (Rainforth et al., 2020), the PBS in particular and basic importance sampling in general may exhibit a large variance since the prior is implicitly used as the proposal, which is generally far from the posterior (MacKay, 2003). In the worst

case, this will result in complete particle degeneracy with $N_{\text{eff}} \simeq 1$. This was clearly the case in our particular experiment, as shown in Figure 2, where the majority of the weight collapsed onto a single particle despite the fact that the range of the prior parameter distribution clearly encompassed the reference posterior MCMC samples obtained via the RAM algorithm. In the model state space, the fact that a single particle carries all the probability leads to a degenerate posterior snow depth ensemble that is not close to the assimilated observations and completely lacks well-calibrated uncertainty quantification.

In this case, the skewed prior on the precipitation perturbation combined with the challenging banana shape of the posterior are key challenges leading to degeneracy since very few particles were sampled in the range of the typical set of the posterior. The situation would improve if a more informative prior were used with a higher median precipitation perturbation parameter and even a positive correlation between the perturbations to better sample the banana shape. The problem, of course, is that these prior hyperparameter settings are rarely, if ever, known a priori, even from expert knowledge, and setting them directly based on the data leads to incoherent ‘double-dipping’ via so-called empirical Bayes (Robert, 2007; Gelman et al., 2013; Murphy, 2023). Adaptive particle methods present a solution that, instead of incoherently changing the prior to avoid degeneracy, simply adapts the proposal towards the target posterior so as to perform more sample efficient approximate Bayesian inference that remains coherent.

It may seem surprising that importance sampling via PBS struggles to accurately capture the reference posterior in this apparently simple experiment with a two-dimensional parameter space. An alternative would be gradient-based optimization techniques, such as gradient descent or Newton’s method, used in both machine learning (Murphy, 2023) and variational DA (Bannister, 2017), to identify the mode with relatively few iterations and find a Laplace approximation of the posterior (MacKay, 2003). However, as is often the case in practical DA in the absence of automatic differentiation (Murphy, 2023), we do not have access to gradients and are focused on the harder problem of ‘black-box’ (i.e., gradient-free) inference that is naturally handled by ensemble-based methods (Sanz-Alonso et al., 2023). While similar low-dimensional calibration problems are routinely encountered in hydrology and can be tackled via methods such as generalized likelihood uncertainty estimation (GLUE; Beven and Binley, 1992), SCEM-UA optimization (Vrugt et al., 2003), or DREAM MCMC sampling (Vrugt et al., 2008), Nonetheless, these methods usually involve orders of magnitude more forward model evaluations than the $N_e = 100$ particles used by PBS in this experiment. For example, Clark et al. (2006) demonstrated the need for a computationally intensive SODA approach, using SCEM-UA in an outer loop to optimize parameters, together with an EnKF inner loop for state estimation, to rigorously calibrate a simple two-parameter snow model. Furthermore, the widely used hydrological calibration technique GLUE (Beven and Binley, 1992), tailored to identify equifinality (Beven, 2006), is equivalent to PBS when using a Gaussian likelihood. As such, GLUE could only outperform PBS by increasing the ensemble size or modifying the likelihood, which are equally applicable to PBS. Both black-box optimization and inference are challenging problems where particle methods have proven to be highly competitive (Duan et al., 2023; Chopin and Papaspiliopoulos, 2020). In fact, these particle methods can be seen as formalizing highly successful meta-heuristic evolutionary algorithms that are widely used for optimization (Del Moral, 2004). The takeaway here is that particle methods can be improved either by increasing the ensemble size, at considerable computational expense given typically poor scaling (Snyder et al., 2008), or by adapting them for example using various iterative methods (Bugallo et al., 2017; Chopin and Papaspiliopoulos, 2020).

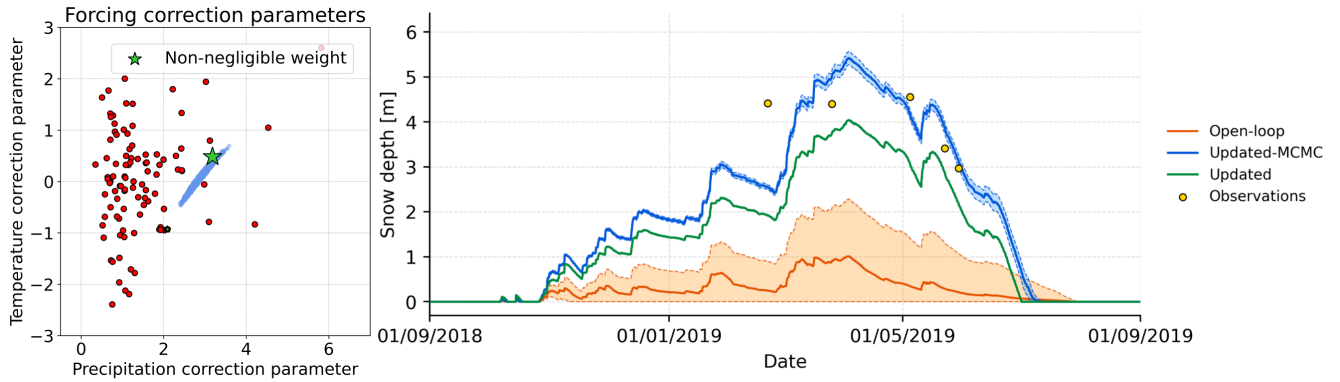


Figure 2. Results from the particle batch smoother (PBS) assimilating drone-based snow depth observations in a temperature index model for a single cell at Izas in water year 2019: The model parameter space visualized as the temperature bias parameter b on the y -axis and the precipitation correction c on the x -axis with the prior ensemble of particles shown with red dots while 2 green stars with size proportional to weight indicate particles with non-negligible PBS weights ($N_{\text{eff}} = 1.26$) and the blue banana-shaped distribution shows the reference posterior MCMC samples obtained via RAM (left panel); The model state space (right panel) showing the trajectory of snow depth with mean (solid line) ± 1 standard deviation (shading) from the prior (orange), PBS posterior (green), and reference MCMC posterior (blue), along with the assimilated observations (yellow dots).

The ES (van Leeuwen and Evensen, 1996), i.e. the batch smoother version of the EnKF (Evensen, 1994), has also been widely used in cryospheric DA particularly in reanalysis settings (Durand et al., 2008; Girotto et al., 2014). Although the EnKF has mainly been used to directly update snow model states (e.g. De Lannoy et al., 2012), the ES is better suited to update the forcing and internal parameters in a forcing formulation of the data assimilation problem (Evensen et al., 2022). Since the update moves the parameters in parameter space through the ensemble Kalman analysis step, with the ES it is necessary to generate an ensemble of snow model simulations twice, once with the prior parameters and once with the posterior parameters. This has the obvious advantage of maintaining model stability and physical consistency as long as the parameters remain within their physical bounds (Alonso-González et al., 2022). The latter boundedness can be easily controlled using transformation techniques such as Gaussian anamorphosis that also help to accommodate the Gaussian assumption (Bertino et al., 2003; Aalstad et al., 2018). However, with the ES this forcing formulation approach entails a non-parallelizable doubling of the computational cost ($2N_e$ simulations) compared to solving the same problem with the same ensemble size using the PBS (N_e simulations). In practice, the fact that the ES is less prone to ensemble collapse allows one to reduce the ensemble size compared to PBS such that in some problems the ES may actually end up being less costly. This gives a lot of flexibility when configuring the assimilation system according to the available computational resources especially in spatio-temporal settings where localization methods are more easily implemented with ensemble Kalman schemes (Evensen et al., 2022). Despite these advantages, the linear assumption in combination with models used in cryospheric sciences, which are typically non-linear, can lead to highly suboptimal posterior approximations with poorly calibrated and even underconfident uncertainty estimates.

835 In fact, this is what we see in Figure 3 with slight and underconfident updates both in parameter and state space, in line with previous studies (Aalstad et al., 2018; Alonso-González et al., 2022).

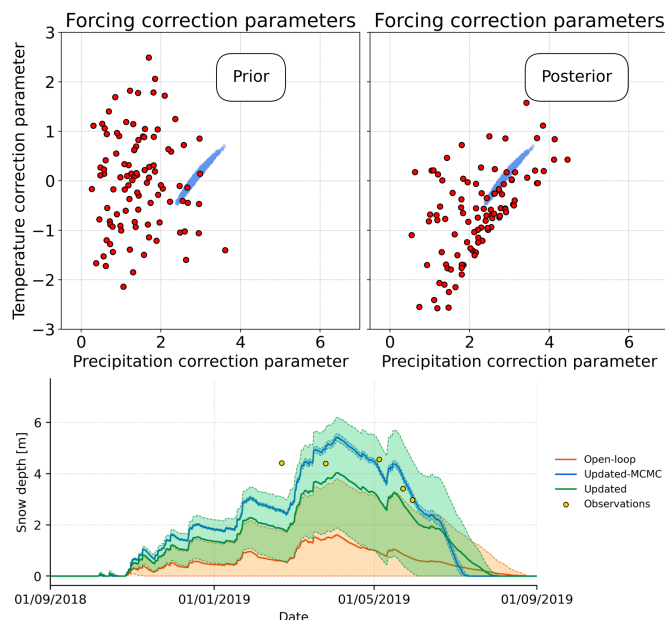


Figure 3. Analogous to Figure 2 but for the ensemble smoother (ES) with the prior (top left) and posterior (top right) in the model parameter space and the trajectory of snow depth (bottom) in the model state space.

The typical underlying linear assumption of ES, which is the main cause of suboptimal results, can be strongly violated in practical cryospheric DA problems. The iterative ES-MDA (Emerick and Reynolds, 2013) has proven to be a powerful cryospheric DA algorithm in the few previous inter-comparisons that have been carried out to date (Aalstad et al., 2018; Alonso-González et al., 2022). Iterations allow for a progressive movement of parameters towards the posterior via likelihood tempering (Chopin and Papaspiliopoulos, 2020; Murphy, 2023), which in practice limits the effects of model non-linearity as shown in Figure 4. One limitation is that the number of iterations must be pre-set in the ES-MDA of Emerick and Reynolds (2013), making it an important hyperparameter that must be carefully adjusted balancing between computational cost and solution robustness. When implemented in a distributed cell-by-cell manner, this is problematic, as it is difficult to adjust the number of iterations cell by cell, opting in practice for a fixed number of iterations for the whole domain. Such a global ‘one size fits all’ approach is very likely to result in a waste of computational resources through an unnecessary number of model realizations compared to a locally adaptive approach. In addition, the computational cost of the linear algebra operations needed for ES-MDA may become non-negligible in the case of assimilating a large number of observations.

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. The AdaPBS has the potential to be a powerful iterative DA method, aspiring to sample

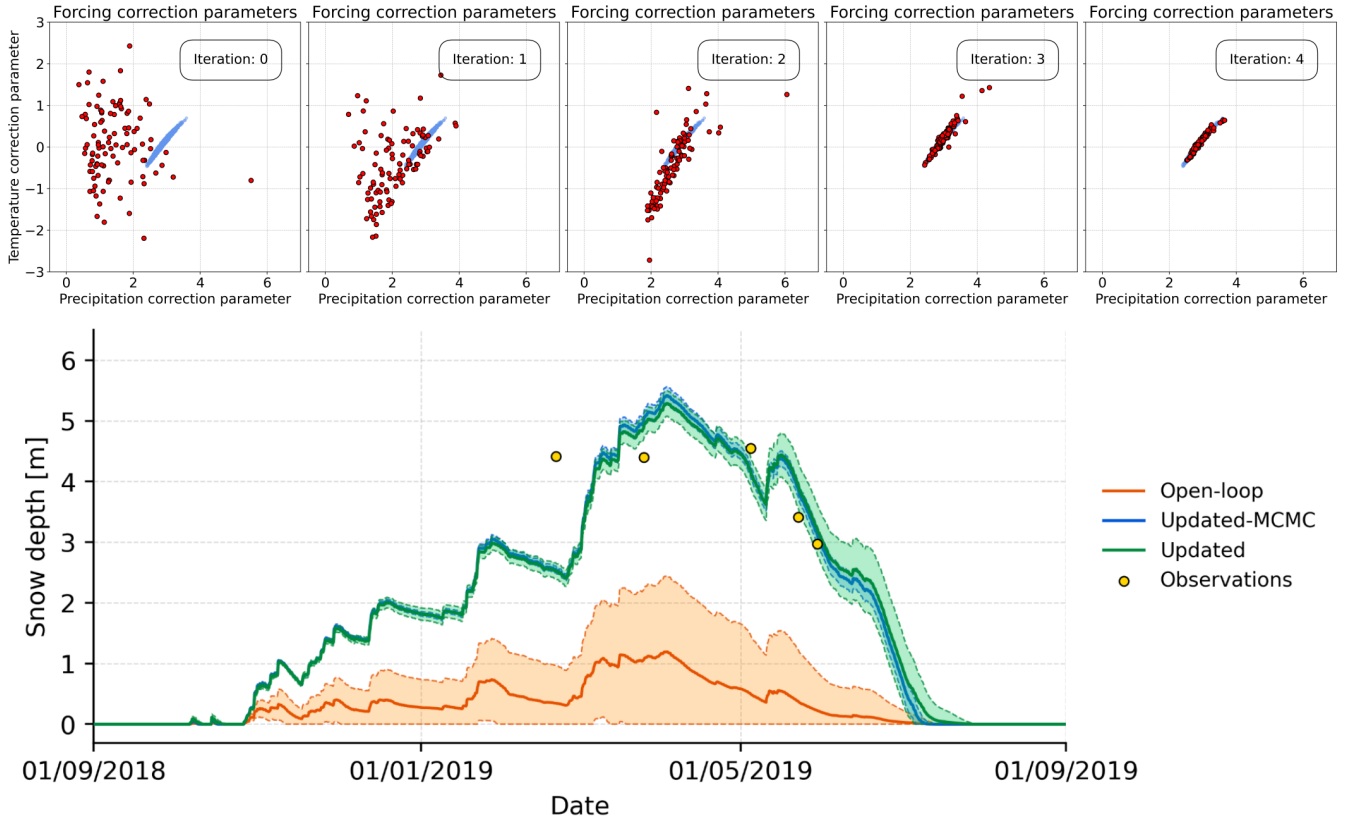


Figure 4. Analogous to Figure 3 but for an iterative ensemble smoother (IES) in the form of the ensemble smoother with multiple data assimilation (ES-MDA). The top panels show the evolution from the prior to the posterior ensemble members (red) across the MDA iterations along with the reference banana-shaped MCMC posterior samples (blue). As before, the bottom panel shows the predicted snow depth in model state space, but with a better calibrated IES posterior (green). The zero-based iteration counter from 0 to 4 corresponds to ℓ in the algorithm description in Section 2.4 with $N_a = 4$ iterations of the update step and $N_a + 1 = 5$ iterations of the prediction step.

from the posterior with performance similar to that of ES-MDA. Unlike the ES-MDA, AdaPBS does not rely on assumptions of linearity or Gaussianity, which facilitates its implementation, especially for users with limited experience in data assimilation. This also opens up for the possibility of using tailored likelihood functions such as those involving zero-inflation (Smith et al., 2010; Tang et al., 2023) that may be better suited to double-bounded cryospheric variables.

855 A clear computational advantage of the AdaPBS over ES-MDA is that it does not require pre-selecting a fixed number of iterations, enabling the use of early stopping strategies throughout the simulation domain. This allows the number of iterations to be automatically adapted for each cell in a spatio-temporally distributed model domain, depending on the difficulty of the local (or localized) inference problem being solved. We have used an early stopping criterion that stops the iterations when a threshold on the classical estimate of the ES (SN_{eff} in Equation (22)), i.e., a minimum number of ensemble members with

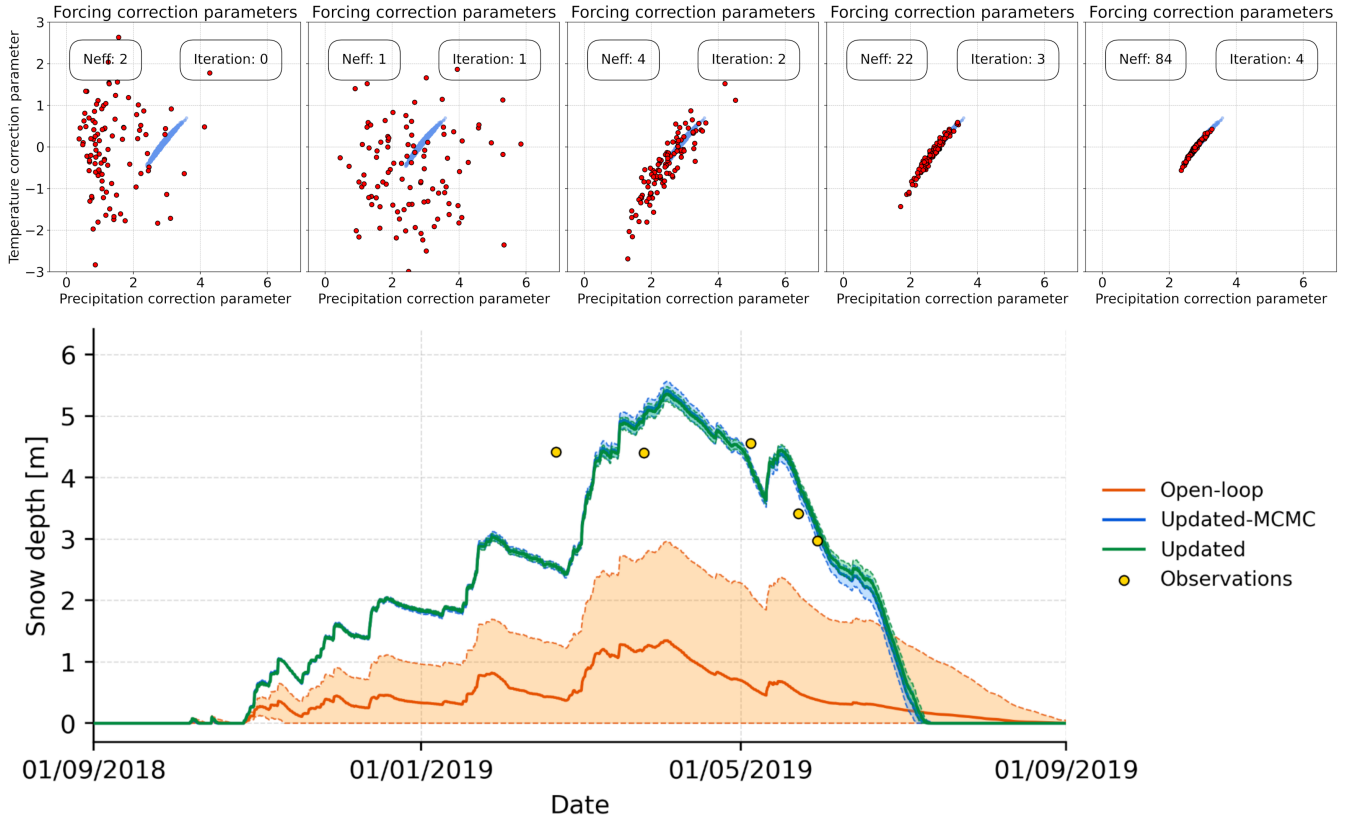


Figure 5. Analogous to Figure 4 but for the adaptive particle batch smoother (AdaPBS) showing the particle and effective sample size (ESS, i.e., N_{eff}) evolution across iterations. Note that the results after the initial iteration (top left) correspond exactly (ignoring Monte Carlo variance) to those that would be obtained from the PBS in Figure 2. The zero-based iteration counter from 0 to 4 corresponds to $\ell - 1$ in the algorithm description in Section 2.6, so here the AdaPBS converged (here $N_{\text{eff}} > 30$) in $\ell = 5$ iterations with exactly the same computational cost as the ES-MDA with $N_a + 1 = 5$ iterations of the prediction step.

considerable weight is reached, making it more robust to collapse than PBS. In this case, we used a relatively high diversity threshold of $\tau = 0.3$ for N_{eff}/N_e , which in practice resulted in the same number of iterations as ES-MDA. However, it should be noted that in this example, a respectable $N_{\text{eff}} = 22$ was already achieved by iteration 3 as shown in Figure 5, which may be sufficient for many applications.

Thus, a key hyperparameter to set in AdaPBS is the diversity threshold τ , which controls the target minimum effective sample size N_{eff} . Since τ is directly related to the desired performance of the algorithm, it is more interpretable for users than the fixed number of iterations N_a required in ES-MDA (Emerick and Reynolds, 2013). This makes AdaPBS an easy-to-implement algorithm, resistant to collapse, and potentially able to significantly reduce computational cost compared to ES-MDA when applied to large domains with many cells. Furthermore, even for very challenging cases with AdaPBS struggling to reach the

threshold diversity τ , the algorithm will still terminate after a predefined maximum number of iterations $\ell_{\max}$ which can be allocated depending on the computational budget at hand. Even in these challenging edge cases we have found that although the particle diversity and resulting uncertainty quantification may be worse than desired, the AdaPBS method will by construction still typically provide more accurate and reliable inference than the PBS.

In Table 2 we compare the KLD distance measures for the marginal posterior distributions obtained for temperature and precipitation parameters using the respective algorithms. This shows that although AdaPBS is not capable of reaching precisely the same low KLD values as the ES-MDA, it shows quite a comparable performance to this state-of-the art iterative ensemble Kalman method. In fact, the quantitative differences appear minimal when the distributions are compared graphically in Figure 6. More generally, it is clear that the iterative methods, namely the AdaPBS and ES-MDA, are much closer to being able to match the performance of MCMC than their non-iterative counterparts, namely the PBS and ES. Crucially, the AdaPBS can adapt to the complexity of the problem at hand and can thus be guaranteed to incur a lower computational or in the worst case equal cost to the ES-MDA by setting $\ell_{\max} = N_a + 1$. This cost difference is highlighted in the cost column of Table 2, where in the best case with a less complex target posterior the AdaPBS has the chance to trigger early stopping already after a single iteration ($L = 1$) and computationally affordable as the PBS on which it is based. Moreover, like the PBS, the AdaPBS can handle more general probabilistic models than the Gaussian models assumed by the ES-MDA and other ensemble Kalman methods.

Table 2. Performance of cryospheric DA algorithms in terms of inferential accuracy and cost. Accuracy is gauged in terms of how well the posterior approximations q from the algorithms match the reference posterior estimate p from RAM as measured by the Gaussian approximation of the marginal reverse KLD $D_{\text{KL}}(q||p)$ in (39) for air temperature bias b and the precipitation correction c in the temperature index model experiments. Computational cost is measured in terms of the number of forward model runs as multiples of the ensemble size $N_e = 100$, ES-MDA iterations $N_a = 4$, AdaPBS iterations $L \leq \ell_{\max}$ where typically $\ell_{\max} = N_a + 1$, and the number of Markov chain steps in RAM $N_s = 20 \times 10^3$

Algorithm	Temperature $D_{\text{KL}}(q p)$	Precipitation $D_{\text{KL}}(q p)$	Cost
Prior	625.80	52893.24	N_e
PBS	85.20	186.69	N_e
ES	878.32	8418.42	$2N_e$
IES (ES-MDA)	3.60	27.66	$(N_a + 1)N_e$
AdaPBS	5.59	47.31	$LN_e \leq \ell_{\max}N_e$
RAM ($q = p$)	0	0	$N_s \gg N_e$

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 compared to that of AdaPBS. At least in the present example, these differences are nonetheless relatively minor, especially in the model 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

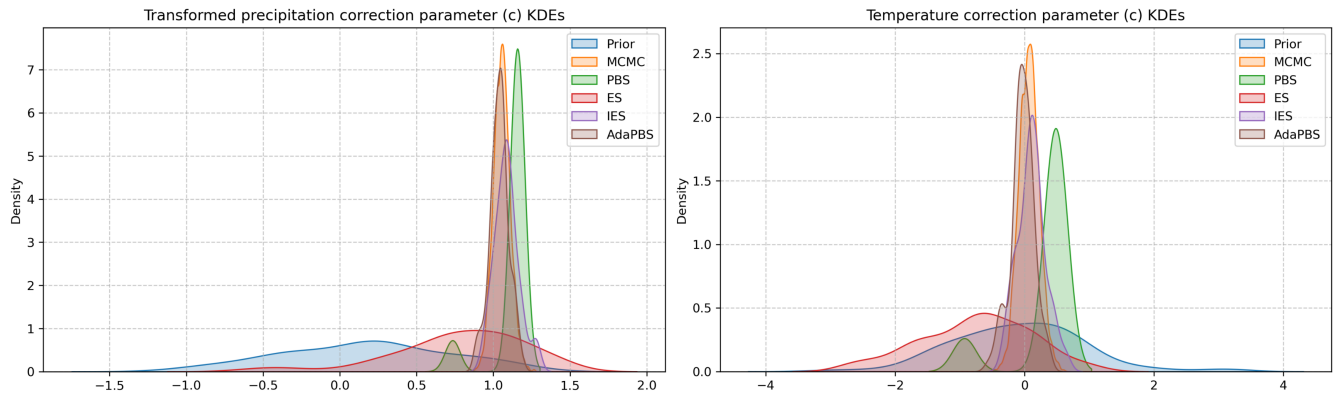


Figure 6. Comparison of Gaussian kernel density estimates (KDEs) obtained via `scipy.stats.gaussian_kde` (Scott’s rule bandwidth; Virtanen et al., 2020) of the marginal prior and marginal approximate posterior distributions of the parameters from the respective algorithms in the transformed (unbounded) space. The MCMC posterior (orange) from RAM is considered the gold-standard.

890 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 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
895 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.

Despite the many advantages of the AdaPBS scheme that we have highlighted, the ES-MDA still retains a key advantage:
900 There are more localization methods developed for ensemble Kalman-based algorithms (Evensen et al., 2022) than for those based on particle methods (Farchi and Bocquet, 2018). This facilitates the development of spatiotemporal assimilation initiatives (Alonso-González et al., 2023; Mazzolini et al., 2025; Alonso-González et al., 2026), as opposed to the purely temporal example presented here, which allows non-local observations in remote cells of a distributed model domain to be considered to update the local cell in question. In this way, information can be propagated in space, correcting areas of the domain even when
905 they have no local observations, or integrating point-scale observations into distributed simulations. A future line of research with great potential will be the development of these spatio-temporal methods for AdaPBS and other particle-based methods, which remains at the frontier of current DA research. Moreover, in conjunction with the localization problem, it remains to be seen if adaptive particle methods can be scaled up to as high-dimensional problems as ensemble Kalman methods (Carrassi et al., 2018; Pirk et al., 2024). Nonetheless, the results from our inter-comparison of cryospheric DA schemes with a tempera-

ture index model show that these adaptive particle methods are likely to perform favorably in local cryospheric DA problems with low dimensional parameter spaces. In the ensuing experiments with FSM2, we push the model complexity, number of observations, and the dimensionality of the parameter space considerably to further test the AdaPBS.

4.2 Applying AdaPBS in FSM2 at SnowMIP sites

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 related to the the parallelization scheme of the computing infrastructure, the hyperparameters (i.e. number of iterations for ES-MDA and 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 N_{eff} threshold is selected, at the cost of increasing the computational 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.

The results of AdaPBS in this more challenging setting with higher-dimensional parameter and observation spaces are similar to those obtained by ES-MDA, as shown in Table 3. It is difficult to conclude confidently which of the two schemes works best, so we conclude tentatively that their performance is similar even in scenarios as complex as the one proposed. The biggest differences are found at SNB, both in terms of RMSE (ES-MDA: 0.21, AdaPBS: 0.44) and CRPS (ES-MDA: 0.15, AdaPBS: 0.29), despite both methods presenting a very similar bias for this site. While the AdaPBS has a tendency to either match or perform slightly worse than ES-MDA in terms of RMSE and CRPS, it generally exhibits a lower bias than ES-MDA particularly for SWA (ES-MDA: -0.16, AdaPBS: -0.06). However, the fact that there are notable improvements with AdaPBS compared to ES-MDA in only one of the sites (SWA), worse performance at one site (SNB), and nearly identical performance at the remaining sites, is clearly not sufficient grounds to claim that there is considerable differences in the performance of these schemes. The results in Table 3 also show that AdaPBS consistently outperforms PBS in this experiment for all metrics across all sites, with the exception of SOD where the low bias differs by only 0.01. Overall, when considering metrics averaged over

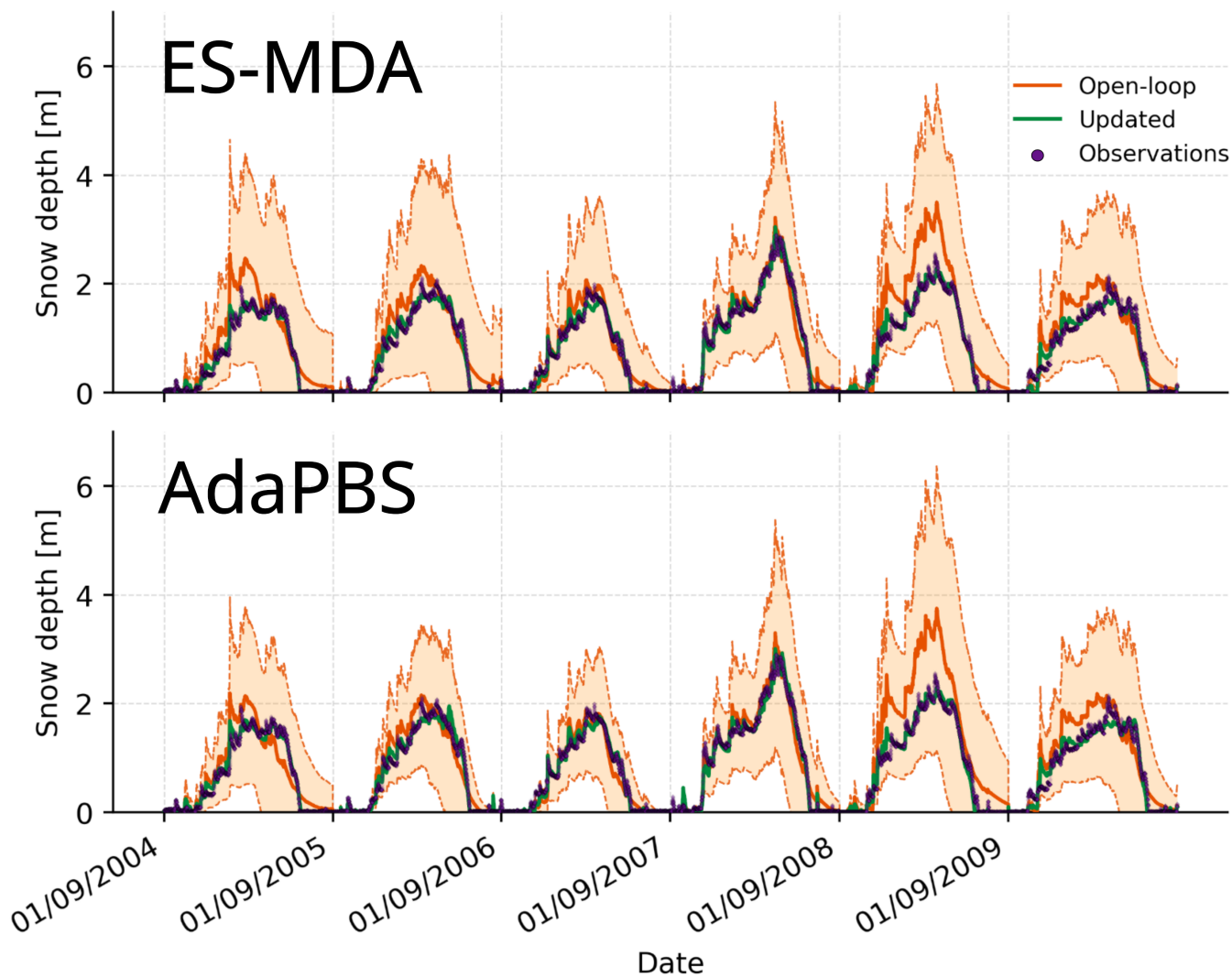


Figure 7. Comparison of the performance of the ESA-MDA (top) and the AdaPBS (bottom) at the Weissfluhjoch (WFJ) ESM-SnowMIP site in the eastern Swiss Alps. The red and green lines with corresponding shading show the prior ('open-loop' in red) and posterior ('updated' in green) mean ± 1 standard deviation along with the hourly assimilated snow depth observations (purple dots).

Table 3. Snow depth evaluation metrics in the form of the root mean square error (RMSE), mean continuous ranked probability score (CRPS), and mean error (bias), for the Prior, PBS, IES (ES-MDA), and AdaPBS across 6 ESM-SnowMIP sites.

ESM-SnowMIP Site (Location)	RMSE				Mean CRPS				Bias			
	Prior	PBS	IES	AdaPBS	Prior	PBS	IES	AdaPBS	Prior	PBS	IES	AdaPBS
RME (Idaho, USA)	0.31	0.22	0.10	0.11	0.29	0.18	0.08	0.10	0.01	-0.06	0.01	0.01
SAP (Hokkaido, Japan)	0.22	0.11	0.07	0.07	0.14	0.14	0.04	0.05	-0.16	-0.03	-0.01	0.00
SNB (Colorado, USA)	0.65	0.48	0.21	0.44	0.36	0.34	0.15	0.29	-0.10	-0.12	-0.07	-0.09
SOD (Lapland, Finland)	0.11	0.07	0.06	0.06	0.08	0.08	0.04	0.05	-0.07	0.00	0.00	0.01
SWA (Colorado, USA)	0.48	0.28	0.28	0.27	0.35	0.33	0.20	0.21	-0.27	-0.09	-0.16	-0.06
WFJ (Swiss Alps)	0.33	0.25	0.10	0.11	0.28	0.27	0.08	0.09	0.15	-0.01	-0.01	-0.01
Mean over sites	0.35	0.24	0.14	0.18	0.25	0.22	0.10	0.13	-0.07	-0.05	-0.04	-0.02

all sites, use of AdaPBS leads to a 25%, 41%, and 60% reduction in RMSE, CRPS, and bias, respectively, compared to PBS.

945 A likely reason for the large gains from adaptation is that the non-iterative PBS struggles to capture the dense observations, as demonstrated by the average of eight iterations needed in AdaPBS. As a result, PBS often degenerates completely to a suboptimal solution. Although not explicitly evaluated here, a single ES update is expected to also lead to a suboptimal solution due to the combined effects of nonlinearity and the highly informative nature of the observations. This expectation is supported by Experiment 1 as well as previous experiments comparing the ES-MDA to the ES in snow data assimilation (Aalstad et al., 950 2018; Alonso-González et al., 2022) and other applications (Pirk et al., 2022, 2024; Keetz et al., 2025). It is thus encouraging to see the AdaPBS, a purely particle-based iterative batch smoother, be able to match the performance of the ES-MDA given that AdaPBS provides added flexibility in terms of adapting to the problem at hand through early stopping and is readily extended to non-Gaussian likelihoods (e.g. Smith et al., 2010). Figure 7 demonstrates the generally very good performance of both the ES-MDA and AdaPBS algorithms at WFJ in terms of matching the temporally dense snow depth observations. The 955 same holds for most of the other ESM-SnowMIP sites visualized in the supplement (e.g. Figure S1). However, suboptimal posterior predictions are obtained in several water years, especially with AdaPBS but also ES-MDA, at the SNB which turns out to be the most challenging among all the sites (see Figure S3). Nonetheless, it should be noted that no explicit ERA5 forcing downscaling has been performed other than implicitly by the assimilation itself by inferring the forcing perturbation parameters. The resolution of ERA5 may not be sufficient to capture all the particularities of local observations even after the 960 implicit downscaling at least with the current probabilistic model configuration.

4.3 Recommendations for applying and extending the AdaPBS

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 965 $0 < \tau < 1$ and the maximum number of iterations $\ell_{\max} \geq 1$ (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 $\ell_{\text{max}} N_e$ 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
970 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 $\ell_{\text{max}} = 1$ 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 $\ell_{\text{max}} = 5$, 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
975 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}
980 to 10 or more. In practice, it is often entirely acceptable to end up with a low percentage (say $\leq 10\%$) 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 ($N_{\text{eff}} = 1$). Since the AdaPBS is designed as a direct extension of the PBS, instances of failed convergence will correspond to cases when
985 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., $N_{\text{eff}} = 30$ with $N_e = 100$) with $N_{\text{eff}} = 84$ 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
990 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
995 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 $\ell_{\text{max}} = 8$, and the diversity
1000 threshold to $\tau = 0.3$. In this two-dimensional parameter space, across all grid cells, convergence was achieved within an aver-

age 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 $\ell_{\max}$ 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.

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 Papaspiliopoulos, 2020). 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., 2026), 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.

5 Conclusions

We introduced the AdaPBS algorithm, an iterative and adaptive particle-based data assimilation scheme that combines a Particle Batch Smoother (Margulis et al., 2015) with Adaptive Multiple Importance Sampling (Cornuet et al., 2012). This formulation improves the resilience against ensemble collapse typically associated with particle-based methods while avoiding the Gaussian linear assumptions associated with ensemble Kalman-based methods. It also allows for the implementation of early stopping strategies, adapting the computational effort automatically to the problem complexity at the local grid-cell and water-year level in distributed multi-year simulations with potentially large cost reductions.

In a simple temperature index snow model assimilating snow depth observations, AdaPBS agreed closely with a costly RAM-based MCMC gold-standard benchmark, consistently outperforming or at least matching commonly used particle and ensemble Kalman-based batch smoothing methods. In more complex experiments assimilating hourly snow depth observations at SnowMIP sites into ensembles generated with the FSM2 model, AdaPBS managed to sample from the posterior simulations in a challenging high dimensional space of 19 uncertain parameters with a similar performance to ES-MDA. All experiments were developed with the open source MuSA toolbox (Alonso-González et al., 2022), in which both the proposed AdaPBS scheme and the RAM MCMC algorithm used for benchmarking are now available. This should facilitate the use and adaptation of AdaPBS to the plethora of data assimilation problems related to the terrestrial cryosphere (Giroto et al., 2020; Westermann et al., 2023; Rounce et al., 2023) and adjacent fields (Pirk et al., 2022; Keetz et al., 2025). Contemporaneous and promising implementations of AdaPBS already exist in the CryoGrid community model (Westermann et al., 2023) as detailed in Willmes et al. (2025) and the coupled PyGEM-OGGM global glacier models (Rounce et al., 2023; Maussion et al., 2019) as detailed in Yang et al. (2025). We encourage cryospheric researchers to help test, develop, refine, and remix data assimilation methods such as AdaPBS so that we as a community can continue to evolve in symbiosis with rapid advances in Bayesian methods (e.g. Chopin and Papaspiliopoulos, 2020; Evensen et al., 2022; Murphy, 2023).

Code and data availability. The latest release of MuSA (v2.3) containing the AdaPBS in conjunction with this paper is open source and can be found at Alonso-González and Aalstad (2025). The code and data needed to reproduce all the results and figures in this study are openly available through Alonso González and Aalstad (2026). The original Izas input data needed to run the first set of experiments in this study can be found at Alonso-González (2022). The original SnowMIP input data for the second set of experiments can be found at Menard and Essery (2019). Future versions of MuSA will continue to be submitted to <https://github.com/ealonsogzl/MuSA>.

1070 *Author contributions.* Conceptualization: KA and EAG. Data curation: EAG. Formal analysis: EAG and KA. Funding acquisition: EAG, KA. Investigation: EAG and KA. Methodology: KA and EAG with contributions from NP, CW, SW, RY. Resources: EAG. Software: EAG and KA. Validation: EAG and KA. Visualization: EAG. Writing (original draft preparation): KA and EAG. Writing (review and editing): KA, EAG, NP, CW, SW, RY.

Competing interests. The contact author has declared that none of the authors has any competing interests.

1075 *Financial support.* Kristoffer Aalstad acknowledges funding from an ESA-CCI Research Fellowship (PATCHES project) and the ERC-2022-ADG under grant agreement No 101096057 GLACMASS. Esteban Alonso-González acknowledges funding from an ESA-CCI Research Fellowship (SnowHotspots project) and the “Ramon y Cajal” Fellowship RYC2023-044416-I. Norbert Pirk acknowledges funding from the European Research Council (ACTIVATE project #101116083). Ruitang Yang acknowledges funding from the Research Council of Norway (GLACMOD project #324131). This work is a contribution to the strategic research initiative LATICE (#UiO/GEO103920), the Center for Biogeochemistry in the Anthropocene, as well as the Center for Computational and Data Science at the University of Oslo.

1080 *Acknowledgements.* We are grateful to the European Centre for Medium-Range Weather Forecasts (ECMWF) for openly providing global atmospheric reanalysis data. In particular, all experiments carried out in this study were forced using ERA5 reanalysis data (Hersbach et al., 2020) obtained from the Copernicus Climate Change Service (C3S) Climate Data Store. These data were generated using modified Copernicus Climate Change Service information. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. All simulations were carried out using the MuSA toolbox (Alonso-González et al., 2022) which is built on open source Python libraries and the Fortran-based snow model FSM2 (Essery et al., 2025). We are also
1085 grateful to ESM-SnowMIP effort for making snow depth data readily available from several snow reference sites through Ménard et al. (2019). This research was made possible through the access granted by the Galician Supercomputing Center (CESGA) to its supercomputing infrastructure. The supercomputer FinisTerra III and its permanent data storage system have been funded by the NextGeneration EU 2021 Recovery, Transformation and Resilience Plan, ICT2021-006904, and also from the Pluriregional Operational Programme of Spain 2014-2020 of the European Regional Development Fund (ERDF), ICTS-2019-02-CESGA-3, and from the State Programme for the Promotion
1090 of Scientific and Technical Research of Excellence of the State Plan for Scientific and Technical Research and Innovation 2013-2016 State subprogramme for scientific and technical infrastructures and equipment of ERDF, CESG15-DE-3114.

References

- Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., and Bertino, L.: Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites, *The Cryosphere*, 12, 247–270, <https://doi.org/10.5194/tc-12-247-2018>, 2018.
- 1095 Alonso-González, E., López-Moreno, J., Gascoin, S., García-Valdecasas Ojeda, M., Sanmiguel-Valladolid, A., Navarro-Serrano, F., Revuelto, J., Ceballos, A., Esteban-Parra, M., and Essery, R.: Daily gridded datasets of snow depth and snow water equivalent for the Iberian Peninsula from 1980 to 2014, *Earth Sys. Sci. Data*, 10, 303–315, <https://doi.org/10.5194/essd-10-303-2018>, 2018.
- Alonso-González, E., Gutmann, E., Aalstad, K., Fayad, A., Bouchet, M., and Gascoin, S.: 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, <https://doi.org/10.5194/hess-25-4455-2021>, 2021.
- 1100 Alonso-González, E., Aalstad, K., Baba, M. W., Revuelto, J., López-Moreno, J. I., Fiddes, J., Essery, R., and Gascoin, S.: The Multiple Snow Data Assimilation System (MuSA v1.0), *Geosci. Model Dev.*, <https://doi.org/10.5194/gmd-15-9127-2022>, 2022.
- Alonso-González, E., Aalstad, K., Pirk, N., Mazzolini, M., Treichler, D., Leclercq, P., Westermann, S., López-Moreno, J. I., and Gascoin, S.: Spatio-temporal information propagation using sparse observations in hyper-resolution ensemble-based snow data assimilation, *Hydrol. Earth Syst. Sci.*, 27, 4637–4659, <https://doi.org/10.5194/hess-27-4637-2023>, 2023.
- 1105 Alonso-González, E., Harpold, A., Lundquist, J. D., Piske, C., Sourp, L., Aalstad, K., and Gascoin, S.: Ensemble-based data assimilation improves hyperresolution snowpack simulations in forests, *The Cryosphere*, 20, 209–225, <https://doi.org/10.5194/tc-20-209-2026>, 2026.
- Alonso-González, E.: Inputs (forcing and observations) ready for use by 'MuSA: The Multiscale Snow Data Assimilation System (v1.0)', <https://doi.org/10.5281/zenodo.7248635>, 2022.
- 1110 Alonso-González, E. and Aalstad, K.: MuSA: v2.3 AdaPBS submission, <https://doi.org/10.5281/zenodo.17292981>, [Computer software]. Zenodo, 2025.
- Alonso González, E. and Aalstad, K.: Code and data to reproduce the results and figures in : Evolving beyond collapse: An adaptive particle batch smoother for cryospheric data assimilation , <https://doi.org/10.5281/zenodo.21244337>, [Dataset]. Zenodo, 2026.
- Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 143, 607–633, <https://doi.org/10.1002/qj.2982>, 2017.
- 1115 Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature*, 438, 303–309, <https://doi.org/10.1038/nature04141>, 2005.
- Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, *International Statistical Review*, 71, 223–241, <https://doi.org/10.1111/j.1751-5823.2003.tb00194.x>, 2003.
- 1120 Beven, K.: A manifesto for the equifinality thesis, *Journal of Hydrology*, 320, 18–36, <https://doi.org/10.1016/j.jhydrol.2005.07.007>, 2006.
- Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, <https://doi.org/10.1002/hyp.3360060305>, 1992.
- Bugallo, M. F., Elvira, V., Martino, L., Luengo, D., Miguez, J., and Djuric, P. M.: Adaptive Importance Sampling: The past, the present, and the future, *IEEE Signal Processing Magazine*, 34, 60–79, <https://doi.org/10.1109/MSP.2017.2699226>, 2017.
- 1125 Campelo, F. and Aranha, C.: Lessons from the Evolutionary Computation Bestiary, *Artificial Life*, 29, 421–432, https://doi.org/10.1162/artl_a_00402, 2023.
- Cao, W., Aalstad, K., Schmidt, L., Westermann, S., and Schuler, T.: Glacier data assimilation on an Arctic glacier: Learning from large ensemble twin experiments, *Journal of Glaciology*, <https://doi.org/10.48550/arXiv.2502.09314>, preprint in review, 2025.

- Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, *WIREs Climate Change*, 9, e535, <https://doi.org/10.1002/wcc.535>, 2018.
- Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G.: On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model, *The Cryosphere*, 10, 1021–1038, <https://doi.org/10.5194/tc-10-1021-2016>, 2016.
- Chopin, N.: A sequential particle filter method for static models, *Biometrika*, 89, 539–552, <https://doi.org/10.1093/biomet/89.3.539>, 2002.
- Chopin, N. and Papaspiliopoulos, O.: *An Introduction to Sequential Monte Carlo*, Springer, <https://doi.org/10.1007/978-3-030-47845-2>, 2020.
- Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., and Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models, *Advances in Water Resources*, 29, 1209–1221, <https://doi.org/10.1016/j.advwatres.2005.10.001>, 2006.
- Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., and Stuart, A.: Calibrate, emulate, sample, *Journal of Computational Physics*, 424, 109 716, <https://doi.org/10.1016/j.jcp.2020.109716>, 2021.
- Cluzet, B., Lafaysse, M., Cosme, E., Albergel, C., Meunier, L.-F., and Dumont, M.: CrocO_v1, *Geosci. Model Dev.*, 14, 1595–1614, <https://doi.org/10.5194/gmd-14-1595-2021>, 2021.
- Cluzet, B., Magnusson, J., Quéno, L., Mazzotti, G., Mott, R., and Jonas, T.: Exploring how Sentinel-1 wet-snow maps can inform fully distributed physically based snowpack models, *The Cryosphere*, 18, 5753–5767, <https://doi.org/10.5194/tc-18-5753-2024>, 2024.
- Cornuet, J.-M., Marin, J.-M., Antonietta, M., and Robert, C.: Adaptive Multiple Importance Sampling, *Scandinavian Journal of Statistics*, 39, 798–812, <https://doi.org/10.1111/j.1467-9469.2011.00756.x>, 2012.
- Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis: El Niño Effects on Andes Snow, *Geophysical Research Letters*, 44, 6859–6867, <https://doi.org/10.1002/2017GL073826>, 2017.
- De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., and Pauwels, V. R. N.: Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resources Research*, 48, W01 522, <https://doi.org/10.1029/2011WR010588>, 2012.
- De Lannoy, G. J. M., Bechtold, M., Busschaert, L., Heyvaert, Z., Modanesi, S., Dunmire, D., Lievens, H., Getirana, A., and Mas-sari, C.: Contributions of Irrigation Modeling, Soil Moisture and Snow Data Assimilation to High-Resolution Water Budget Estimates Over the Po Basin: Progress Towards Digital Replicas, *Journal of Advances in Modeling Earth Systems*, 16, e2024MS004 433, <https://doi.org/10.1029/2024MS004433>, 2024.
- Del Moral, P.: *Feynman-Kac Formulae*, Springer, <https://doi.org/10.1007/978-1-4684-9393-1>, 2004.
- Doucet, A., Godsill, S., and Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*, 10, 197–208, <https://doi.org/10.1023/A:1008935410038>, 2000.
- Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world’s mountains, *Wiley Interdisciplinary Reviews: Water*, 3, 461–474, <https://doi.org/10.1002/wat2.1140>, 2016.
- Duan, J.-C., Li, S., and Xu, Y.: Sequential Monte Carlo optimization and statistical inference, *WIREs Computational Statistics*, 15, e1598, <https://doi.org/10.1002/wics.1598>, 2023.
- Durand, M., Molotch, N. P., and Margulis, S. A.: A Bayesian approach to snow water equivalent reconstruction, *Journal of Geophysical Research: Atmospheres*, 113, D20 117, <https://doi.org/10.1029/2008JD009894>, 2008.

- Elias Chereque, A., Kushner, P. J., Mudryk, L., Derksen, C., and Mortimer, C.: A simple snow temperature index model exposes discrepancies between reanalysis snow water equivalent products, *The Cryosphere*, 18, 4955–4969, <https://doi.org/10.5194/tc-18-4955-2024>, 2024.
- Elvira, V., Martino, L., Bugallo, M. F., and Djuric, P. M.: Elucidating the Auxiliary Particle Filter via Multiple Importance Sampling [Lecture Notes], *IEEE Signal Processing Magazine*, 36, 145–152, <https://doi.org/10.1109/MSP.2019.2938026>, 2019.
- Elvira, V., Martino, L., and Robert, C.: Rethinking the Effective Sample Size, *International Statistical Review*, 90, 525–550, <https://doi.org/10.1111/insr.12500>, 2022.
- Emerick, A. A. and Reynolds, A. C.: Combining the Ensemble Kalman Filter with Markov Chain Monte Carlo for Improved History Matching and Uncertainty Characterization, <https://doi.org/10.2118/141336-MS>, 2011.
- Emerick, A. A. and Reynolds, A. C.: Ensemble smoother with multiple data assimilation, *Computers & Geosciences*, 55, 3–15, <https://doi.org/10.1016/j.cageo.2012.03.011>, 2013.
- Essery, R.: A factorial snowpack model (FSM 1, *Geosci. Model Dev.*, 8, 3867–3876, <https://doi.org/10.5194/gmd-8-3867-2015>, 2015.
- Essery, R., Mazzotti, G., Barr, S., Jonas, T., Quaife, T., and Rutter, N.: A Flexible Snow Model (FSM 2.1.1) including a forest canopy, *GMD*, 18, 3583–3605, <https://doi.org/10.5194/gmd-18-3583-2025>, 2025.
- Euskirchen, E. S., Goodstein, E. S., and Huntington, H. P.: An estimated cost of lost climate regulation services caused by thawing of the Arctic cryosphere, *Ecological Applications*, 23, 1869–1880, <https://doi.org/10.1890/11-0858.1>, 2013.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, 99, 10 143, <https://doi.org/10.1029/94JC00572>, 1994.
- Evensen, G., Vossepoel, F. C., and van Leeuwen, P. J.: *Data Assimilation Fundamentals*, Springer, <https://doi.org/10.1007/978-3-030-96709-3>, 2022.
- Farchi, A. and Bocquet, M.: Review article: Comparison of local particle filters and new implementations, *Nonlinear Processes in Geophysics*, 25, 765–807, <https://doi.org/10.5194/npg-25-765-2018>, 2018.
- Fiddes, J., Aalstad, K., and Westermann, S.: Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering, *Hydrol. Earth Syst. Sci.*, 23, 4717–4736, <https://doi.org/10.5194/hess-23-4717-2019>, 2019.
- Garbuno-Inigo, A., Hoffmann, F., Li, W., and Stuart, A.: Interacting Langevin Diffusions: Gradient Structure and Ensemble Kalman Sampler, *SIAM Journal on Applied Dynamical Systems*, 19, 412–441, <https://doi.org/10.1137/19M1251655>, 2020.
- Gascoin, S.: A call for an accurate presentation of glaciers as water resources, *WIREs Water*, 11, e1705, <https://doi.org/10.1002/wat2.1705>, 2024.
- Gascoin, S., Luojus, K., Nagler, T., Lievens, H., Masiokas, M., Jonas, T., Zheng, Z., and De Rosnay, P.: Remote sensing of mountain snow from space: status and recommendations, *Frontiers in Earth Science*, 12, 1381 323, 2024.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D.: *Bayesian Data Analysis*, CRC Press, 3 edn., 2013.
- Gilks, W. R. and Berzuini, C.: Following a moving target-Monte Carlo inference for dynamic Bayesian models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 127–146, <https://doi.org/10.1111/1467-9868.00280>, 2001.
- Giroto, M., Margulis, S., and Durand, M.: Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques, *Hydrological Processes*, 28, 3875–3895, <https://doi.org/10.1002/hyp.9887>, 2014.
- Giroto, M., Musselman, K. N., and Essery, R. L. H.: Data Assimilation Improves Estimates of Climate-Sensitive Seasonal Snow, *Current Climate Change Reports*, 6, 81–94, <https://doi.org/10.1007/s40641-020-00159-7>, 2020.

- Giroto, M., Formetta, G., Azimi, S., Bachand, C., Cowherd, M., De Lannoy, G., Lievens, H., Modanesi, S., Raleigh, M. S., Rigon, R., and Massari, C.: Identifying snowfall elevation patterns by assimilating satellite-based snow depth retrievals, *Science of The Total Environment*, 906, 167 312, <https://doi.org/10.1016/j.scitotenv.2023.167312>, 2024.
- Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association*, 102, 359–378, <https://doi.org/10.1198/016214506000001437>, 2007.
- Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T.: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation, *Monthly Weather Review*, 133, 1098–1118, 2005.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F Radar and Signal Processing*, 140, 107, <https://doi.org/10.1049/ip-f-2.1993.0015>, 1993.
- Gottlieb, A. R. and Mankin, J. S.: Evidence of human influence on Northern Hemisphere snow loss, *Nature*, 625, 293–300, <https://doi.org/10.1038/s41586-023-06794-y>, publisher: Nature Publishing Group, 2024.
- Groenke, B., Langer, M., Nitzbon, J., Westermann, S., Gallego, G., and Boike, J.: Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer, *The Cryosphere*, 17, 3505–3533, <https://doi.org/10.5194/tc-17-3505-2023>, 2023.
- Günther, D., Marke, T., Essery, R., and Strasser, U.: 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, <https://doi.org/10.1029/2018WR023403>, 2019.
- Hammersley, J. and Morton, K.: Poor Man’s Monte Carlo, *Journal of the Royal Statistical Society: Series B (Methodological)*, 16, 23–38, <https://doi.org/10.1111/j.2517-6161.1954.tb00145.x>, 1954.
- Hastings, W.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 57, <https://doi.org/10.2307/2334940>, 1970.
- Hennig, P., Osborne, M., and Kersting, H.: Probabilistic Numerics, Cambridge, <https://doi.org/10.1017/9781316681411>, 2022.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hock, R.: Temperature index melt modelling in mountain areas, *Journal of Hydrology*, 282, 104–115, [https://doi.org/10.1016/S0022-1694\(03\)00257-9](https://doi.org/10.1016/S0022-1694(03)00257-9), 2003.
- Hock, R. et al.: High Mountain Areas, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by Pörtner, H.-O. et al., pp. 131–202, Cambridge, <https://doi.org/10.1017/9781009157964.004>, 2019.
- Holland, J.: *Adaptation in Natural and Artificial Systems*, MIT, 2 edn., ISBN 9780262581110, 1992.
- Immerzeel, W., Lutz, A., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B., Elmore, A., Emmer, A., Feng, M., Fernánde, A., Haritashya, U., Kargel, J., Koppes, M., Kraaijenbrink, P., Kulkarni, A., Mayewski, P., Nepal, S., Pacheco, P., Painter, T., Pellicciotti, F. Rajaram, H., Rupper, S., Sinisalo, A., Srestha, A., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J.: Importance and vulnerability of the world’s water towers, *Nature*, 577, 364–369, <https://doi.org/10.1038/s41586-019-1822-y>, 2020.
- Jaynes, E.: *Probability Theory*, Cambridge, <https://doi.org/10.1017/CBO9780511790423>, 2003.
- Jazwinski, A.: *Stochastic Processes and Filtering Theory*, Academic Press, 1970.
- Keetz, L. T., Aalstad, K., Fisher, R. A., Poppe Terán, C., Naz, B., Pirk, N., Yilmaz, Y. A., and Skarpaas, O.: Inferring Parameters in a Complex Land Surface Model by Combining Data Assimilation and Machine Learning, *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004 542, <https://doi.org/10.1029/2024MS004542>, 2025.

- 1240 Kitagawa, G.: Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, *Journal of Computational and Graphical Statistics*, 5, 1–25, <https://doi.org/10.1080/10618600.1996.10474692>, 1996.
- Koblents, E. and Míguez, J.: A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models, *Statistics and Computing*, 25, 407–425, <https://doi.org/10.1007/s11222-013-9440-2>, 2015.
- Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B.,
1245 Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chérut, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V., Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li, W., Marke, T., Marks, D., Ménégoz, M., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T. G., Stacke, T., Strasser, U., Svenson, S., Turkov, D., Wang, T., Wever, N., Yuan, H., Zhou, W., and Zhu, D.: ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, *Geosci. Model Dev.*, 11, 5027–5049,
1250 <https://doi.org/10.5194/gmd-11-5027-2018>, 2018.
- Landmann, J. M., Künsch, H. R., Huss, M., Ogier, C., Kalisch, M., and Farinotti, D.: Assimilating near-real-time mass balance stake readings into a model ensemble using a particle filter, *The Cryosphere*, 15, 5017–5040, <https://doi.org/10.5194/tc-15-5017-2021>, 2021.
- Largerion, C., Dumont, M., Morin, S., Boone, A., Lafaysse, M., Metref, S., Cosme, E., Jonas, T., Wisntral, A., and Margulis, S.: Toward Snow Cover Estimation in Mountainous Areas Using Modern Data Assimilation Methods: A Review, *Frontiers in Earth Science*, 8, 325,
1255 <https://doi.org/10.3389/feart.2020.00325>, 2020.
- Law, K. and Stuart, A.: Evaluating Data Assimilation Algorithms, *Monthly Weather Review*, 140, 3757–3782, <https://doi.org/10.1175/MWR-D-11-00257.1>, 2012.
- Leisenring, M. and Moradkhani, H.: Snow water equivalent prediction using Bayesian data assimilation methods, *Stochastic Environmental Research and Risk Assessment*, 25, 253–270, <https://doi.org/10.1007/s00477-010-0445-5>, 2011.
- 1260 Li, T., Bolic, M., and Djuric, P.: Resampling Methods for Particle Filtering: Classification, implementation, and strategies, *IEEE Signal Processing Magazine*, 32, 70–86, <https://doi.org/10.1109/MSP.2014.2330626>, 2015.
- Liston, G. E. and Elder, K.: A meteorological distribution system for high-resolution terrestrial modeling (MicroMet, *Journal of Hydrometeorology*, 7, 217–234, <https://doi.org/10.1175/JHM486.1>, 2006b.
- Liu, Y., Fang, Y., and Margulis, S. A.: Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year
1265 Landsat–MODIS era snow reanalysis dataset, *The Cryosphere*, 15, 5261–5280, <https://doi.org/10.5194/tc-15-5261-2021>, 2021.
- Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, *Earth Sys. Sci. Data*, 10, 235–249, <https://doi.org/10.5194/essd-10-235-2018>, 2018.
- MacKay, D. J. C.: *Information Theory, Inference, and Learning Algorithms*, Cambridge, 2003.
- Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., and Jonas, T.: Improving physically based snow simulations by assimilating snow
1270 depths using the particle filter, *Water Resources Research*, 53, 1125–1143, <https://doi.org/10.1002/2016WR019092>, 2017.
- Margulis, S. A., Giroto, M., Cortés, G., and Durand, M.: A particle batch smoother approach to snow water equivalent estimation, *Journal of Hydrometeorology*, 16, 1752–1772, <https://doi.org/10.1175/JHM-D-14-0177.1>, 2015.
- Margulis, S. A., Cortés, G., Giroto, M., and Durand, M.: A landsat-era Sierra Nevada snow reanalysis (1985–2015, *Journal of Hydrometeorology*, 17, 1203–1221, <https://doi.org/10.1175/JHM-D-15-0177.1>, 2016.
- 1275 Marin, J.-M., Pudlo, P., and Sedki, M.: Consistency of adaptive importance sampling and recycling schemes, *Bernoulli*, 25, 1977–1998, <https://doi.org/10.3150/18-BEJ1042>, 2019.

- Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, *Geoscientific Model Development*, 12, 909–931, <https://doi.org/10.5194/gmd-12-909-2019>, 2019.
- 1280 Mazzolini, M., Aalstad, K., Alonso-González, E., Westermann, S., and Treichler, D.: Spatio-temporal snow data assimilation with the ICESat-2 laser altimeter, *The Cryosphere*, 19, 3831–3848, <https://doi.org/10.5194/tc-19-3831-2025>, 2025.
- Menard, C. and Essery, R.: ESM-SnowMIP meteorological and evaluation datasets at ten reference sites (in situ and bias corrected reanalysis data), <https://doi.org/10.1594/PANGAEA.897575>, supplement to: Menard, Cecile; Essery, Richard; Barr, Alan; Bartlett, Paul; Derry, Jeff; Dumont, Marie; Fierz, Charles; Kim, Hyungjun; Kontu, Anna; Lejeune, Yves; Marks, Danny; Niwano, Masashi; Raleigh, Mark; Wang, Libo; Wever, Nander (2019): Meteorological and evaluation datasets for snow modelling at 10 reference sites: description of in situ and bias-corrected reanalysis data. *Earth Sys. Sci. Data*, 11(2), 865–880, <https://doi.org/10.5194/essd-11-865-2019>, 2019.
- 1285 Ménard, C. B., Essery, R., Barr, A., Bartlett, P., Derry, J., Dumont, M., Fierz, C., Kim, H., Kontu, A., Lejeune, Y., Marks, D., Niwano, M., Raleigh, M., Wang, L., and Wever, N.: Meteorological and evaluation datasets for snow modelling at 10 reference sites: description of in situ and bias-corrected reanalysis data, *Earth Sys. Sci. Data*, 11, 865–880, <https://doi.org/10.5194/essd-11-865-2019>, 2019.
- 1290 Meredith, M. et al.: Polar Regions, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by Pörtner, H.-O. et al., p. 203–320, Cambridge, <https://doi.org/10.1017/9781009157964.004>, 2019.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equation of State Calculations by Fast Computing Machines, *J. Chem. Phys.*, 26, 1087–1092, <https://doi.org/10.1063/1.1699114>, 1953.
- Morzfeld, M., Hodyss, D., and Snyder, C.: What the collapse of the ensemble Kalman filter tells us about particle filters, *Tellus A: Dynamic Meteorology and Oceanography*, 69, 1–14, <https://doi.org/10.1080/16000870.2017.1283809>, 2017.
- 1295 Murphy, K.: Probabilistic Machine Learning: Advanced Topics, MIT, <https://probml.github.io/book2>, (last access: 18.12.2023), 2023.
- Navari, M., Margulis, S. A., Bateni, S. M., Tedesco, M., Alexander, P., and Fettweis, X.: Feasibility of improving a priori regional climate model estimates of Greenland ice sheet surface mass loss through assimilation of measured ice surface temperatures, *The Cryosphere*, 10, 103–120, <https://doi.org/10.5194/tc-10-103-2016>, 2016.
- 1300 Neal, R.: Annealed importance sampling, *Statistics and Computing*, 11, 125–139, <https://doi.org/10.1023/A:1008923215028>, 2001.
- Neal, R.: MCMC Using Hamiltonian Dynamics, in: *Handbook of Markov Chain Monte Carlo*, edited by Brooks, S. et al., chap. 5, pp. 113–162, CRC Press, <https://doi.org/10.1201/b10905>, 2011.
- Oberrauch, M., Cluzet, B., Magnusson, J., and Jonas, T.: Improving Fully Distributed Snowpack Simulations by Mapping Perturbations of Meteorological Forcings Inferred From Particle Filter Assimilation of Snow Monitoring Data, *Water Resources Research*, 60, e2023WR036994, <https://doi.org/10.1029/2023WR036994>, 2024.
- 1305 Owen, A. and Zhou, Y.: Safe and Effective Importance Sampling, *Journal of the American Statistical Association*, 95, 135–143, <https://doi.org/10.1080/01621459.2000.10473909>, 2000.
- Piazzì, G., Thirel, G., Campo, L., and Gabellani, S.: A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment, *The Cryosphere*, 12, 2287–2306, <https://doi.org/10.5194/tc-12-2287-2018>, 2018.
- 1310 Pirk, N., Aalstad, K., Westermann, S., Vatne, A., van Hove, A., Tallaksen, L., Cassiani, M., and Katul, G.: Inferring surface energy fluxes using drone data assimilation in large eddy simulations, *Atmospheric Measurement Techniques*, 15, 7293–7314, <https://doi.org/10.5194/amt-15-7293-2022>, 2022.

- Pirk, N., Aalstad, K., Mannerfelt, E. S., Clayer, F., de Wit, H., Christiansen, C. T., Althuisen, I., Lee, H., and Westermann, S.: Disaggregating the Carbon Exchange of Degrading Permafrost Peatlands Using Bayesian Deep Learning, *Geophysical Research Letters*, 51, e2024GL109283, <https://doi.org/https://doi.org/10.1029/2024GL109283>, e2024GL109283 2024GL109283, 2024.
- 1315 Rainforth, T., Golinski, A., Wood, F., and Zaidi, S.: Target–Aware Bayesian Inference: How to Beat Optimal Conventional Estimators, *JMLR*, 21, 1–54, 2020.
- Reich, S. and Cotter, C.: *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge, <https://doi.org/10.1017/CBO9781107706804>, 2015.
- 1320 Revuelto, J., Azorin-Molina, C., Alonso-González, E., Sanmiguel-Valladolid, A., Navarro-Serrano, F., Rico, I., and Ignacio López-Moreno, J.: Meteorological and snow distribution data in the Izas Experimental Catchment (Spanish Pyrenees) from 2011 to 2017, *Earth Sys. Sci. Data*, 9, 993–1005, <https://doi.org/10.5194/essd-9-993-2017>, 2017.
- Revuelto, J., Alonso-Gonzalez, E., Vidaller-Gayan, I., Lacroix, E., Izagirre, E., Rodríguez-López, G., and López-Moreno, J.: Intercomparison of UAV platforms for mapping snow depth distribution in complex alpine terrain, *Cold Regions Science and Technology*, 190, 103344, <https://doi.org/10.1016/j.coldregions.2021.103344>, 2021.
- 1325 Riihelä, A., Bright, R. M., and Anttila, K.: Recent strengthening of snow and ice albedo feedback driven by Antarctic sea-ice loss, *Nature Geoscience*, 14, 832–836, <https://doi.org/10.1038/s41561-021-00841-x>, 2021.
- Robert, C.: *The Bayesian Choice*, Springer, 2 edn., <https://doi.org/10.1007/0-387-71599-1>, 2007.
- Robert, C. and Casella, G.: *Monte Carlo Statistical Methods*, Springer, <https://doi.org/10.1007/978-1-4757-4145-2>, 2004.
- 1330 Rounce, D. R., Hock, R., Maussion, F., Hugonnet, R., Kochtitzky, W., Huss, M., Berthier, E., Brinkerhoff, D., Compagno, L., Copland, L., Farinotti, D., Menounos, B., and McNabb, R. W.: Global glacier change in the 21st century: Every increase in temperature matters, *Science*, 379, 78–83, <https://doi.org/10.1126/science.abo1324>, 2023.
- Sanz-Alonso, D., Stuart, A. M., and Taeb, A.: *Inverse problems and data assimilation*, vol. 107, Cambridge, <https://doi.org/10.1017/9781009414319>, 2023.
- 1335 Särkkä, S. and Svensson, L.: *Bayesian Filtering and Smoothing*, Cambridge, 2 edn., <https://doi.org/10.1017/9781108917407>, 2023.
- Smith, A. F. M. and Gelfand, A. E.: Bayesian Statistics without Tears: A Sampling–Resampling Perspective, *The American Statistician*, 46, 84–88, <https://doi.org/10.1080/00031305.1992.10475856>, 1992.
- Smith, T., Sharma, A., Marshall, L., Mehrotra, R., and Sisson, S.: Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resources Research*, 46, <https://doi.org/10.1029/2010WR009514>, 2010.
- 1340 Smyth, E. J., Raleigh, M. S., and Small, E. E.: Particle Filter Data Assimilation of Monthly Snow Depth Observations Improves Estimation of Snow Density and SWE, *Water Resources Research*, 55, 1296–1311, <https://doi.org/10.1029/2018WR023400>, 2019.
- Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to High-Dimensional Particle Filtering, *Monthly Weather Review*, 136, 4629 – 4640, <https://doi.org/10.1175/2008MWR2529.1>, 2008.
- Stordal, A. S. and Elsheikh, A. H.: Iterative ensemble smoothers in the annealed importance sampling framework, *Advances in Water Resources*, 86, 231–239, <https://doi.org/10.1016/j.advwatres.2015.09.030>, 2015.
- 1345 Sun, H., Fang, Y., Margulis, S. A., Mortimer, C., Mudryk, L., and Derksen, C.: Evaluation of the Snow Climate Change Initiative (Snow CCI) snow-covered area product within a mountain snow water equivalent reanalysis, *The Cryosphere*, 19, 2017–2036, <https://doi.org/10.5194/tc-19-2017-2025>, 2025.
- Sörensen, K.: Metaheuristics—the metaphor exposed, *International Transactions in Operational Research*, 22, 3–18, <https://doi.org/10.1111/itor.12001>, 2015.
- 1350

- Tang, B., Frye, H., Gelfand, A., and Silander, J. A.: Zero-Inflated Beta Distribution Regression Modeling, *Journal of Agricultural, Biological and Environmental Statistics*, 28, 117–137, 2023.
- van Hove, A., Aalstad, K., Lind, V., Arndt, C., Odongo, V., Ceriani, R., Fava, F., Hulth, J., and Pirk, N.: Inferring methane emissions from African livestock by fusing drone, tower, and satellite data, *Biogeosciences*, 22, 4163–4186, <https://doi.org/10.5194/bg-22-4163-2025>, 2025.
- van Hove, A., Aalstad, K., and Pirk, N.: Actively inferring methane sources with drones, *Environmental Data Science*, 5, e2, <https://doi.org/10.1017/eds.2026.10029>, 2026.
- van Leeuwen, P. J.: Particle Filtering in Geophysical Systems, *Monthly Weather Review*, 137, 4089–4114, <https://doi.org/10.1175/2009MWR2835.1>, 2009.
- van Leeuwen, P. J. and Evensen, G.: Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation, *Monthly Weather Review*, 124, 2898–2913, [https://doi.org/10.1175/1520-0493\(1996\)124<2898:DAAIMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2), 1996.
- van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S.: Particle filters for high-dimensional geoscience applications: A review, *Quarterly Journal of the Royal Meteorological Society*, 145, 2335–2365, <https://doi.org/10.1002/qj.3551>, 2019.
- Vihola, M.: Robust adaptive Metropolis algorithm with coerced acceptance rate, *Statistics and Computing*, 22, 997–1008, <https://doi.org/10.1007/s11222-011-9269-5>, 2012.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y.: SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nature Methods*, 17, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, 39, <https://doi.org/10.1029/2002WR001642>, 2003.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, 44, W00B09, <https://doi.org/10.1029/2007WR006720>, 2008.
- Westermann, S., Ingeman-Nielsen, T., Scheer, J., Aalstad, K., Aga, J., Chaudhary, N., Etzelmüller, B., Filhol, S., Käab, A., Renette, C., Schmidt, L. S., Schuler, T. V., Zweigel, R. B., Martin, L., Morard, S., Ben-Asher, M., Angelopoulos, M., Boike, J., Groenke, B., Miesner, F., Nitzbon, J., Overduin, P., Stuenzi, S. M., and Langer, M.: The CryoGrid community model (version 1.0) – a multi-physics toolbox for climate-driven simulations in the terrestrial cryosphere, *Geosci. Model Dev.*, 16, 2607–2647, <https://doi.org/10.5194/gmd-16-2607-2023>, 2023.

- Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, *Physica D: Nonlinear Phenomena*, 230, 1–16, <https://doi.org/10.1016/j.physd.2006.09.017>, 2007.
- Willmes, C., Aalstad, K., and Westermann, S.: Assimilating high-resolution satellite snow cover data in a permafrost model, *EGUsphere*, <https://doi.org/10.5194/egusphere-2025-3142>, 2025.
- Yang, R., Ultee, L., Aalstad, K., Debolsky, M., Hock, R., Schmitt, P., and Li, T.: Calibrating Coupled Frontal Ablation and Surface Mass Balance in Global Glacier Models Using an Adaptive Particle Smoother, In prep, 2025.
- 1395 Zschenderlein, L., Luojus, K., Takala, M., Venäläinen, P., and Pulliainen, J.: Evaluation of passive microwave dry snow detection algorithms and application to SWE retrieval during seasonal snow accumulation, *Remote Sensing of Environment*, 288, 113476, <https://doi.org/https://doi.org/10.1016/j.rse.2023.113476>, 2023.