Journal of Glaciology

Bayesian data assimilation on an Arctic glacier: learning from large ensemble twin experiments

Wenxue Cao, Kristoffer Aalstad, Louise Steffensen Schmidt, Sebastian Westermann and Thomas V. Schuler

Department of Geosciences, University of Oslo, Oslo, Norway

Article

Cite this article: Cao W, Aalstad K, Schmidt LS, Westermann S, Schuler TV (2025) Bayesian data assimilation on an Arctic glacier: learning from large ensemble twin experiments. *Journal of Glaciology* **71**, e121, 1–17. https://doi.org/10.1017/jog.2025.10101

Received: 7 February 2025
Revised: 1 October 2025
Accepted: 22 October 2025

Keywords:
Arctic glaciology; glaciological model experiments; glacier mass balance

Corresponding author: Wenxue Cao;
Email: wenxue.cao@geo.uio.no

Abstract

Numerical modeling is crucial for quantifying the evolution of cryospheric processes. At the same time, uncertainties hamper process understanding and predictive accuracy. Here, we suggest improving glacier surface mass balance simulations for the Kongsvegen glacier in Svalbard through the application of Bayesian data assimilation techniques in a set of large ensemble twin experiments. Noisy synthetic observations of albedo and snow depth, generated using the multilayer CryoGrid community model with a full energy balance, are assimilated using two ensemble-based data assimilation schemes: the particle batch smoother and the ensemble smoother. A comprehensive evaluation exercise demonstrates that the joint assimilation of albedo and snow depth improves the simulation skill by up to 86% relative to the prior in specific glacier regions. The particle batch smoother excels in representing albedo dynamics, while the ensemble smoother is marginally more effective for snow depth under low snowfall conditions in the ablation area. By combining the strengths of both observations, the joint assimilation achieves improved surface mass balance simulations across different glacier zones using either assimilation scheme. This work underscores the potential of ensemble-based data assimilation methods for refining glacier models by offering a robust framework to enhance predictive accuracy and reduce uncertainties in cryospheric simulations. Further advances in glacier data assimilation research with both synthetic and real observations will be critical to better understanding the fate and role of Arctic glaciers in a changing climate

1. Introduction

Glaciers are regarded as key indicators of climate change. Over the past three decades, global glacier ice loss has contributed nearly 1 mm annually to sea level rise (Zemp and others, 2019; IPCC, 2022). At the same time, glaciers serve as a critical component of mountain water towers, helping to provide a more consistent and reliable water supply to downstream regions (Immerzeel and others, 2019; Zhang and others, 2023). Arctic glaciers are experiencing an accelerated mass loss (Østby and others, 2017; van Pelt and others, 2019; Rounce and others, 2023; Schmidt and others, 2023) because warming is amplified in the Arctic at two to four times the global average through various positive feedback mechanisms (e.g. Lind and others, 2018; Rantanen and others, 2022). Freshwater runoff from melting Arctic glaciers can have considerable impacts on ocean circulation and ocean-atmosphere interaction globally (Devilliers and others, 2024; Pontes and Menviel, 2024; Schiller-Weiss and others, 2024; Malles and others, 2025), as well as on regional marine biogeochemistry and productivity (Hopwood and others, 2020; Ezat and others, 2024). Thus, accurate knowledge of glacier surface mass balance is vital for understanding, detecting, and predicting the impacts of climate change.

Numerical modeling is the main method to reconstruct past or project future glacier surface mass balance at sites with scarce in situ observations. Glacier surface mass balance models forced by meteorological data include temperature-index models (e.g. Hock, 2003; Marzeion and others, 2012) and energy balance models (e.g. Hock and Holmgren, 2005; Westermann and others, 2023). Temperature-index models approximate the melt rate based on air temperature (Huss and Hock, 2015), while physically-based energy balance models explicitly calculate the energy fluxes on the glacier surface and therefore provide a more detailed representation of the processes controlling the surface mass balance. The accuracy of glacier models is limited by uncertainties related to meteorological forcing (Marzeion and others, 2020), incomplete model physics (Schmidt and others, 2023), and parameter uncertainty (Rounce and others, 2020; Schuster and others, 2023). Constraining each uncertainty source remains a significant challenge.

© The Author(s), 2025. Published by Cambridge University Press on behalf of International Glaciological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

Data assimilation methods can incorporate observations into modeling to improve accuracy and constrain simulation uncertainty (Evensen and others, 2022, Chapter 2). In situ and remotely sensed observations can be individually or jointly assimilated into glacier models, leading to a reduction of the aforementioned uncertainties (Gillet-Chaulet, 2020; Choi and others, 2023). The assimilation of ground-based glaciological measurements into surface mass balance models is gradually becoming a recognized approach for updating glacier model parameters or initial states (Landmann and others, 2021; Sjursen and others, 2023). Despite this recognition, there are still relatively few studies that have implemented data assimilation in surface mass balance modeling. Moreover, in situ measurements are available for only a minority of glaciers worldwide, which presents a significant challenge to transfer information to the unmeasured majority of glaciers. In addition to the direct assimilation of surface mass balance measurements, other quantities that can directly influence surface mass balance changes, such as remotely sensed albedo or snow depth, can also be ingested within a data assimilation framework. Satellite-derived observations of albedo and snow depth offer greater spatial and temporal coverage, enabling broader applicability despite their susceptibility to factors such as cloud cover, sensor resolution, and atmospheric interference, which can impact data quality and availability (Vionnet and others, 2012; Østby and others, 2014; Deschamps-Berger and others, 2023).

Albedo, defined as the reflectivity of the Earth's surface to shortwave insolation, is a controlling variable in the surface energy balance of glaciers. It significantly impacts the radiation budget, thereby influencing the rate of melt and overall surface mass balance of glaciers (e.g. Budyko, 1969; Ye and others, 2024). In a pioneering study, a variational assimilation scheme was used to incorporate Moderate Resolution Imaging Spectroradiometer (MODIS) derived albedo into a snowpack model to reconstruct the spatial surface mass balance distribution for an Alpine glacier (Dumont and others, 2012). More recently, Sentinel-2 albedo estimates were assimilated into a glacio-hydrological model to improve the simulation of streamflow in two glacierized basins in the Canadian Rockies (Bertoncini and others, 2024).

Snowfall is another major driver of the surface mass balance, as it is the primary source of glacier mass gain (Hock, 2003; Pramanik and others, 2019). Satellite-based snow depth retrievals, such as from the ICESat-2 laser altimeter, are a potentially globally available constraint on uncertainties in snowfall forcing which is being explored for seasonal snow data assimilation (Mazzolini and others, 2024). However, to the best of our knowledge, no experiment has explored the joint assimilation of remotely sensed albedo and snow depth into an energy balance model for surface mass balance simulation. Moreover, the current state of the art in using Bayesian data assimilation to infer surface mass balance has focused on static parameters in temperature index models using relatively costly Markov chain Monte Carlo methods (Rounce and others, 2020; Sjursen and others, 2023). This stands in contrast to other recent cryospheric work on glacier flow (Brinkerhoff, 2022), ice sheet (Navari and others, 2000–2014), seasonal snow (Alonso-González and others, 2022), and permafrost (Groenke and others, 2023), which employ a greater diversity of modern Bayesian data assimilation (also known as inversion) schemes that allow for the use of more complex models.

In this study, we performed twin experiments (Arnold and Dey, 1986; Masutani and others, 2010), also known as synthetic experiments or Observing System Simulation Experiments, to explore the benefits of assimilating albedo and snow depth on surface mass balance simulations. This allowed us to test the data assimilation workflow in a series of targeted experiments while avoiding challenges of real observations and model discrepancies (Masutani and others, 2010). In particular, as satellite-based measurements of albedo and snow depth and their associated error characteristics are not always available or consistent due to variable weather conditions and observational limitations. Here, synthetic observations are instead generated using synthetic truth (also known as nature) runs of the energy balance model CryoGrid (Schmidt and others, 2023; Westermann and others, 2023). It should be noted from the title of our study that it revolves entirely around conducting twin experiments using synthetic truth runs and noisy synthetic observations thereof. For brevity, we will henceforth often simply refer to these as truth and observations. These synthetic observations serve as idealized representations of satellite measurements, with their spatiotemporal resolution designed to mirror that of actual satellite data, forming the foundation for observing system simulation. While these synthetic observations currently facilitate controlled testing of our methodology, the ultimate objective is to assimilate actual satellite observations in future applications. To assimilate the synthetic observations, we employed and compared two Bayesian data assimilation schemes, namely the Particle Batch Smoother (PBS; Margulis and others, 2015) and the Ensemble Smoother (ES; van Leeuwen and Evensen, 1996). The synthetic observations were derived from synthetic truth runs for four distinct scenarios, each representing different climatic conditions. These four scenarios were selected to better capture the varying information content of the assimilated observations. They are represented by four different climatic conditions generated by four different parameter settings (Table 1). The simulations were driven by reanalysis data from the Copernicus Arctic Regional Reanalysis (CARRA) dataset over 12 hydrological years from September 2010 to September 2022. Kongsvegen glacier, one of the best studied glaciers in High Arctic Svalbard, was selected as the study area due to the availability of data and its extensive size, encompassing diverse glacier zones that offer a comprehensive basis for representing a broad range of Arctic glaciers. By conducting a large number of twin experiments, we compared the effectiveness of the particle-based PBS scheme to the ensemble Kalman-based ES in improving simulated glacier surface mass balance across different glacier zones and climatic scenarios. The sensitivity of the surface mass balance estimates to ensemble size was also tested in terms of both accuracy and precision via bootstrap resampling experiments using the PBS at essentially no extra cost since (unlike the ES) no costly model

Table 1. Truth parameter value settings used to generate the synthetic truth for each scenario . The selection of these 'extreme' parameter values was based on the values corresponding to three standard deviations away from the prior mean after applying a logit transformation to the respective parameter ensembles

Truth parameter	Albedo evolution rate τ_a	Snowfall factor β_s
Rapid albedo evolution rate and high snowfall factor	0.03	1.7
Slow albedo evolution rate and low snowfall factor	0.0003	0.7
Rapid albedo evolution rate and low snowfall factor	0.03	0.7
Slow albedo evolution rate and high snowfall factor	0.0003	1.7

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

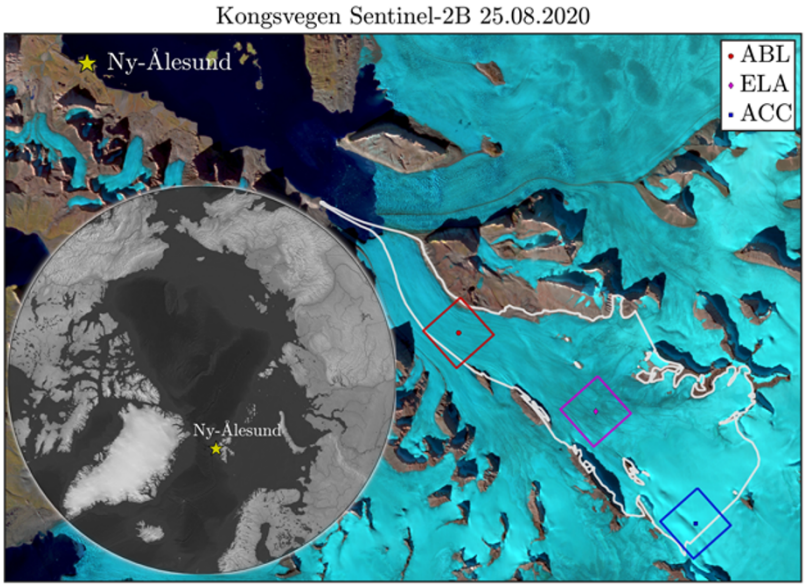


Figure 1. Atmospherically corrected shortwave infrared false color image over the area surrounding Kongsvegen glacier near Ny-Ålesund in the Svalbard archipelago captured by the Sentinel-2B satellite at 13:07 UTC on the 25th of August 2020. The image shows the locations of Ny-Ålesund (yellow star) and the Kongsvegen glacier outline from RGI 7.0 (white) as well as the locations of 2.5 by 2.5km grid cells that were extracted from CARRA to represent the ablation zone (ABL, red), Equilibrium Line Altitude (ELA, purple), and the accumulation zone (ACC, blue) of Kongsvegen. The inset shows the location of Ny-Ålesund (yellow star) in the Arctic (here roughly defined as latitudes above 60°N) on a polar stereographic map using open Gray Earth data from Natural Earth.

reruns were required in this PBS sensitivity analysis. Synthetic observations of albedo and snow depth with synthetic truth surface mass balance enabled us to perform a robust evaluation of a novel glacier data assimilation framework in CryoGrid through a large ensemble of twin experiments that allowed us to control for both observation and model error.

2. Data and methods

2.1. Study site

The Svalbard archipelago is one of the most climatically sensitive regions in the world (Noël and others, 2020; Geyman and others, 2022). For example, it is the region in Europe that has experienced the greatest warming in the past three decades (Nordli and others, 2014; Isaksen and others, 2016). Kongsvegen is a marine-terminating glacier, located on the northwestern coast of Svalbard close to the research station of Ny-Ålesund (Fig. 1). The glacier has an area of around 100 km^2 and a length of 26 km, with slopes ranging from 0.5 to 2.5° (Karner and others, 2013). The ice flows towards the northwest from its ice-divide at about 800 m a.s.l. down to sea level at the head of Kongsfjorden (Hagen and others, 1999; Karner and others, 2013). The three grid cells used in this study are shown in Fig. 1 and we used these grids to represent different glacier zones, namely the ablation area, equilibrium line altitude (ELA), and accumulation area.

2.2. Forcing data

This study uses the Copernicus Arctic Regional Reanalysis (CARRA) dataset (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2024) as meteorological forcing data. The CARRA forcing fields considered are the 2 m air temperature, 2 m specific humidity, 10 m windspeed, incoming longwave and shortwave radiation, precipitation, and atmospheric pressure. CARRA is derived from the HARMONIE-AROME numerical weather prediction system (Bengtsson and others, 2017). This regional reanalysis covers two domains in the European sector of the Arctic, CARRA-West and CARRA-East, employing ERA5 reanalysis as boundary conditions (Yang and others, 2021). The CARRA output has a horizontal resolution of 2.5km and a 3-hour temporal resolution covering the period from 1991 to present. Following Schmidt and others (2023), this study employs meteorological data from the CARRA-East domain over 12 hydrological years from the 16th of September 2010 to the 15th of September 2022.

2.3. Surface mass balance model

2.3.1. Model description

The CryoGrid community model is an open-source model developed for climate-driven multiphysics simulations of the terrestrial cryosphere (Westermann and others, 2023), which uses a full surface energy-balance scheme that can be coupled to different multilayer subsurface modules of varying complexity. We used the glacier surface mass balance configuration of CryoGrid with a snow and firn module that was first employed by Schmidt and others (2023). The model calculates the full energy balance at the surface E_s:

$$E_s = (1-\alpha)S_{\text{in}} + L_{\text{in}} - L_{\text{out}} - Q_h - Q_e, \quad (1)$$

where S_{in} is the incoming shortwave radiation, α is the surface albedo, L_{in} and L_{out} is the incoming and outgoing longwave radiation, Q_h is the sensible heat flux and Q_e is the latent heat flux. The surface albedo of snow is calculated in three spectral bands following Vionnet and others (2012). In the UV and visible range (0.3-0.8 μm), the albedo is calculated as:

$$\alpha = \max(0.6, \alpha_i - \tau_a A), \quad (2)$$

where α_i is the albedo contribution from snow microstructure, represented by the optical diameter, τ_a is the albedo decay rate, and A is the snow age in days. The albedo in the near-infrared bands ([0.8–1.5] and [1.5–2.8] μm) only depend on snow microstructure. The albedo of ice is given a constant value of 0.4.

The outgoing longwave radiation is calculated as

$$L_{\text{out}} = \epsilon\sigma T^4 - (1-\epsilon)L_{\text{in}}, \quad (3)$$

where ϵ is the surface emissivity, σ is the Stefan Boltzmann constant and T is the 2m air temperature in Kelvin.

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

Finally, the turbulent fluxes are calculated using the Monin-Obukhov similarity theory (Monin and Obukhov, 1954), and depend on the surface pressure and wind speed. The sensible heat flux is further calculated using the temperature gradient between the 2m temperature and the surface, while the latent heat flux is calculated from the gradient in specific humidity.

If the energy balance is positive $E_s > 0$, the excess energy can either be used to warm up or melt the glacier surface. If the surface consists of pure ice, the melt will immediately run off. If a layer of snow is present, the water will percolate down the snowpack and, depending on the snow conditions, either be retained, refrozen, or run off.

While glacier ablation is calculated through the energy balance, the accumulation is taken directly from the CARRA reanalysis. The precipitation is divided into rainfall and snowfall using a temperature threshold. CryoGrid has the option of a relative bias correction for snowfall and rainfall using multiplicative factors β_s and β_r, respectively.

2.3.2. Uncertain parameters

In this study, we added ensemble-based data assimilation methods to the glacier surface mass balance simulations, thereby creating a comprehensive probabilistic modeling package. Through data assimilation, we aim to improve simulated glacier surface mass balance. The most important physical parameters which are currently tunable within the CryoGrid model are given in Table 2.

In numerical weather forecasting and climate modeling, precipitation, particularly snowfall, is associated with significant uncertainties that can contribute to considerable errors in Arctic surface mass balance models (e.g. Forbes and others, 2011; Schmidt and others, 2017; Van Pelt and others, 2019; Lenaerts and others, 2020). Albedo is a controlling variable of the surface energy balance, and therefore accurately simulating albedo is important for modeling the surface mass balance (e.g Schmidt and others, 2017; Gunnarsson and others, 2023). We therefore determine that the parameters in Table 2 which control these variables (β_s and τ_a) are the most uncertain and will have the largest impact on the surface mass balance simulations. Our ensemble assimilation approach thus involved constraining these uncertain parameters, to enhance the accuracy of the simulated snowfall, the snow albedo evolution, and the associated surface mass balance.

2.3.3. Model initialization

We initialized the model using a 5-year spin-up from 2006 to 2010, using a snow albedo evolution rate of $\tau_a = 0.005$ day^{-1} and no bias correction of the snowfall, i.e. $\beta_s = 1$. This allows the near-surface ice temperatures to respond to the model forcing and the buildup of a small firn layer of 3 m w.e. in the accumulation zone and improves the physical consistency of the subsequent experiments, particularly in the accumulation area. The primary objective of the spin-up phase is to achieve model relaxation, allowing the glacier system to reach a quasi-equilibrium state that aligns with realistic initial conditions consistent with climatological data and available observations. This ensures that subsequent simulations start from a physically consistent baseline, minimizing artifacts from arbitrary initializations and better reflecting natural variability in snow and ice dynamics. Regarding the zonal differences, in the accumulation area, where perennial snow cover predominates, the spin-up is crucial for building a stable snowpack and ensuring that processes like firn densification and albedo evolution are initialized correctly, as these areas are less likely to complete melt-out.

2.4. Synthetic observations

In this study, synthetic albedo and snow depth observations were assimilated to constrain simulations of glacier surface mass balance using the CryoGrid model. We generated albedo and snow depth time series for each of the three grid cells, selected to represent the ablation area, ELA, and accumulation area of Kongsvegen. By prescribing true parameters and running the model for these three grid cells, we obtained a synthetic truth from which synthetic observations have been generated. Figure 2 shows the workflow in our experimental design. We primarily divided the workflow into three steps. First, the generation of synthetic truth data was achieved by prescribing 'true' parameters representing different conditions. To mimic realistic observational data that is inherently noisy, we added Gaussian noise to both the albedo and snow depth truth to represent observation error. For the albedo, this noise has a mean of 0 and an observation error standard deviation of $\sigma_\alpha = 0.1$. This standard deviation is chosen based on the upper limit values of the reported root mean square error, which is an average of 0.7, obtained by comparing MODIS albedo retrievals to in situ measurements (Stroeve and others, 2005). The noise added to snow depth has a mean of 0 m and a standard deviation of $\sigma_s = 0.5$ m, which is based on the findings of Deschamps-Berger and others (2023) for ICESat2 snow depth retrievals on low slopes that the median residual is up to 0.65m. The synthetic albedo observations were sampled based on the effective temporal resolution of MODIS onboard the Terra and Aqua satellites. We simulated the impact of polar night on the availability of optical albedo retrievals in our research area by removing the values of synthetic albedo during this time (November to February). In addition, the occluding impact of cloud cover is considered in this study. According to findings by Marshall and others (1993), statistically only 22% to 24% of days between April and September in Svalbard are classified as clear-sky conditions, making MODIS albedo products usable only for those days. Also, Østby and others (2014) found that only 26% of MODIS products are acquired under clear sky conditions on Austfonna, Svalbard. Thus, we use 30 daily albedo observations that were randomly distributed in time excluding polar night, representing approximately 20% of the total days of each year between mid of April to mid of October. In our analysis of snow depth data, we considered the temporal resolution provided by the ICESat-2 satellite in the Arctic region. Typically, ICESat-2 operates on a 91-day revisit cycle at the equator. However, due to its high-inclination orbit, the ground tracks of the satellite converge towards the poles, significantly enhancing the frequency of overpasses in polar regions. Consequently, in the Arctic, the temporal resolution increases, with revisit intervals reduced to approximately 1–2 weeks (Markus and others, 2017). This enhancement in revisit frequency was utilized to simulate the temporal resolution in our synthetic snow depth data, providing a more accurate representation of snow accumulation and change over time in this region.Table 2.

2.5. Data assimilation

In this section, we describe the data assimilation methods used and their implementation in CryoGrid to infer glacier surface mass balance. For a more comprehensive treatment of Bayesian data assimilation methods, we refer to the extensive work of Evensen and others (2022, Chapters 2 and 6) and Sanz-Alonso and others (2023) and the overview in Alonso-González and others (2022) for details pertinent to cryospheric applications. Data assimilation is

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

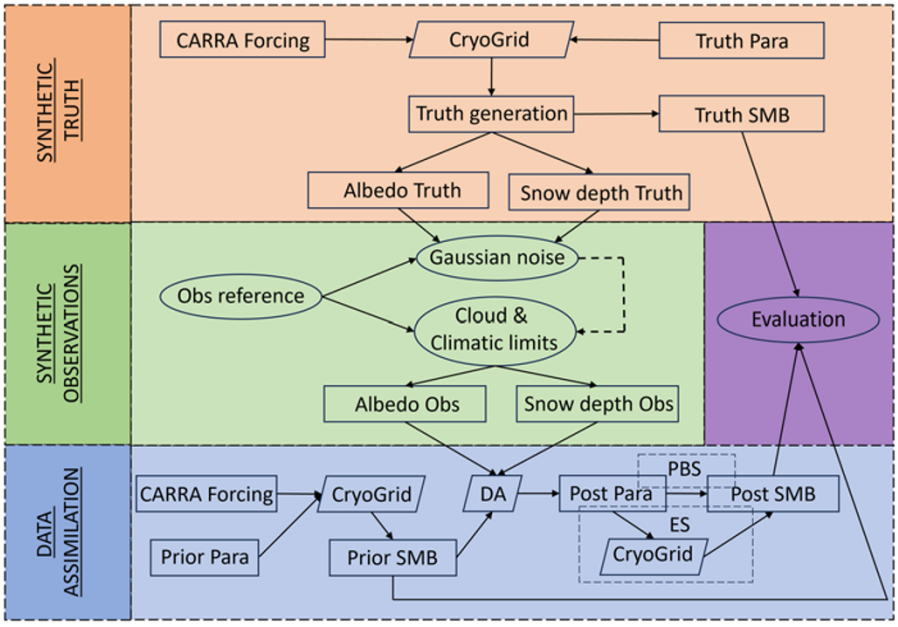


Figure 2. Workflow in the twin experiments involving the sequential generation of: synthetic truth runs (orange), noisy synthetic observations (green), and data assimilation experiments (blue) followed by the evaluation of each experiment (purple).

Table 2. Tuneable parameters within glacier and snow modules of the CryoGrid model

Parameter	Description	Default value	Reference
τ_a	Albedo decay rate	0.0033 day^{-1}	Vionnet and others (2012)
α_{ice}	Ice albedo	0.4	e.g. Schmidt and others (2023)
z_0^{snow}	Roughness length of snow	1 mm	e.g. Schmidt and others (2023)
z_0^{ice}	Roughness length of ice	1 mm	e.g. Schmidt and others (2023)
θ	Field capacity of snow/firn	0.05	e.g. Westermann and others (2023)
β_s	Snow factor	1	
β_r	Rain factor	1	

loosely defined as the fusion of data and models that can be mathematically formalized using the probabilistic framework of Bayesian inference.

There are multiple sources of model uncertainty related to the choice of model parameters, forcing, initial conditions, and model structure. Here, we are primarily concerned with the two first sources of uncertainty which we lump into an uncertain input parameter vector $\boldsymbol{\theta}$ with $N_p = 2$ elements, namely β_s and τ_a. By using the synthetic truth generated by the same underlying model, CryoGrid in this case, we avoid model structural uncertainty by confining ourselves to so-called identical twin experiments (Arnold and Dey, 1986). Within these identical twin experiments, we are thus by construction justified in restricting ourselves to solving the strong constraint data assimilation problem that assumes a perfect data generating model (CryoGrid) that can map perfectly onto reality if the true input vector $\boldsymbol{\theta}^\star$ were known (Evensen and others, 2022, Chapter 2). Note that the strong assumption has been widely and successfully adopted in atmospheric (Hersbach and others, 2020), land (Keetz and others, 2025), and cryospheric (Alonso-González and others, 2022) data assimilation even if the underlying perfect model assumption is always violated to some extent. Moreover, research in snow hydrology points to forcing, particularly precipitation, as being the dominant source of uncertainty (Günther and others, 2019; Tang and others, 2023). Translating this into glacier surface mass balance modeling helps further justifies our use of the strong constraint assumption by encoding the forcing uncertainty into the parameter vector $\boldsymbol{\theta}$ in line with common practice in snow data assimilation (Alonso-González and others, 2022).

In this strong constraint setting, we can model the observation vector $\mathbf{y}$ containing N_o noisy observations for a given temporal data assimilation window using the following data generating process:

$$\mathbf{y} = \mathcal{G}(\boldsymbol{\theta}^\star) + \boldsymbol{\epsilon}, \tag{4}$$

where $\mathcal{G}(\cdot)$ denotes the data generating model, $\boldsymbol{\theta}^\star$ is the aforementioned true parameter vector, and $\boldsymbol{\epsilon}$ is a noise term representing observation error. Given some observations, the task at hand is to invert $\mathcal{G}(\cdot)$ to recover $\boldsymbol{\theta}^\star$. This task is challenging since $\mathcal{G}(\cdot)$ is often a nonlinear and relatively computationally costly model instantiated in a long piece of typically non-differentiable code, namely CryoGrid in our case (Westermann and others, 2023). To complicate matters further, it is also a fundamentally ill-posed inverse problem since the solution is not unique due to the presence of irreducible noise in the form of observation error $\boldsymbol{\epsilon}$ (Sanz-Alonso and others, 2023). As such, we abandon the ill-conceived notion of a single optimal solution $\boldsymbol{\theta}^\star$ and instead adopt a probabilistic perspective where we seek a distribution over solutions $\boldsymbol{\theta}$ that are compatible with the noisy data $\mathbf{y}$ that we are given. Adopting a probabilistic perspective naturally leads to casting this ill-posed inverse problem in terms of Bayesian inference (Sanz-Alonso and others, 2023), and the computational challenge motivates the adoption of efficient ensemble-based data assimilation algorithms to make inference tractable (Evensen and others, 2022, Chapter 8). Formally the entire exercise of data assimilation can now be boiled down to using Bayes' rule as follows

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \tag{5}$$

to infer the posterior probability distribution $p(\boldsymbol{\theta}|\mathbf{y})$ over parameters $\boldsymbol{\theta}$ given data $\mathbf{y}$. The likelihood quantifies how well the model

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

predictions with parameters $\boldsymbol{\theta}$ fit the noisy observations $\mathbf{y}$, the prior regularizes the problem using background information about $\boldsymbol{\theta}$, and the evidence $p(\mathbf{y})$ is a normalizing constant (MacKay, 2003).

Once prior and likelihood are defined, Bayesian inference is theoretically straightforward and is just a matter of applying (5) to a grid of parameter vectors $\boldsymbol{\theta}$. Practical geophysical applications of Bayesian inference for data assimilation tend to require more efficient methods than computationally expensive grid approximations. The current state-of-the-art data assimilation approaches can generally be split into ensemble-based (Monte Carlo) and variational methods (Evensen and others, 2022, Chapter 2). The latter requires a differentiable model which is often, as is the case with this CryoGrid version (Westermann and others, 2023), not available. As such, we use ensemble-based data assimilation methods that are widely used in cryospheric applications (e.g. Navari and others, 2000–2014; Alonso-González and others, 2022; Groenke and others, 2023), but have not been widely applied to glacier surface mass balance modeling. In particular, we adopt both the PBS and the ES to compare their performance for a large ensemble of twin experiments. In addition to being used in the literature (Alonso-González and others, 2022), these methods are relatively straightforward to implement and can serve as kernels for more sophisticated schemes, such as the hybrid particle-adjusted iterative ensemble smoother (Pirk and others, 2022), which remain too costly for large ensemble twin experiments.

2.5.1. Prior and likelihood

In this study, we focus on two uncertain parameters within the glacier configuration of CryoGrid, namely the albedo evolution rate τ_a and the snowfall factor β_s. The former factor τ_a is an inverse timescale that controls the rate at which the visible albedo in the Crocus albedo parametrization decays (Vionnet and others, 2012), with larger (smaller) values indicating a faster (slower) decay rate. Here, we are not trying to be within the bounds given by Vionnet and others (2012), but rather to find a value that better represents the albedo evolution due to snow age for Svalbard. The latter multiplier β_s explicitly accounts for biases in the snowfall (solid precipitation) forcing from the CARRA reanalysis while also implicitly accounting for unresolved processes in this instantiation of CryoGrid in the form of wind-driven snow redistribution. Both parameters are treated as fixed (time-invariant) within a given mass balance year which is used as the data assimilation window. As such, the Bayesian inference step in (5) can be carried out independently for each such mass balance year window using the same prior $p(\boldsymbol{\theta})$ but different observations $\mathbf{y}$ and thus varying likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ resulting in a posterior $p(\boldsymbol{\theta}|\mathbf{y})$ that varies from (balance) year to year. In this study, the parameter vector $\boldsymbol{\theta}$ has $N_p = 2$ elements so we consider a 2D parameter space. On the one hand, this is quite a low-dimensional parameter space. On the other hand, CryoGrid which we use as the data generating model is relatively expensive to evaluate. Moreover, these $N_p = 2$ parameters were selected based on several modeling studies of surface mass balance where related parameters were deemed among the most uncertain yet sensitive parameters (e.g. Schmidt and others, 2017; 2023; Lenaerts and others, 2019; van Pelt and others, 2019; Raoult and others, 2023).

To encode uncertainty in these parameters we need to specify a prior distribution $p(\boldsymbol{\theta})$ that reflects our prior knowledge concerning possible values for these parameters. Herein, building on several related studies (Aalstad and others, 2018; Guidicelli and others, 2024; Mazzolini and others, 2024; Keetz and others, 2025), we use the generalized logit-normal prior distribution that is a double bounded transformed version of a normal distribution allowing for upper and lower bounds (a,b), a central location parameter μ_0^*, and a scale parameter σ_0 reflecting the spread in possible values. Following Keetz and others (2025), this prior is defined as follows for a scalar parameter θ

$$p(\theta|\mu_0, \sigma_0, a, b) = \frac{|J|}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{(\phi - \mu_0)^2}{2\sigma_0^2}\right), \tag{6}$$

where $|J| = (b-a)/(\theta-a)(b-\theta)$ is a Jacobian term and ϕ is the generalized logit transform of θ

$$\phi = \psi(\theta, a, b) = \ln\left(\frac{\theta - a}{b - a}\right) - \ln\left(\frac{b - \theta}{b - a}\right), \tag{7}$$

with corresponding inverse transform

$$\theta = \psi^{-1}(\phi, a, b) = a + \frac{b - a}{1 + \exp(-\phi)}. \tag{8}$$

The generalized logit normal distribution in (6) has an associated normal distribution, namely the distribution of the logit transformed parameter ϕ with mean μ_0 and standard deviation σ_0. We use the σ_0 parameter to define the scale (spread) of the logit-normal prior for a bounded parameter θ. For the location parameter of the logit normal we use the median of the logit normal distribution μ_0^* that can be transformed to the mean of the associated normal distribution through $\mu_0 = \psi(\mu_0^*, a, b)$ and vice versa.

The generalized logit prior in (6) is specified independently for each of the parameters τ_a and β_s using the hyperparameters in Table 3 resulting in a weakly informative prior (Banner and others, 2020). Assuming independence, the joint prior $p(\boldsymbol{\theta})$ is simply given by the product of the marginal priors $p(\boldsymbol{\theta}) = p(\tau_a)p(\beta_s)$. It is possible to relax this prior independence assumption by adding non-zero correlations to a multivariate logit-normal distribution (Mazzolini and others, 2024). More generally, adding dependence structure to the joint prior using background knowledge can improve inference (Pirk and others, 2022). At the same time, adding this kind of structure requires that such background knowledge is available, and here, we had no prior reasons to suspect a general dependence between the albedo aging factor and the snowfall multiplier. Crucially, prior independence does not imply posterior independence since via the likelihood the data allow us to infer the posterior dependence structure between parameters. To sample from the generalized logit normal distribution (6), we apply the generalized logit transform (7) to the prior median μ_0^* to obtain the mean of the associated normal μ_0 then add N_e samples of randomly generated Gaussian noise with standard deviation σ_0 to μ_0 and apply the inverse transform (8) to obtain prior samples $\theta_i \sim p(\theta)$ for the parameter $\theta \in \boldsymbol{\theta}$ (i.e., either τ_a or β_s) in question. After having done this for both parameters, we are left with an ensemble of N_e particles from the joint prior $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$.

As is commonly done in data assimilation (Carrassi and others, 2018), we use a simple additive zero-mean Gaussian observation error model of the form $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$ where $\mathbf{R}$ is an $N_o \times N_o$ observation error covariance matrix. This can be justified as a useful default first-order error model using both the central limit theorem and maximum entropy arguments (Jaynes, 2003). Using this error model, allows us to formulate the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$. By definition, this is the probability density of the (fixed) observations $\mathbf{y}$ given that the parameter set $\boldsymbol{\theta}$ is true. By inspection of (4) conditional on $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$ the observation error becomes $\boldsymbol{\epsilon} = \mathbf{y} - \mathcal{G}(\boldsymbol{\theta})$

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

Table 3. The hyperparameters for the independent logit-normal priors used for each of the $N_p = 2$ uncertain parameters in the parameter vector $\boldsymbol{\theta}$ considered in this study. The hyperparameters are the lower bound a, the upper bound b, the location parameter which is the median μ_0^*, and the scale (spread) which is the dimensionless standard deviation σ_0 of the associated normal distribution

Parameter name	Symbol	Units	Lower bound a	Upper bound b	Location μ_0^*	Scale σ_0
Albedo evolution rate	τ_a	day^{-1}	0.0001	0.05	0.005	1
Snowfall factor	β_s	–	0.5	2	1	1

and by inserting this into the Gaussian observation error model we obtain a Gaussian likelihood $p(\mathbf{y}|\boldsymbol{\theta}) = \mathrm{N}(\mathbf{y}|\mathcal{G}(\boldsymbol{\theta}), \mathbf{R})$ of the form

$$p(\mathbf{y}|\boldsymbol{\theta}) = c_y \exp\left(-\frac{1}{2}\left[\mathbf{y}-\hat{\mathbf{y}}\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}-\hat{\mathbf{y}}\right]\right), \tag{9}$$

where $c_y = \det(2\pi\mathbf{R})^{-1/2}$ is a constant and $\hat{\mathbf{y}} = \mathcal{G}(\boldsymbol{\theta})$ denotes the predicted observations from the data generating model given a particular parameter set $\boldsymbol{\theta}$. Following the likelihood principle in Bayesian inference (Jaynes, 2003), the likelihood should be viewed as a function of the uncertain parameters $\boldsymbol{\theta}$ (here through $\hat{\mathbf{y}} = \mathcal{G}(\boldsymbol{\theta})$) rather than a distribution over the fixed (albeit noisy) observations $\mathbf{y}$ that we are assimilating. Although the likelihood in (9) is Gaussian, our data generating model $\hat{\mathbf{y}} = \mathcal{G}(\boldsymbol{\theta})$ makes it nonlinear.

To further simplify the likelihood (9) we also make a standard assumption that the observation errors are conditionally independent (Carrassi and others, 2018; Särkkä and Svensson, 2023). As such, our $N_o \times N_o$ observation error covariance matrix $\mathbf{R}$ becomes diagonal with entries corresponding to the observation error variance $\sigma^2_{y_m}$ associated with each of the $m = 1, \ldots, N_o$ observations y_m in the observation vector $\mathbf{y}$. When we only assimilate one type of observation, these entries are constant and equal to the observation error variance of either snow depth (σ_d^2) or albedo (σ_α^2). For joint assimilation, where both types of observation are assimilated, both error variances appear along the diagonal of $\mathbf{R}$ in accordance with the entries in $\mathbf{y}$.

2.5.2. Particle batch smoother

The Particle Batch Smoother was introduced in the snow literature by Margulis and others (2015) as a batch smoother version of the widely used particle filter (see Chopin and Papaspiliopoulos, 2020; Särkkä and Svensson, 2023). Algorithmically, the PBS boils down to performing basic sequential importance sampling (van Leeuwen, 2009) which effectively represents the posterior through a particle approximation

$$p(\boldsymbol{\theta}|\mathbf{y}) \simeq \sum_{i=1}^{N_e} w_i \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_i\right), \tag{10}$$

where $\mathrm{w_i}$ are the weights associated with each of the $i = 1, \ldots, N_e$ particles (samples) $\boldsymbol{\theta}_i$ in parameter space. These particles weights are self-normalized such that $\sum_{i=1}^{N_e} w_i = 1$. The $\delta(\cdot)$ in (10) denotes the Dirac delta which is a generalized function with properties $\int \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i)\,\mathrm{d}\boldsymbol{\theta} = 1$ and $\int g(\boldsymbol{\theta})\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i)\,\mathrm{d}\boldsymbol{\theta} = g(\boldsymbol{\theta}_i)$ for some function of the parameters $g(\boldsymbol{\theta})$. Thereby, the particle approximation represents the continuous posterior probability density function as a sum of discrete particles with probability mass given by their weights w_i. Posterior expectations become straightforward to compute, for example setting $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$ we recover the particle approximation to the posterior mean of the parameters as the weighted sum over particles $\boldsymbol{\theta}_i$. The corresponding posterior expectations in state space are obtained analogously. With minimal loss of accuracy, simpler unweighted posterior statistics are computed by first resampling particles based on the weights (Alonso-González and others, 2022). The weights $\mathrm{w_i}$ in the PBS are obtained through basic importance sampling approach using the prior as a proposal distribution to sample particles $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$ so that the weights effectively become the likelihood ratio

$$w_i = \frac{p(\mathbf{y}|\boldsymbol{\theta}_i)}{\sum_{k=1}^{N_e} p(\mathbf{y}|\boldsymbol{\theta}_k)}, \tag{11}$$

which when we insert for our Gaussian likelihood becomes (Aalstad and others, 2018)

$$w_i = \frac{\exp\left(-\frac{1}{2}\left[\mathbf{y}-\hat{\mathbf{y}}_i\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}-\hat{\mathbf{y}}_i\right]\right)}{\sum_{k=1}^{N_e}\exp\left(-\frac{1}{2}\left[\mathbf{y}-\hat{\mathbf{y}}_k\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}-\hat{\mathbf{y}}_k\right]\right)}, \tag{12}$$

where $\hat{\mathbf{y}}_i = \mathcal{G}(\boldsymbol{\theta}_i)$ denotes the vector of $\mathrm{N_o}$ predicted observables from CryoGrid for particle i with associated parameter vector $\boldsymbol{\theta}_i$, $[\cdot]^{\mathrm{T}}$ denotes the transpose, and $\mathbf{R}^{-1}$ is the inverse of the $N_o \times N_o$ observation error covariance matrix. In practice, we first compute the logarithm of the PBS weights in (12) to ensure numerical stability as described in Alonso-González and others (2022). Both the PBS and ES are batch smoothers in the sense that they assimilate a single batch of observations in a long data assimilation window, unlike a filter which updates sequentially as observations become available. The length of the window is typically defined by a typical timescale of the system being modeled, which we here take to be one mass balance year. This smoothing property is crucial since it allows the future to update the past: observations in the accumulation season can inform model states in the *preceding* accumulation season (Margulis and others, 2015; Aalstad and others, 2018). A computational advantage of the PBS is that it only requires running a single ensemble model integration of $\mathrm{N_e}$ particles sampled from the parameter prior. A particle approximation of the posterior for model parameters and state variables can then be obtained solely using the weights in (12) followed by a resampling step. As such, the computational cost of the PBS is incurred almost entirely by the need to run $\mathrm{N_e}$ forward simulations of the data generating model $\mathcal{G}$. It is this feature that helped motivate our design of a large ensemble of twin experiments, in that it is straightforward to test a large number of observation types and parameter scenarios based on a single large ensemble run by using a (fixed) prior distribution $p(\boldsymbol{\theta})$ as the proposal.

2.5.3. Ensemble smoother

We also test the ensemble smoother (ES) scheme that was originally proposed by van Leeuwen and Evensen (1996) as a batch smoother version of the widely used ensemble Kalman filter (EnKF, Evensen and others, 2022, Chapter 6). Here we use the classic stochastic version of the ES with perturbed observations to avoid underestimating ensemble covariances (van Leeuwen, 2020). The general framework of ensemble Kalman methods, which the ES falls under, extends the domain of applicability of classical Kalman filtering methods (Särkkä and Svensson, 2023), that require Gaussian linear data generating models, to Gaussian nonlinear models (Evensen and others, 2022). The Gaussian assumption in the prior and likelihood can also be relaxed through transformations using Gaussian

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

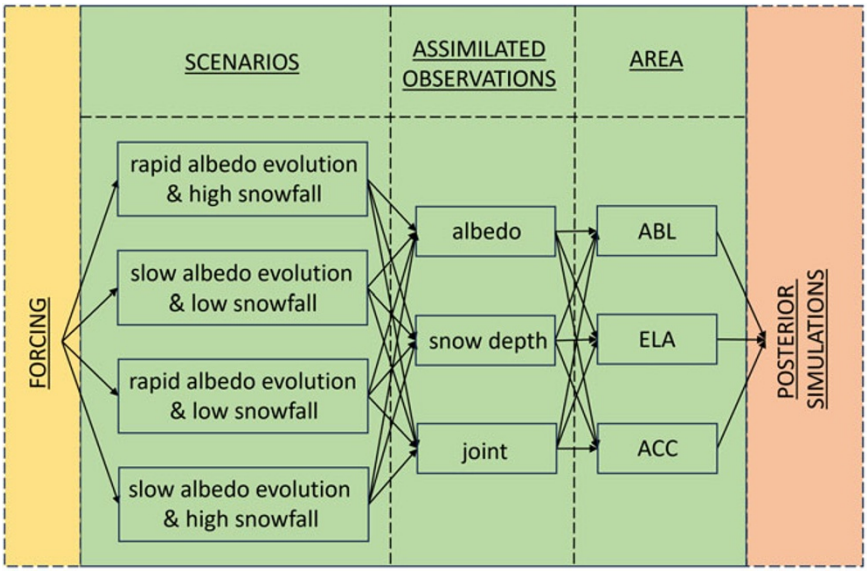


Figure 3. Structure of the large ensemble twin experiments based on permutations of four parameter scenarios, three types of assimilated observation vectors, and three experimental areas generating a total of 36 twin experiments. The scenarios combine either a rapid or slow albedo evolution rate with either a high or low snowfall factor. The assimilated observation vectors are either albedo only, snow depth only, or joint assimilation of albedo and snow depth. The experimental areas are either the ablation (ABL), equilibrium line altitude (ELA), and accumulation (ACC) areas depicted in Fig. 1.

anamorphosis functions (Bertino and others, 2003). Herein we use an analytical approach to Gaussian anamorphosis using the generalized logit transform in (7). Among the ensemble Kalman methods, the ES is most widely used for parameter estimation such as the strong constraint problem that we are tackling here.

The ES is initialized by sampling an ensemble of N_e parameter vectors $\boldsymbol{\theta}_i^{(0)}$ from the prior $\boldsymbol{\theta}_i^{(0)} \sim p(\boldsymbol{\theta})$. Using this prior parameter ensemble, following Aalstad and others (2018) the stochastic ES with analytical anamorphosis proceeds in the following steps while looping over ensemble members $i = 1, \ldots, N_e$:

(1) Generate an ensemble of prior predicted observables by running the parameters through the data generating model $\hat{\mathbf{y}}_i^{(0)} = \mathcal{G}\left(\boldsymbol{\theta}_i^{(0)}\right)$ which implicitly also involves generating an ensemble of prior model state vectors $\mathbf{x}_i^{(0)}$ for the whole data assimilation window (i.e., mass balance year in our case).
(2) Transform the prior parameter ensemble to Gaussian space using Gaussian anamorphosis $\boldsymbol{\phi}_i^{(0)} = \boldsymbol{\Psi}(\boldsymbol{\theta}_i^{(0)})$ in the form of the generalized logit transform (7).
(3) Perform the ensemble Kalman analysis step to update the parameters

$$\boldsymbol{\phi}_i^{(1)} = \boldsymbol{\phi}_i^{(0)} + \mathbf{K}^{(0)}\left(\mathbf{y} + \boldsymbol{\epsilon}_i - \hat{\mathbf{y}}_i^{(0)}\right) \tag{13}$$

where the ensemble Kalman gain $\mathbf{K}^{(0)}$ is obtained using ensemble covariance matrices together with $\mathbf{R}$ as outlined in Aalstad and others (2018) while realizations of Gaussian observation noise $\boldsymbol{\epsilon}_i \sim \mathrm{N}(0, \mathbf{R})$ are used to perturb the observations $\mathbf{y}$ in this stochastic scheme (van Leeuwen, 2020).
(4) Apply the inverse transformations using (8) to recover the posterior parameter ensemble in the original model parameter space $\boldsymbol{\theta}_i^{(1)} = \boldsymbol{\Psi}^{-1}(\boldsymbol{\phi}_i^{(1)})$.
(5) Rerun the data generating model to obtain an ensemble of posterior predicted observables $\hat{\mathbf{y}}_i^{(1)} = \mathcal{G}\left(\boldsymbol{\theta}_i^{(1)}\right)$ which also implicitly yields an ensemble of posterior model state vectors $\mathbf{x}_i^{(1)}$.

Note that the parameters $\boldsymbol{\theta}$ are updated directly while the model state $\mathbf{x}$ are updated indirectly. As such, to recover the posterior state $\mathbf{x}$ with the ES it is necessary to run the data generating model twice for each ensemble member, first with the prior parameters $\boldsymbol{\theta}_i^{(0)}$ in step 1 and subsequently with the posterior parameters $\boldsymbol{\theta}_i^{(1)}$ in step 5. Thereby, for posterior state estimation the ES is twice as costly as the PBS in that it requires running the data generating model $2N_e$ times.

2.6. Twin experiments

The conceptual diagram in Fig. 3 shows the structure of the twin experiments where we generated synthetic truth scenarios along with synthetic noisy observations. We constructed four different scenarios by using different true parameter vectors $\boldsymbol{\theta}^\star$ with different values for the true snow albedo evolution rate $\tau_a^\star$ and true snowfall factor $\beta_s^\star$. These true parameter vector scenarios include combinations of high and low values for each of the two parameters. The true parameter vector scenarios are then used in CryoGrid to generate synthetic true state $\mathbf{x}^\star$ scenarios including the true observables $\hat{\mathbf{y}}^\star = \mathcal{G}(\boldsymbol{\theta}^\star)$. These diverse true parameter scenarios are used to effectively mimic the variability of meteorological conditions and location-specific characteristics under differnet climatic scenarios. The synthetic albedo and snow depth obtained under these four different climatic scenarios were generated and perturbed with Gaussian noise that was scaled with the appropriate variances (σ_α^2 and σ_d^2) to mimic observation error. These noisy synthetic observations are then assimilated to constrain the prior CryoGrid simulations. Note that in this assimilation exercise, the model has no access to the hidden synthetic truth ($\boldsymbol{\theta}^\star$, $\mathbf{x}^\star$, $\hat{y}^\star$) other than through the corrupted information present in the noisy synthetic observations. This is the standard setup for widely used identical twin experiments where the same model is used to generate the observations and in the subsequent assimilation experiments (Arnold and Dey, 1986; Masutani and others, 2010). The generated synthetic observations, albedo and snow depth, can be assimilated either individually or jointly, amounting to a total of three assimilated observation scenarios. All experiments are applied in three different glacier zones, namely the ablation, ELA,

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

and accumulation areas. In total, we conducted 864,000 unique forward model realizations by using 4 scenarios × 3 glacier zones × 3 combinations of assimilated observations × 2 assimilation methods × 12 years × 1000 ensemble members.

The prior ensemble of CryoGrid simulations consists of $N_e = 1000$ ensemble members that were generated by perturbing albedo evolution rate and snowfall factor. We implemented the prior simulation with the same initial conditions for all data assimilation experiments. When initiating the model, we performed a 5 years spin-up to eliminate initialization shocks. Two different ensemble-based data assimilation methods, the PBS and ES, were compared in the twin experiments.

2.7. Evaluation of the experiments

To evaluate the performance of all experiments, we use the Continuous Ranked Probability Score (CRPS) to compare the posterior surface mass balance distribution to the synthetic truth surface mass balance. As outlined in Hersbach (2000), the CRPS is a statistical metric that compares probabilistic ensemble predictions to deterministic ground-truth values. Compared with Root Mean Square Error (RMSE) which is mainly used for deterministic forecasts, the CRPS is designed for probabilistic forecasts, which can evaluate the entire predictive distribution and provide a comprehensive assessment of the quality of predictions that include uncertainty quantification. The CRPS evaluates both the accuracy and the precision of the ensemble. The precision is a gauge of how well calibrated the ensemble is by punishing ensembles that are overconfident (too narrow) and underconfident (too wide). The CRPS is a negatively oriented score where a score of zero means that the probabilistic prediction is perfect, which only occurs for deterministic forecasts centered on the truth, while a larger CRPS entails a lower score. The CRPS is given by (Gneiting and others, 2005)

$$\mathrm{CRPS}(P, x^\star) = \int_{-\infty}^{\infty} \left(P(x) - H(x - x^\star)\right)^2 \, \mathrm{d}x \qquad (14)$$

where $P(x)$ is the cumulative distribution function of the ensemble prediction for variable x, $x^\star$ is the reference value which can be a synthetic truth or an observation, and $H(x - x^\star)$ is the Heaviside function, which is 1 if $x \geq x^\star$ and 0 otherwise. The CRPS inherits the same units as the variable x whose ensemble prediction is being evaluated.

3. Results

3.1. Influence of observations on surface mass balance modeling by PBS

Here, we present the results of surface mass balance (SMB) simulations in several twin experiments achieved by assimilating two observational datasets, albedo and snow depth, using the PBS scheme on an ensemble with $N_e = 1000$ members. Figure 4 shows the posterior annual SMB for the ablation area for the four different scenarios. The prior and posterior CRPS are calculated by comparing the prior and posterior SMB estimates with the synthetic truth over a 12-year period across these scenarios. Since this is the ablation area, the effect of the snow albedo evolution rate on the SMB (compare panels b to c and panels a to d) is clearly negligible because the surface albedo is mostly that of bare ice in either scenario. As such, snow depth is a stronger constraint on the SMB than albedo in this case. After assimilation, the average CRPS values (Fig. 5) for the posterior SMB estimates are 0.05 m for albedo assimilation, 0.03 for snow depth assimilation, and 0.02 m for jointly assimilating both observations. For all assimilated observation scenarios this is a marked improvement from the prior CRPS of 0.16 m. These improvements represent CRPS reductions of 69%, 80%, and 86%, respectively, compared to the prior, showing the enhanced skill of the posterior estimates obtained after data assimilation. On the one hand, snow depth assimilation is particularly effective in bringing the posterior ensemble median SMB closer to the truth. On the other hand, the 95th percentile of the posterior ensemble after albedo assimilation often scarcely encompassed the true values. A comparison of error reduction demonstrates (Fig. 4) that joint assimilation of albedo and snow depth provides the most substantial improvements in the performance of posterior SMB simulations in the ablation area. When comparing the four scenarios, it becomes evident that snow depth assimilation performs better under high snowfall conditions, yielding a CRPS of 0.03 m compared to 0.04 m in low snowfall conditions. Conversely, albedo assimilation performs better in low snowfall scenarios, with a CRPS of 0.04 m compared to 0.06 m under high snowfall conditions.

Figure 5 provides a comprehensive evaluation of the experiments using the PBS assimilation scheme across all scenarios and areas. In both the ablation and equilibrium line altitude areas, the assimilation of either albedo or snow depth substantially reduces CRPS compared to the prior estimates. The average CRPS reduction is 71% and 74%, respectively, when albedo and snow depth are assimilated individually. The difference in performance improvement in terms of CRPS between albedo and snow depth assimilation is particularly notable, ranging from a 10% to 19% difference, especially under the high snowfall scenario. In most experiments, joint assimilation of albedo and snow depth consistently yields the lowest CRPS values. However, in the accumulation area, results indicate an increase in CRPS following snow depth assimilation under low snowfall scenarios, relative to the prior. In contrast, albedo assimilation still improves performance, though the improvements are less pronounced than in the ablation and equilibrium line altitude areas, especially under high snowfall conditions. Similarly, joint assimilation of albedo and snow depth exhibits behavior similar to snow depth assimilation alone in the accumulation area.

The results indicate that the assimilation of joint albedo and snow depth observations within the PBS framework improves the skill of surface mass balance simulations, particularly in scenarios with high snowfall. The results show that snow depth assimilation tends to perform better under high snowfall conditions in the ablation area, while albedo assimilation is more effective under low snowfall scenarios in the ablation and accumulation areas. In the ELA area, albedo assimilation outperforms snow depth assimilation under high snowfall scenarios, while the two methods yield comparable results in low snowfall conditions. Moreover, joint assimilation tends to yield the best (including ties) results across the majority of experiments (10 out of 12) in Fig. 5, providing the greatest improvements both in terms of reducing uncertainty and bringing the posterior closer to the truth. These findings highlight the importance of selecting appropriate observations to assimilate based on specific climatic conditions to optimize the performance of SMB simulations. In particular, the most robust choice is generally joint data assimilation which can automatically handle trade-offs in the information content of different types of observations.

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

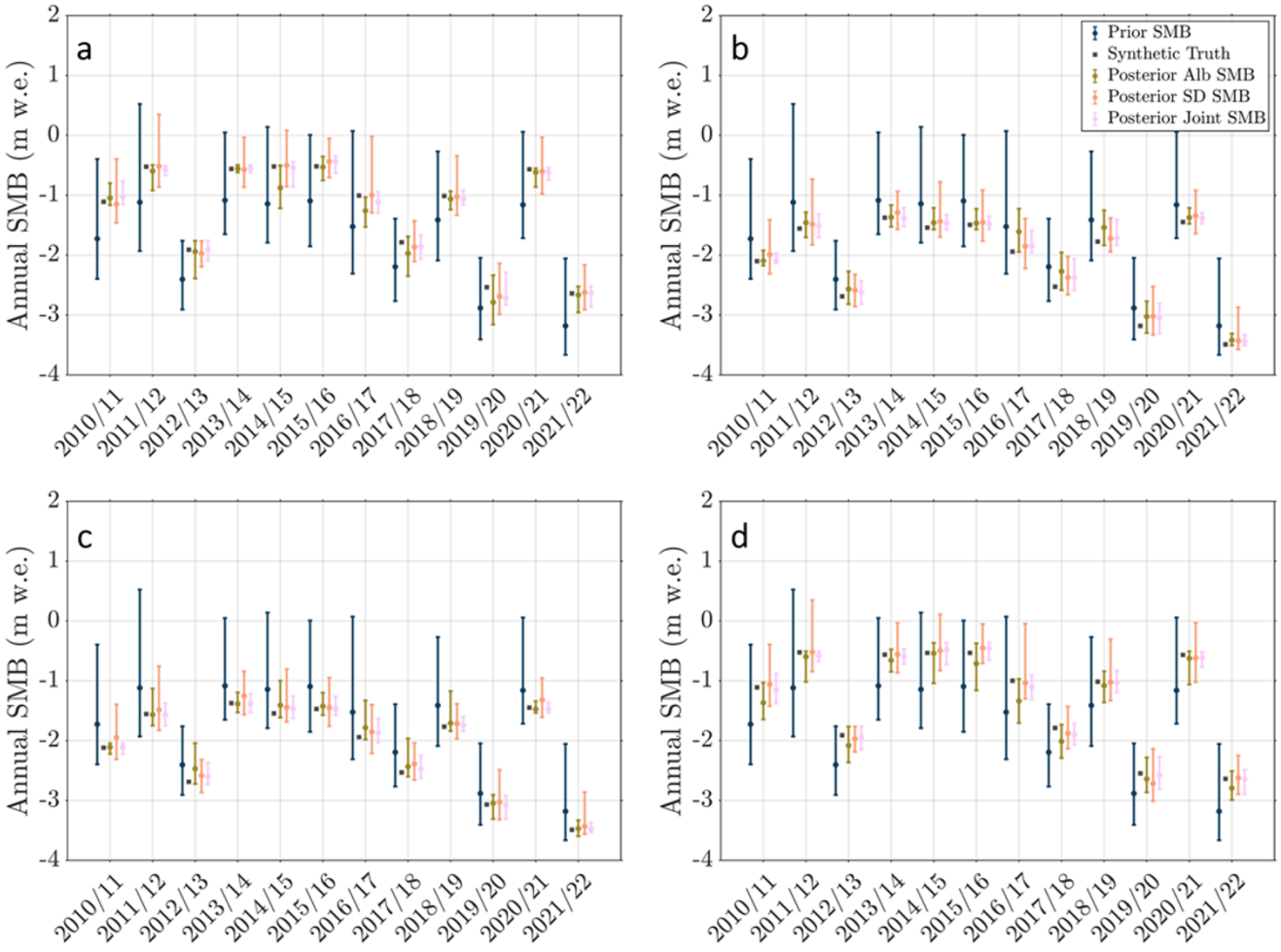


Figure 4. Comparison of prior, posterior, and true surface mass balance in the ablation area when using the PBS to assimilate albedo only, snow depth only, and both observations jointly. The figure presents four scenarios based on the snow albedo evolution rates and snowfall factors: (a) Rapid snow albedo evolution with high snowfall. (b) Slow snow albedo evolution with low snowfall. (c) Rapid snow albedo evolution with low snowfall. (d) Slow snow albedo evolution with high snowfall. Error bars represent the 95th central percentile range of the ensemble with the points indicating the median value for surface mass balance estimates.

The ES scheme overall exhibits performance similar to that of the PBS scheme after assimilating albedo and snow depth (Fig. S1 and Fig. S2). Joint assimilation yields the best results, while the assimilation of albedo and snow depth individually shows varying outcomes across different scenarios. A detailed comparison of the two assimilation schemes follows in the next section.

3.2. Comparison of two data assimilation schemes

We evaluated both the PBS and ES schemes against synthetic truth using an ensemble size of 1000 members for all experiments. Table 4 presents the improvement in CRPS performance in terms of surface mass balance achieved by the two data assimilation schemes relative to the prior. The values represent the average improvement across four truth scenarios, calculated by comparing the posterior results with the prior. For albedo assimilation, PBS shows a significantly better overall performance compared to ES across all glacier zones. However, for snow depth assimilation, ES performs slightly better than PBS, except in the accumulation zone. Joint assimilation of both albedo and snow depth yields the best performance, regardless of the assimilation method used. In terms of different glacier zones, the performance in the ablation area is generally the best across both data assimilation schemes. In the ELA region, the results slightly underperform those in the ablation area, considering the average performance of three distinct assimilated observation scenarios. The accumulation zone yields the lowest accuracy improvement among the glacier zones for both schemes. Nonetheless, the posterior always improved over the prior in terms of surface mass balance CRPS. Note that these results represent averages across four scenarios and considerable differences exist between individual scenarios, particularly when assimilating snow depth generated under different snowfall factors.

Figure 6 presents statistical properties of the posterior annual surface mass balance results derived from the two data assimilation schemes in the ELA region under a scenario of rapid snow albedo evolution and high snowfall. Joint assimilation under the ES scheme demonstrates the best overall performance, achieving the lowest RMSE and standard deviation compared to other configurations. While the PBS scheme also performs well with joint assimilation, but the variability in standard deviation across years is notably higher than in the ES results. For albedo assimilation, both methods considerably enhance the accuracy and reduce

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

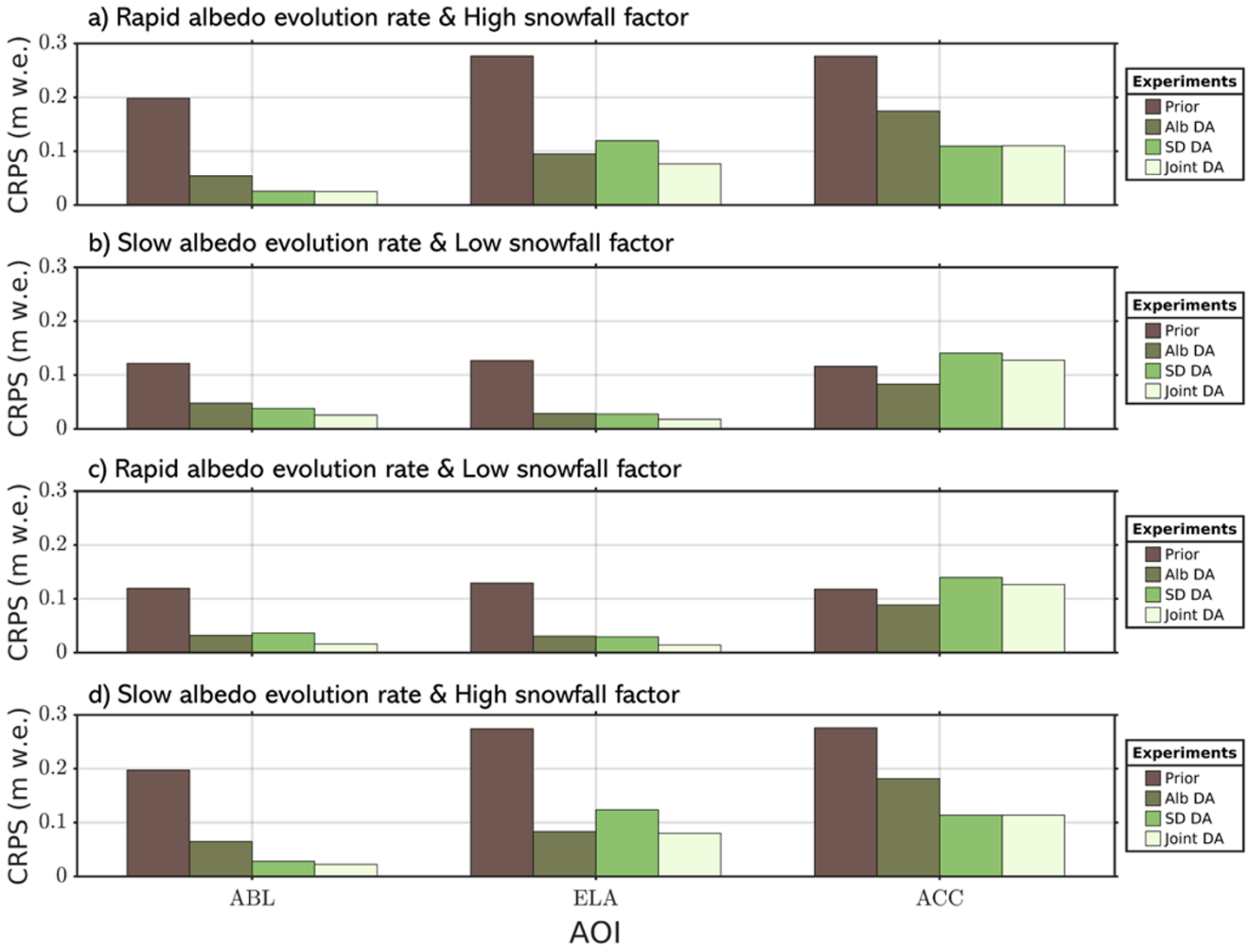


Figure 5. Continuous ranked probability score (CRPS) for the prior and posterior surface mass balance after assimilating albedo, snow depth, and both observations jointly using the PBS, for all scenarios: (a) rapid albedo evolution rate & high snowfall factor, (b) slow albedo evolution rate & low snowfall factor, (c) rapid albedo evolution rate & slow snowfall factor, (d) slow albedo evolution rate & high snowfall factor, and three different areas of interest (AOI): ablation area (ABL), equilibrium-line area (ELA), accumulation area (ACC), compared to synthetic true surface mass balance.

uncertainty compared to the prior. However, the PBS scheme yields slightly lower RMSE and standard deviation than the ES scheme but exhibits higher interannual variability. In contrast, the ES scheme demonstrates a more stable annual performance. Regarding snow depth assimilation, the posterior results from the PBS scheme reveal overconfidence, characterized by an ensemble spread near zero and high annual variability, along with a higher average RMSE than the ES scheme. In comparison, the ES scheme produces a smaller standard deviation and maintains stable annual performance after snow depth assimilation, with no marked interannual fluctuations. However, this disadvantage of the results from PBS may partially stem from the bias introduced by overconfidence, especially influenced by snow depth assimilation. In PBS, particles are weighted based on their likelihood given the observations. If the observations strongly favor one particle, that particle's weight approaches 1 (Margulis and others, 2015). During resampling, particles with low weights are discarded, and the dominant particle is replicated multiple times. This reduces the effective number of particles, collapsing the posterior to a single point or a very narrow range (Aalstad and others, 2018). It is also unclear if the higher computational cost of $2N_e$ CryoGrid running with the ES, compared to just N_e with the PBS, justifies the slight gain in performance in this case. If the prior ensemble is so biased that it does not encompass the observations, the PBS is incapable of correcting the posterior towards the observations outside the bounds of the prior (Aalstad and others, 2018), while ES can address this problem by re-run the model with updated parameters and higher computational cost.

Table 4. Comparison of two data assimilation methods in improving the average CRPS of surface mass balance simulations by assimilating different observations for various glacier zones. The values in the table represent the average CRPS improvement, calculated by comparing the percentage improvement of the posterior CRPS results to that of the prior CRPS results, across all four scenarios

		Albedo	Snow depth	Joint
PBS	**ABL**	68.4%	77.8%	85.4%
	ELA	72.2%	66.9%	79.5%
	ACC	31.0%	19.9%	25.5%
ES	**ABL**	48.3%	79.0%	85.6%
	ELA	47.9%	67.4%	76.7%
	ACC	24.1%	11.6%	25.4%

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

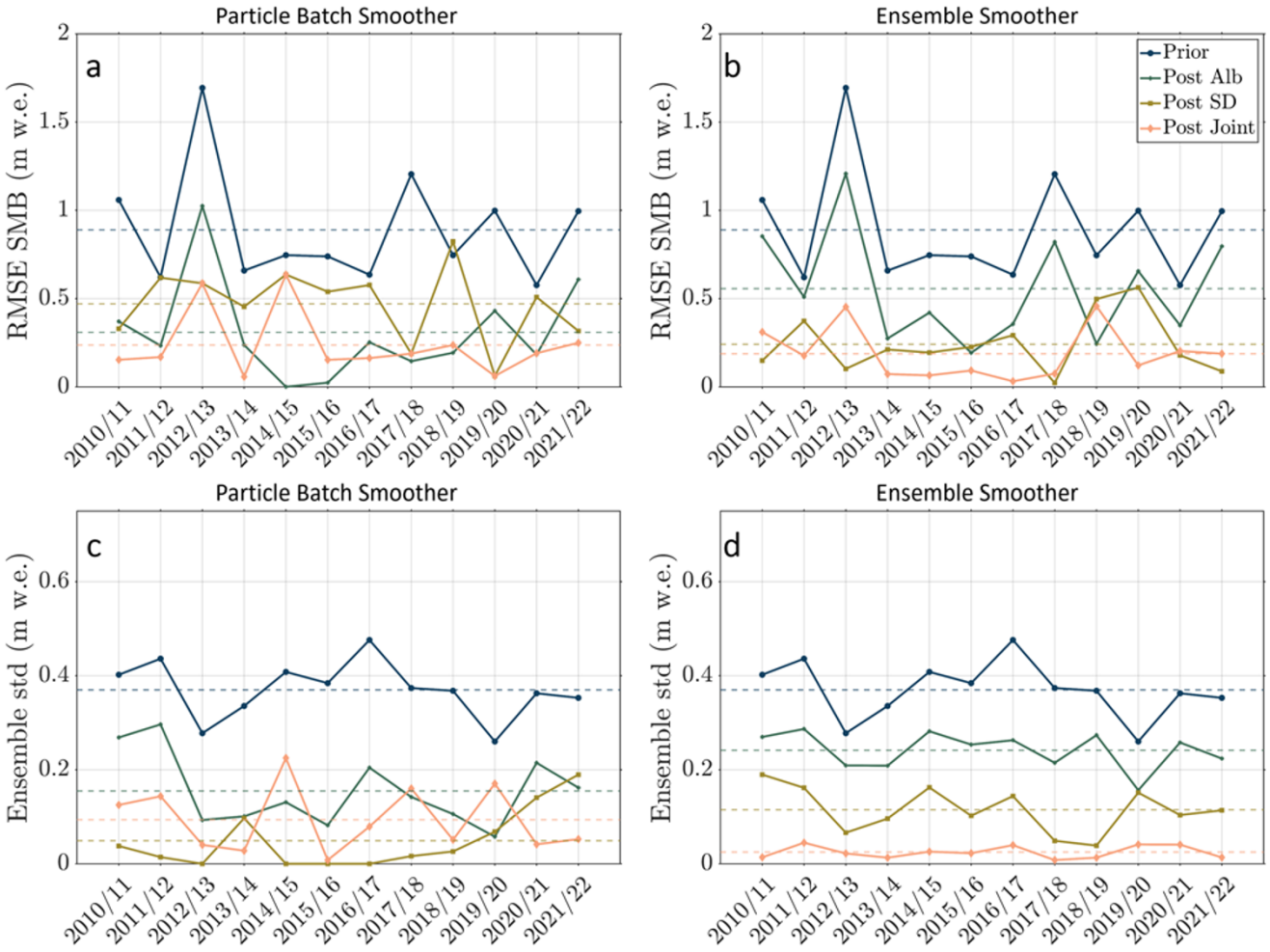


Figure 6. Comparison of the performance of two assimilation schemes applied to the ELA area under rapid albedo evolution and high snowfall scenario in terms of RMSE (top row) and ensemble standard deviation (bottom row) for the Particle Batch Smoother (left panels a and c) and the Ensemble Smoother (right panels b and d). Dashed lines represent the mean value of RMSE and ensemble standard deviation of the simulation performance.

3.3. Sensitivity of data assimilation performance to the ensemble size

In this section, we present the results of analysing the sensitivity of data assimilation performance to the number of ensemble members N_e. The range of ensemble sizes N_e investigated was selected to be regular on a logarithmic scale, generating a vector of seven logarithmically spaced values between 10^1 and 10^3 ensemble members. Figure 7 illustrates the mean and variance of CRPS values obtained from 100 iterations of bootstrapping (resampling with replacement) prior ensembles of variable size N^e from the original large ensemble of 1000 prior parameters (used in the rest of the study) followed by the assimilation of joint albedo and snow depth under the PBS scheme. The results indicate that, across all experiments, both the average CRPS and its Monte Carlo variance decrease as the ensemble size increases. This replicates the improvement in performance, both in terms of mean and variance, with increased ensemble size as expected from Monte Carlo methods. Moreover, as expected, the rate of error reduction diminishes considerably, particularly after the ensemble size reaches 100, the mean CRPS starts to show clear convergent behavior towards an asymptote around 0.025 (m w.e.) with a steadily decreasing variance. However, unlike the Monte Carlo variance, interannual variability remains relatively stable and does not exhibit any clear dependence on ensemble size.

4. Discussion

4.1. Influence of observations on surface mass balance modeling

Across all scenarios and regions, the assimilation of albedo consistently brings the ensemble median of the SMB simulations closer to the true values while effectively reducing the ensemble spread. This improvement is consistent with the findings of Dumont and others (2012), which demonstrated that assimilating MODIS-derived albedo in a snowpack model improves the accuracy of the SMB simulation for an alpine glacier in the French Alps through variational assimilation. Despite claims to this effect, Dumont and others (2012) did not show how their variational data assimilation scheme constrained uncertainty. In contrast, our ensemble-based data assimilation results show that both the PBS and ES schemes effectively constrain the ensemble, leading to significant reductions in uncertainty. Moreover, unlike variational methods, the ensemble-based schemes pursued herein do not require a differentiable data-generating model and are thus more widely applicable.

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

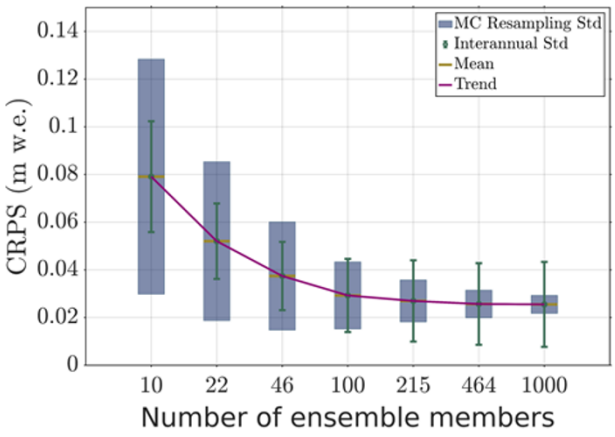


Figure 7. Sensitivity of the posterior surface mass balance CRPS to ensemble size following joint assimilation of albedo and snow depth using the PBS scheme under the scenario of rapid albedo evolution rate and high snowfall factor in the ablation area. For each ensemble size (N_e) the CRPS statistics were estimated by resampling with replacement (i.e., bootstrapping) an ensemble of N_e particles from the complete large ensemble (1000 members) 100 times, evaluating the CRPS for each of these 100 bootstrapped ensembles, and subsequently computing sample statistics.

Figure 6 highlights the improvement in accuracy and the reduction in uncertainty achieved by albedo assimilation, with PBS outperforming ES in both metrics.

The impact of snow depth assimilation on SMB simulations exhibits some spatial variability, but overall, snow depth assimilation generally enhances SMB accuracy, with more consistent improvements observed outside the accumulation area. Under high snowfall factor scenarios, snow depth assimilation markedly improves SMB simulation accuracy across all regions, aligning with the general findings by Landmann and others (2021) for surface mass balance and Magnusson and others (2017) for seasonal snow.

Our results show a substantial impact of snow depth assimilation on model performance in terms of SMB , with an average improvement of 74% in SMB accuracy in both the ablation and ELA regions. This demonstrates similar performance gains to previous studies. For example, Landmann and others (2021) reported a relatively low CRPS of 0.012 m w.e. surface mass balance compared with cumulative observations, while Magnusson and others (2017) observed a 64% reduction in SWE error from snow depth assimilation across 40 sites in Switzerland. Unlike these studies that use particle filtering techniques, we apply the smoothing-based PBS and ES schemes that allow information from the observations to propagate backward in time which has been shown to be advantageous for retrospective snow data assimilation (Alonso-González and others, 2022).

Under low snowfall scenarios, snow depth assimilation alone yields less favorable results, particularly in the accumulation area. As illustrated in Fig. 8, posterior estimates in PBS collapse to a single particle with snow depth assimilation in this scenario. This phenomenon is likely due to limitations inherent in the PBS scheme (Robinson and others, 2018; Pirk and others, 2022) and the nature of the low snowfall setting, which produces some SMB truth values that fall outside the prior ensemble range. This discrepancy prevents the posterior from fully encompassing true values, and, when coupled with the ensemble's overconfidence, results in an increased CRPS due to bias and overconfident predictions. Under the same conditions, ES outperforms PBS due to fundamental differences in both the assumptions and updates steps in these methods (Margulis and others, 2015; Aalstad and others, 2018; Alonso-González and others, 2022). Additionally, synthetic snow depth data were perturbed with Gaussian noise, maintaining a uniform standard deviation of 50 cm across all glacier zones. In practice, the RMSE of in situ snow depth measurements in Svalbard is around 5cm, particularly in accumulation zones with rare rainfall, which is substantially lower than the applied noise level (An and others, 2020). Assimilation of lower-quality observational data into the model may compromise the posterior estimates, resulting in suboptimal performance.

To address data availability challenges, we generated synthetic observational data for albedo and snow depth, potentially providing daily coverage over a full year. Subsequently, we applied the specific methods mentioned above to select data points that mimic the temporal availability of ICESat-2 and MODIS measurements. This approach enabled us to control the experimental environment under conditions of parameter uncertainty, thereby facilitating the execution of large ensemble experiments. While this approach theoretically fulfilled continuous data requirements, achieving similar completeness with real observational data remains challenging (Gabarró and others, 2023; Sandven and others, 2023). Satellite-based measurements, such as those from ICESat-2 and MODIS, face limitations due to cloud contamination, which degrades data quality and restricts data acquisition (Ø stby and others, 2014; Neuenschwander and Magruder, 2019; Kotarba, 2022). Additionally, optical satellites that provide albedo data are limited by daylight availability (Wang and others, 2018), resulting in data gaps in areas with heavy cloud cover or reduced sunlight. Consequently, while synthetic data can theoretically satisfy continuous data requirements, real-world data collection remains inherently constrained by these observational challenges that we aimed to replicate in the design of our twin experiments.

4.2. Performance of data assimilation schemes

For all the given observations and research areas, both data assimilation schemes contribute to considerably reductions in uncertainty and error in SMB simulations. The PBS showed superior performance in albedo assimilation, offering a more confident and accurate ensemble. Conversely, ES generally outperformed PBS in snow depth assimilation scenarios, particularly where the model's prior did not bracket the truth value (Fig. 8). The PBS operates by weighing the ensemble of states based on their likelihood (Margulis and others, 2015; Aalstad and others, 2018), avoiding the need to move particles in parameter space. This results in lower computational demands for state estimation as it only requires one model run per ensemble member. In our study, PBS was particularly effective for albedo assimilation, offering significant uncertainty reduction with less computational effort. However, the performance of PBS can be limited when the true state falls outside the range of the prior ensemble(Robinson and others, 2018; Pirk and others, 2022), as seen in scenarios with low snowfall where the posterior ensemble sometimes became degenerate and overconfident. To address this limitation, a potential approach involves implementing hybrid schemes that combine ensemble Kalman methods and particle methods to alleviate this issue (Pirk and others, 2022), although this would increase computational costs.

In contrast, ES updates the state by moving particles in parameter space, which can lead to better coverage of the true state, especially when it lies outside the prior ensemble range (van Leeuwen and Evensen, 1996; Evensen and others, 2022). This adaptability

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

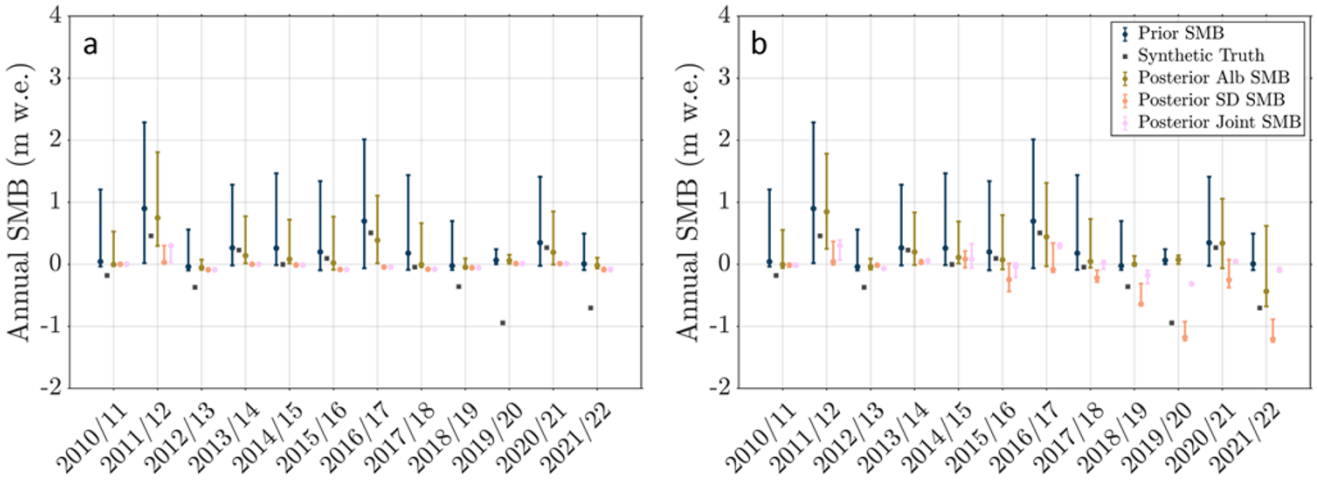


Figure 8. Comparison of the prior, posterior, and true annual surface mass balance in the accumulation area when assimilating different types of observations with the PBS (a) and the ES (b) under a low snowfall factor and slow albedo evolution rate scenario.

was evident in our results, where ES performed better in assimilating snow depth, particularly under low snowfall scenarios. The ES method requires twice the number of model runs compared to PBS because it requires rerunning the model with updated parameters, which increases computational cost, but can lead to more accurate results in certain scenarios.

Whereas traditional Markov Chain Monte Carlo (MCMC) methods involve numerous sequential iterations to converge on a solution, both PBS and ES use parallelizable ensemble approaches, significantly reducing computational time. For instance, in the study by Rounce and others (2020), MCMC methods were used to quantify parameter uncertainty in glacier models, involving costly iterations to sample the posterior distribution. Our methods avoid the mainly iterative sampling of MCMC by directly updating an ensemble of parameter vectors, providing a faster convergence to a posterior estimate. While MCMC methods can be very accurate due to their thorough sampling of parameter space, they are often computationally heavy for complex models like CryoGrid, where each model simulation is expensive. Our approach integrates the complexity of CryoGrid with efficient data assimilation methods, allowing for more frequent updates or larger ensembles without a proportional increase in computational demand.

4.3. Sensitivity to ensemble size

The sensitivity is evaluated based on two components: resampling (Monte Carlo) variance and interannual variability. As reported in the results, the average CRPS decreases with increasing ensemble size. Notably, the variance of CRPS from interannual variability remains unchanged, whereas the variance of CRPS associated with Monte Carlo resampling error follows the overall decreasing trend of the total error. Interannual variance reflects the natural variability in the system over the different years, capturing the system's response to varying climatic conditions (Malone and others, 2019; Wei and others, 2019). In this study, all experiments are forced using the same meteorological data source, meaning that interannual variance is inherent to the system and remains unaffected by ensemble size. The problem of small ensemble sizes resulting in large resampling variance is well-documented, as subsets sampled from smaller ensembles may fail to adequately represent the full diversity of a larger ensemble, leading to greater variance in the results (Choi and Lee, 2025). In our study, as the ensemble size increases to 100, the rate of improvement in CRPS (result accuracy) and the reduction in resampling variance both exhibit a diminishing trend. While larger ensembles generally reduce sampling errors, they come at the cost of increased computational demands (Sacher and Bartello, 2008). The optimal ensemble size, however, depends on the specific design of the experiment and the acceptable trade-off between computational cost and error tolerance for the user (Milinski and others, 2020). When the ensemble size reaches 1000, the resampling variation might be expected to approach zero in our experiment design, as the entire large ensemble pool comprises 1000 unique members. However the resampling variation remains non-zero even when the ensemble size is 1000. This is because the CRPS statistics are estimated through a bootstrapping process, where an ensemble of N particles is resampled *with replacement* from the complete large ensemble. Even when the number of samples matches the original pool size, the randomness introduced by bootstrapping with replacement ensures that the resampled subset does not perfectly replicate the original pool (Davison and Hinkley, 1997). Some particles may appear multiple times, while others may be excluded entirely. This stochastic nature of the bootstrapping process introduces Monte Carlo sampling error, leading to persistent variability in the results and ensuring a non-zero resampling variance that mimics the actual Monte Carlo variance that would arise when individual particles are sampled multiple times within 1000 ensemble members. The bootstrap technique used herein is a computationally affordable way to probe Monte Carlo sampling error that could otherwise be prohibitively expensive to evaluate in that it would require running multiple distinct large ensembles through CryoGrid.

5. Conclusions

In this study, we applied two data assimilation schemes, the PBS and ES, to simulate glacier surface mass balance from 2010 to 2022 across different glacier zones in Kongsvegen with 2.5 km grid cells through extensive large ensemble twin experiments. The posterior results were evaluated by comparing them with synthetic true surface mass balance values using the CRPS metric with the prior CRPS as a reference from which improvement was measured. Cross-comparisons across different scenarios further illustrated

https://doi.org/10.1017/jog.2025.10101 Published online by Cambridge University Press

the impact of various observational data on surface mass balance simulations under different assimilation schemes. From this study, the following conclusions can be drawn:

- Assimilating albedo generally improves SMB simulation across all glacier zones, with 68.4% improvement by PBS and 48.3% improvement by ES. However, the degree of improvement varies between different glacier areas. In particular, results in the ablation area show an average improvement of 58.4%, which is greater than the 27.5% improvement observed in the accumulation area.
- The assimilation of snow depth yields results comparable to those of albedo assimilation, particularly in the ablation and ELA zones, 77.8% (79%) and 66.9% (67.4%) respectively, for the PBS (ES) data assimialtion scheme. However, under the low snowfall scenarios within the PBS scheme, methodological limitations cause the posterior results to collapse to a single point in the accumulation zone, resulting in an overly constrained ensemble. This excessive constraint leads to outcomes that are both overconfident and biased.
- Both assimilation schemes lead to marked improvements in surface mass balance simulations. While the PBS outperforms the ES in assimilating albedo, the ES demonstrates marginally better performance over the PBS when assimilating snow depth.
- The joint assimilation of both observation types gives the best performance across all experiments except those given by low snowfall level in the accumulation area. The average improvement in CRPS after joint assimilation across all different glacier areas is 63.5%.
- Resampling from the large 1000 members ensemble using varying ensemble sizes, the rate of improvement, reflected in both the variance of the Monte Carlo resampling and the median CRPS, slows considerably when the ensemble size reaches 100 indicating diminishing performance gains with further computationally costly increases in ensemble size.

The twin experiments in this study demonstrated strong performance gains in most scenarios, including various glacier zones and observational data. This establishes the assimilation approach as effective in synthetic experiments and suggests that it is potentially transferable for estimating surface mass balance of all glaciers on Svalbard. Corroborating this claim will require further experiments with real observations. However, observational data can be inconsistent in real-world applications, posing further implementation challenges when relying on satellite-based observations due to factors such as gaps and retrieval uncertainty.

Data availability statement. CARRA data was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store at https://doi.org/10.24381/cds.d29ad2c6. The results are generated using Copernicus Climate Change Service information (2025). Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. The CryoGrid community model is hosted on Github. The source code is available at https://github.com/CryoGrid/CryoGridCommunity_source.

Acknowledgements. This study was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie action COMPSCI (grant no. 945371), the EEA and Norway initiative HarSval (grant no. UMO-2023/43/7/ST10/00001) and by the European Union's Horizon Europe program through the project LIQUIDICE (grant no. 101184962). Louise S. Schmidt was funded by the Research Council of Norway through the Nansen Legacy project (NFR-276730) and the MAMMAMIA project (NFR-301837). Kristoffer Aalstad acknowledges funding from the ERC-2022-ADG under grant agreement No 101096057 GLACMASS and an ESA CCI Research Fellowship (PATCHES project).The simulations were performed on resources provided by the Department of Geosciences, University of Oslo.

Competing interests. The authors declare that they have no conflict of interest.

References

Aalstad K, Westermann S, Schuler TV, Boike J and Bertino L (2018) Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *Cryosphere* **12**(1), 247–270. doi: 10.5194/tc-12-247-2018.

Alonso-González E and 7 others (2022) The multiple snow data assimilation system (MuSA v1.0). *Geoscientific Model Development* **15**(24), 9127–9155. doi: 10.5194/GMD-15-9127-2022.

An J and **6 others** (2020) Snow depth variations in Svalbard derived from GNSS interferometric reflectometry. *Remote Sensing* **12**(20), 3352. doi: 10.3390/RS12203352.

Arnold CP and Dey CH (1986) Observing-systems simulation experiments: Past, present and future. *Bulletin of the American Meteorological Society* **67**(6), 687–695. doi: 10.1175/1520-0477(1986)067<0687:OSSEPP>2.0.CO;2.

Banner KM, Irvine KM and Rodhouse TJ (2020) The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution* **11**(8), 882–889. doi: 10.1111/2041-210X.13407.

Bengtsson L and 22 others (2017) The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Monthly Weather Review* **145**(5), 1919–1935. doi: 10.1175/MWR-D-16-0417.1.

Bertino L, Evensen G and Wackernagel H (2003) Sequential data assimilation techniques in oceanography. *International Statistical Review* **71**(2), 223–241. doi: 10.1111/j.1751-5823.2003.tb00194.x.

Bertoncini A, Pomeroy J and Pomeroy JW (2024) Assimilation of satellite albedo to improve simulations of glacier hydrology. *Authorea Preprints*, doi: 10.22541/AU.171128344.49725407/V1

Brinkerhoff DJ (2022) Variational inference at glacier scale. *Journal of Computational Physics* **459**, 111095. doi: 10.1016/J.JCP.2022.111095. .

Budyko MI (1969) The effect of solar radiation variations on the climate of the Earth. *Tellus* **21**(5), 611–619. doi: 10.1111/j.2153-3490.1969.tb00466.x.

Carrassi A, Bocquet M, Bertino L and Evensen G (2018) Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change* **9**, e535. doi: 10.1002/wcc.535.

Choi B and Lee Y (2025) Sampling error mitigation through spectrum smoothing: First experiments with ensemble transform Kalman filters and Lorenz models. *Physica D: Nonlinear Phenomena* **472**, 134436. doi: 10.1016/j.physd.2024.134436.

Choi Y, Seroussi H, Morlighem M, Schlegel NJ and Gardner A (2023) Impact of time-dependent data aVssimilation on ice flow model initialization and projections: A case study of Kjer Glacier, Greenland. *Cryosphere* **17**(12), 5499–5517. doi: 10.5194/TC-17-5499-2023.

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

Copernicus Climate Change Service (C3S) (2024) Climate Data Store (CDS) Arctic regional reanal- ysis on single levels from 1991 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi: 10.24381/cds.713858f6.

Davison AC and Hinkley DV (1997) Bootstrap methods and their application. *Bootstrap Methods and Their Application*. doi: 10.1017/CBO9780511802843

Deschamps-Berger C, Gascoin S, Shean D, Besso H, Guiot A and López-Moreno JI (2023) Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data. *Cryosphere* **17**(7), 2779–2792. doi: 10.5194/TC-17-2779-2023.

Devilliers M, Yang S, Drews A, Schmith T and Olsen SM (2024) Ocean response to a century of observation-based freshwater forcing around Greenland in EC-Earth3. *Climate Dynamics* **62**(6), 4905–4923. doi: 10.1007/S00382-024-07142-0/METRICS.

Dumont M, Durand Y, Arnaud Y and Six D (2012) Variational assimilation of albedo in a snowpack model and reconstruction of the spatial mass-balance distribution of an alpine glacier. *Journal of Glaciology* **58**(207), 151–164. doi: 10.3189/2012JOG11J163.

Evensen G, Vossepoel FC and van Leeuwen PJ (2022) *Data Assimilation Fundamentals*. In *Earth Science, Geography, and Environment*, Springer International Publishing, Cham, doi: 10.1007/978-3-030-96709-3. ISBN 978-3-030-96708-6.

Ezat MM, Fahl K and Rasmussen TL (2024) Arctic freshwater outflow suppressed Nordic Seas overturning and oceanic heat transport during the Last Interglacial. *Nature Communications* **15**(1), 1–9. doi: 10.1038/s41467-024-53401-3.

Forbes R, Tompkins AM and Untch A (2011) A new prognostic bulk microphysics scheme for the IFS. *Technical Memorandum* **649** 28. doi: 10.21957/bf6vjvxk.

Günther D, Marke T, Essery R and Strasser U (2019) 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**(4), 2779–2800. doi: 10.1029/2018WR023403.

Gabarró C and 11 others (2023) Improving satellite-based monitoring of the polar regions: Identification of research and capacity gaps. *Frontiers in Remote Sensing* **4**, 952091. doi: 10.3389/FRSEN.2023.952091/BIBTEX.

Geyman EC, van Pelt JJ Maloof W Aas HF and Kohler J (2022) Historical glacier change on Svalbard predicts doubling of mass loss by 2100. *Nature* **601**(7893), 374–379. doi: 10.1038/s41586-021-04314-4.

Gillet-Chaulet F (2020) Assimilation of surface observations in a transient marine ice sheet model using an ensemble Kalman filter. *Cryosphere* **14**(3), 811–832. doi: 10.5194/TC-14-811-2020.

Gneiting T, Raftery AE, Westveld AH and Goldman T (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review* **133**(5), 1098–1118. doi: 10.1175/MWR2904.1.

Groenke B, Langer M, Nitzbon J, Westermann S, Gallego G and Boike J (2023) Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer. *Cryosphere* **17**(8), 3505–3533. doi: 10.5194/TC-17-3505-2023.

Guidicelli M, Aalstad K, Treichler D and Salzmann N (2024) A combined data assimilation and deep learning approach for continuous spatio-temporal SWE reconstruction from sparse ground tracks. *Journal of Hydrology X* **25**, 100190. doi: 10.1016/j.hydroa.2024.100190.

Gunnarsson A, Gardarsson SM and Pálsson F (2023) Modeling of surface energy balance for Icelandic glaciers using remote-sensing albedo. *Cryosphere* **17**(9), 3955–3986. doi: 10.5194/TC-17-3955-2023.

Hagen JO, Melvold K, Eiken T, Isaksson E and Lefauconnier B (1999) Mass balance methods on Kongsvegen, Svalbard. *Geografiska Annaler: Series A, Physical Geography* **81**(4), 593–601. doi: 10.1111/j.0435-3676.1999.00087.x.

Hersbach H and **43 others** (2020) The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**(730), 1999–2049. doi: 10.1002/qj.3803.

Hersbach H (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**(5), 559–570. doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hock R (2003) Temperature index melt modelling in mountain areas. *Journal of Hydrology* **282**(1-4), 104–115. doi: 10.1016/S0022-1694(03)00257-9.

Hock R and Holmgren B (2005) A distributed surface energy-balance model for complex topography and its application to Storglaciären, Sweden. *Journal of Glaciology* **51**(172), 25–36. doi: 10.3189/172756505781829566.

Hopwood MJ and 16 others (2020) Review article: How does glacier discharge affect marine biogeochemistry and primary production in the Arctic?. *Cryosphere* **14**(4), 1347–1383. doi: 10.5194/TC-14-1347-2020.

Huss M and Hock R (2015) A new model for global glacier change and sea-level rise. *Frontiers in Earth Science* **3**(September), 1–22. doi: 10.3389/feart.2015.00054.

Immerzeel WW and 31 others (2019) Importance and vulnerability of the world's water towers. *Nature* **577**(7790), 364–369. doi: 10.1038/s41586-019-1822-y.

IPCC (2022) *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. Cambridge, UK: Cambridge University Press. doi: 10.1017/9781009157964.

Isaksen K, Nordli FEJ, Łupikasza E, Eastwood S and Niedźwiedź T (2016) Recent warming on Spitsbergen—Influence of atmospheric circulation and sea ice cover. *Journal of Geophysical Research: Atmospheres* **121**(20), 913–11. doi: 10.1002/2016JD025606.

Jaynes E (2003) *Probability Theory*. Cambridge University Press. doi: 10.1017/CBO9780511790423.

Karner F, Obleitner F, Krismer T, Kohler J and Greuell W (2013) A decade of energy and mass balance investigations on the glacier Kongsvegen, Svalbard. *Journal of Geophysical Research: Atmospheres* **118**(10), 3986–4000. doi: 10.1029/2012JD018342.

Keetz LT and 7 others (2025) Inferring parameters in a complex land surface model by combining data assimilation and machine learning. *Journal of Advances in Modeling Earth Systems* **17**(6), e2024MS004542. doi: 10.1029/2024MS004542.

Kotarba AZ (2022) Impact of the revisit frequency on cloud climatology for CALIPSO, EarthCARE, Aeolus, and ICESat-2 satellite lidar missions. *Atmospheric Measurement Techniques* **15**(14), 4307–4322. doi: 10.5194/AMT-15-4307-2022.

Landmann JM, Künsch HR, Huss M, Ogier C, Kalisch M and Farinotti D (2021) Assimilating near-real-time mass balance stake readings into a model ensemble using a particle filter. *Cryosphere* **15**(11), 5017–5040. doi: 10.5194/tc-15-5017-2021.

Lenaerts JT, Drew Camron M, Wyburn-Powell CR and Kay JE (2020) Present-day and future Greenland Ice Sheet precipitation frequency from CloudSat observations and the Community Earth System Model. *Cryosphere* **14**(7), 2253–2265. doi: 10.5194/tc-14-2253-2020.

Lenaerts JTM, Medley B, van den Broeke MR and Wouters B (2019) Observing and modeling ice sheet surface mass balance. *Reviews of Geophysics* **57**(2), 376–420. doi: 10.1029/2018RG000622.

Lind S, Ingvaldsen RB and Furevik T (2018) Arctic warming hotspot in the northern Barents Sea linked to declining sea-ice import. *Nature Climate Change* **8**(7), 634–639. doi: 10.1038/s41558-018-0205-y.

MacKay DJC (2003) *Information Theory, Inference, and Learning Algorithms.*, Cambridge, UK: Cambridge University Press.

Magnusson J, Winstral A, Stordal AS, Essery R and Jonas T (2017) Improving physically based snow simulations by assimilating snow depths using the particle filter. *Water Resources Research* **53**(2), 1125–1143. doi: 10.1002/2016WR019092.

Malles JH, Marzeion B and Myers PG (2025) Freshwater input from glacier melt outside Greenland alters modeled northern high-latitude ocean circulation. *Earth System Dynamics* **16**(2), 347–377. doi: 10.5194/ESD-16-347-2025.

Malone AG, Doughty AM and MacAyeal DR (2019) Interannual climate variability helps define the mean state of glaciers. *Journal of Glaciology* **65**(251), 508–517. doi: 10.1017/JOG.2019.28.

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

Markus T and 24 others (2017) The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment* **190**, 260–273. doi: 10.1016/J.RSE.2016.12.029.

Marshall G, Rees W and Dowdeswell J (1993) Limitations imposed by cloud cover on multitemporal visible band satellite data sets from polar regions. *Annals of Glaciology* **17**, 113–120. doi: 10.3189/S0260305500012696.

Marzeion B and 16 others (2020) Partitioning the uncertainty of ensemble projections of global glacier mass change. *Earth's Future* **8**(7), e2019EF001470. doi: 10.1029/2019EF001470.

Marzeion B, Jarosch AH and Hofer M (2012) The cryosphere past and future sea-level change from the surface mass balance of glaciers. *Cryosphere* **6**, 1295–1322. doi: 10.5194/tc-6-1295-2012.

Masutani M and 9 others (2010) *Observing System Simulation Experiments*, Springer, pp.647–679. doi: 10.1007/978-3-540-74703-1_24.

Mazzolini M, Aalstad K, Alonso-González E, Westermann S and Treichler D (2024) Spatio-temporal snow data assimilation with the ICESat-2 laser altimeter. *EGUsphere* **19**, 3831–3848. doi: 10.5194/EGUSPHERE-2024-1404.

Milinski S, Maher N and Olonscheck D (2020) How large does a large ensemble need to be?. *Earth System Dynamics* **11**(4), 885–901. doi: 10.5194/ESD-11-885-2020.

Monin AS and Obukhov AM (1954) Basic laws of turbulent mixing in the surface layer of the atmosphere. *Transactions of the Academy of Sciences of the USSR, Geophysical Institute* **24**(151), 163–187.

Navari M, Margulis SA, Tedesco M, Fettweis X and van de Wal RS (2000–2014) Reanalysis surface mass balance of the Greenland ice sheet along K-transect. *Geophysical Research Letters* **48**(17), e2021GL094602. doi: 10.1029/2021GL094602.

Neuenschwander AL and Magruder LA (2019) Canopy and terrain height retrievals with ICESat-2: A first look. *Remote Sensing* **11**(14), 1721. doi: 10.3390/RS11141721.

Noël B and 10 others (2020) Low elevation of svalbard glaciers drives high mass loss variability. *Nature Communications* **11**(1), 1–8. doi: 10.1038/s41467-020-18356-1.

Nordli O, Przybylak R, Ogilvie AE and Isaksen K (2014) Long-term temperature trends and variability on Spitsbergen: The extended svalbard airport temperature series, 1898-2012. *Polar Research* **33**(1), 1–23. doi: 10.3402/polar.v33.21349.

Østby TI, Schuler TV and Westermann S (2014) Severe cloud contamination of MODIS land surface temperatures over an arctic ice cap, Svalbard. *Remote Sensing of Environment* **142**, 95–102. doi: 10.1016/J.RSE.2013.11.005.

Østby TI, Vikhamar Schuler T, Ove Hagen J, Hock R, Kohler J and Reijmer CH (2017) Diagnosing the decline in climatic mass balance of glaciers in Svalbard over 1957-2014. *Cryosphere* **11**(1), 191–215. doi: 10.5194/tc-11-191-2017.

Pirk N and 7 others (2022) Inferring surface energy fluxes using drone data assimilation in large eddy simulations. *Atmospheric Measurement Techniques* **15**(24), 7293–7314. doi: 10.5194/amt-15-7293-2022.

Pontes GM and Menviel L (2024) Weakening of the Atlantic Meridional Overturning Circulation driven by subarctic freshening since the mid-twentieth century. *Nature Geoscience* **17**(12), 1291–1298. doi: 10.1038/s41561-024-01568-1.

Pramanik A, Kohler J, Schuler TV, van Pelt W and Cohen L (2019) Comparison of snow accumulation events on two High-Arctic glaciers to model-derived and observed precipitation. *Polar Research* **38**, 3364. doi: 10.33265/POLAR.V38.3364.

Rantanen M and 7 others (2022) The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment* **3**(1), 1–10. doi: 10.1038/s43247-022-00498-3.

Raoult N, Charbit S, Dumas C, Maignan F, Ottlé C and Bastrikov V (2023) Improving modelled albedo over the Greenland ice sheet through parameter optimisation and modis snow albedo retrievals. *Cryosphere* **17**(7), 2705–2724. doi: 10.5194/tc-17-2705-2023.

Robinson G, Grooms I and Kleiber W (2018) Improving Particle Filter Performance by Smoothing Observations. *Monthly Weather Review* **146**(8), 2433–2446. doi: 10.1175/MWR-D-17-0349.1.

Rounce DR and 12 others (2023) Global glacier change in the 21st century: Every increase in temperature matters. *Science* **379**(6627), 78–83. doi: 10.1126/SCIENCE.ABO1324/SUPPL_FILE/SCIENCE.ABO1324_SM.PDF.

Rounce DR, Khurana T, Short MB, Hock R, Shean DE and Brinkerhoff DJ (2020) Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: Application to High Mountain Asia. *Journal of Glaciology* **66**(256), 175–187. doi: 10.1017/JOG.2019.91.

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

Sacher W and Bartello P (2008) Sampling errors in ensemble Kalman filtering. Part I: Theory. *Monthly Weather Review* **136**(8), 3035–3049. doi: 10.1175/2007MWR2323.1.

Sandven S and 6 others (2023) Sea ice remote sensing—recent developments in methods and climate data sets. *Surveys in Geophysics* **44**(5), 1653–1689. doi: 10.1007/S10712-023-09781-0/METRICS.

Sanz-Alonso D, Stuart A and Taeb A (2023) *Inverse Problems and Data Assimilation.*, London Math: Soc. Student Texts, Cambridge University Press. doi: 10.1017/9781009414319.

Schiller-Weiss I, Martin T and Schwarzkopf FU (2024) Emerging influence of enhanced greenland melting on boundary currents and deep convection regimes in the labrador and Irminger seas. *Geophysical Research Letters* **51**(9), e2024GL109022. doi: 10.1029/2024GL109022;PAGEGROUP:STRING:PUBLICATION.

Schmidt LS and 7 others (2017) The importance of accurate glacier albedo for estimates of surface mass balance on Vatnajökull: Evaluating the surface energy budget in a regional climate model with automatic weather station observations. *Cryosphere* **11**(4), 1665–1684. doi: 10.5194/TC-11-1665-2017.

Schmidt LS, Schuler TV, Thomas EE and Westermann S (2023) Meltwater runoff and glacier mass balance in the high Arctic: 1991-2022 simulations for Svalbard. *EGUsphere* **2023**, 1–32.

Schuster L, Rounce DR and Maussion F (2023) Glacier projections sensitivity to temperature-index model choices and calibration strategies. *Annals of Glaciology* **64**(92), 293–308. doi: 10.1017/AOG.2023.57.

Sjursen KH, Dunse T, Tambue A, Schuler TV and Andreassen LM (2023) Bayesian parameter estimation in glacier mass-balance modelling using observations with distinct temporal resolutions and uncertainties. *Journal of Glaciology* **69**(278), 1803–1822. doi: 10.1017/JOG.2023.62.

Stroeve J, Box JE, Gao F, Liang S, Nolin A and Schaaf C (2005) Accuracy assessment of the MODIS 16-day albedo product for snow: Comparisons with Greenland in situ measurements. *Remote Sensing of Environment* **94**(1), 46–60. doi: 10.1016/j.rse.2004.09.001.

Tang G and 9 others (2023) The impact of meteorological forcing uncertainty on hydrological modeling: A global analysis of cryosphere basins. *Water Resources Research* **59**(6), e2022WR033767. doi: 10.1029/2022WR033767.

van Leeuwen P (2009) Particle filtering in geophysical systems. *Monthly Weather Review* **137**, 4089–4114. doi: 10.1175/2009MWR2835.1.

van Leeuwen PJ and Evensen G (1996) Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review* **124**(12), 2898–2913. doi: 10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

van Leeuwen PJ (2020) A consistent interpretation of the stochastic version of the ensembleKalman filter. *Quarterly Journal of the Royal Meteorological Society* **146**(731), 2815–2825. doi: 10.1002/qj.3819.

van Pelt W and 10 others (2019) A long-term dataset of climatic mass balance, snow conditions, and runoff in Svalbard (1957-2018). *Cryosphere* **13**(9), 2259–2280. doi: 10.5194/tc-13-2259-2019.

Vionnet V and 7 others (2012) The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geoscientific Model Development* **5**(3), 773–791. doi: 10.5194/GMD-5-773-2012.

Wang S, Lu X, Cheng X, Li X, Peichl M and Mammarella I (2018) Limitations and challenges of MODIS-derived phenological metrics across different landscapes in pan-arctic regions. *Remote Sensing* **10**(11), 1784. doi: 10.3390/RS10111784.

Wei M and 6 others (2019) Quantifying the importance of interannual, interdecadal and multidecadal climate natural variabilities in the modulation of global warming rates. *Climate Dynamics* **53**(11), 6715–6727. doi: 10.1007/S00382-019-04955-2/FIGURES/5.

Westermann S and 23 others (2023) The CryoGrid community model (version 1.0) - a multi-physics toolbox for climate-driven simulations in the terrestrial cryosphere. *Geoscientific Model Development* **16**(9), 2607–2647. doi: 10.5194/GMD-16-2607-2023.

Yang C and 15 others (2021) Sea surface temperature intercomparison in the framework of the Copernicus Climate Change Service (C3S). *Journal of Climate* **34**(13), 5257–5283. doi: 10.1175/JCLI-D-20-0793.1.

Ye F, Cheng Q, Hao W, Hu A and Liang D (2024) Unveiling glacier mass balance: Albedo Aggregation insights for Austrian and Norwegian Glaciers. *Remote Sensing*, **16**(11), 1914. doi: 10.3390/RS16111914.

Zemp M and 14 others (2019) Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**(7752), 382–386. doi: 10.1038/S41586-019-1071-0.

Zhang Q and 7 others (2023) Oceanic climate changes threaten the sustainability of Asia's water tower. *Nature* **615**(7950), 87–93. doi: 10.1038/s41586-022-05643-8.