



Article

Cite this article: Moss G, Višnjević V, Eisen O, Oraschewski FM, Schröder C, Macke JH, Drews R (2025) Simulation-based inference of surface accumulation and basal melt rates of an Antarctic ice shelf from isochronal layers. *Journal of Glaciology* **71**, e44, 1–21. <https://doi.org/10.1017/jog.2025.13>

Received: 18 November 2024

Revised: 27 January 2025

Accepted: 11 February 2025

Keywords:

Bayesian inference; Ice shelves; Melt - basal; Mass-balance reconstruction; Machine learning; Simulation-based inference

Corresponding author: Guy Moss;
Email: guy.moss@uni-tuebingen.de

†Joint supervision

Simulation-based inference of surface accumulation and basal melt rates of an Antarctic ice shelf from isochronal layers

Guy Moss¹ , Vjeran Višnjević² , Olaf Eisen^{3,4} , Falk M. Oraschewski² ,
Cornelius Schröder^{1,†} , Jakob H. Macke^{1,5,†}  and Reinhard Drews^{2,†} 

¹Machine Learning in Science, University of Tübingen and Tübingen AI Center, Tübingen, Germany; ²Department of Geosciences, University of Tübingen, Tübingen, Germany; ³Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar und Meeresforschung, Bremerhaven, Germany; ⁴Faculty of Geosciences, University of Bremen, Bremen, Germany and ⁵Max Planck Institute for Intelligent Systems, Tübingen, Germany

Abstract

The ice shelves buttressing the Antarctic ice sheet determine the rate of ice-discharge into the surrounding oceans. Their geometry and buttressing strength are influenced by the local surface accumulation and basal melt rates, governed by atmospheric and oceanic conditions. Contemporary methods quantify one of these rates, but typically not both. Moreover, information about these rates is only available for recent time periods, reaching at most a few decades back since measurements are available. We present a new method to simultaneously infer the surface accumulation and basal melt rates averaged over decadal and centennial timescales. We infer the spatial dependence of these rates along flow line transects using internal stratigraphy observed by radars, using a kinematic forward model of internal stratigraphy. We solve the inverse problem using simulation-based inference (SBI). SBI performs Bayesian inference by training neural networks on simulations of the forward model to approximate the posterior distribution, therefore also quantifying uncertainties over the inferred parameters. We validate our method on a synthetic example, and apply it to Ekström Ice Shelf, Antarctica, for which independent validation data are available. We obtain posterior distributions of surface accumulation and basal melt averaging over up to 200 years before 2022.

1. Introduction

The majority of the Antarctic ice sheet is buttressed by floating ice shelves (Bindschadler and others, 2011) which provide large contact areas for ice–ocean interactions. Approximately half of the ice shelves' total mass loss is attributed to ocean-induced melting at the underside of ice shelves (Depoorter and others, 2013), and its spatiotemporal variability imprints ice flow dynamics farther upstream (Reese and others, 2017; Gudmundsson and others, 2019). Consequently, ice flow and ocean models need to be coupled for future projections; frameworks (Goldberg and others, 2019; Gladstone and others, 2021), parameterizations (Burgard and others, 2022; Goldberg and Holland, 2022) and benchmarks (Asay-Davis and others, 2016) for this task have been developed. Similarly, the local snow accumulation is influenced by atmospheric conditions and is crucial in determining ice shelf thickness (Winkelmann and others, 2012). As a result, ice flow models are also coupled to climate models for future projections (Goelzer and others, 2016; Pattyn and others, 2017). It is crucial to confront ice flow models with observations to validate them and investigate their ability to explain observed phenomena. Here, we present a new method that infers surface accumulation (also known as 'surface mass balance' (Lenaerts and others, 2019)) and basal melt rates (collectively, the mass-balance parameters) from the ice shelves' internal stratigraphy, which can be routinely mapped by radio-echo sounding.

Typically, surface accumulation is the more accessible mass-balance parameter (Eisen and others, 2008); it can be measured in situ using stake farms and can also be derived from multiple firn cores (Lenaerts and others, 2019). Many of these observations validate atmospheric models such as RACMO (van Wessem and others, 2018) and MAR (Gallée and Schayes, 1994; Agosta and others, 2019), which estimate surface accumulation on 35 km grids (Lenaerts and others, 2019) (with few locations being estimated at a higher resolution of 5.5 km). Estimating the basal melt is more challenging and is typically dependent on knowledge of surface accumulation. For example, estimates of surface accumulation have been used along with mass conservation arguments to estimate basal melt (Neckel and others, 2012; Depoorter and others, 2013; Berger and others, 2017; Adusumilli and others, 2020). These approaches have provided Antarctic-wide time series of the last few decades of basal melt rates (Adusumilli and others, 2020). The spatial resolution is currently limited to the kilometer scale, which may miss fine grained processes occurring within ice shelf channels (Drews, 2015; Marsh and others, 2016) or near basal terraces



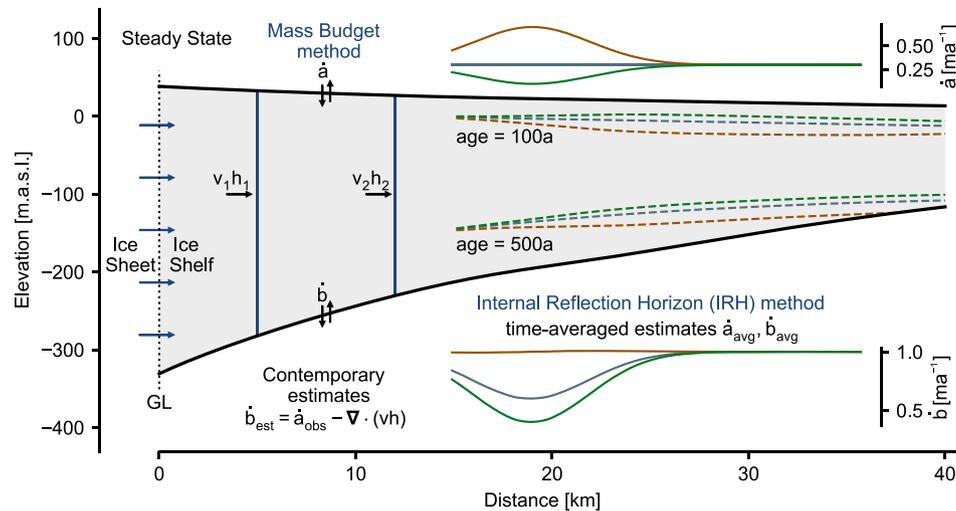


Figure 1. Estimation of mass-balance parameters from a steady-state ice shelf with two methods. The Eulerian Mass Budget method (left) detects the difference of surface accumulation and basal melt within two flux gates (blue vertical lines) by considering flux divergence $\nabla \cdot (\mathbf{v}h)$. Often, the basal melt rates \hat{b} are inferred assuming that the surface accumulation (\hat{a}_{obs}) is known. In the internal reflection horizon (IRH) method, we are given information on the internal stratigraphy of the shelf. This information is used to separate the known total mass balance into individual estimates of surface accumulation and basal melt (\hat{a}_{avg} , \hat{b}_{avg} respectively). These estimates correspond to the time-averaged value over the age of the IRH to the present. The inset plots show different surface accumulation and basal melt parameterizations which give rise to the same total mass balance and overall shape of the ice shelf, but different internal stratigraphy.

(Dutrieux and others, 2014). Measurements of basal melt which are independent of the surface accumulation are also available, but typically only on short temporal scales, for example, with time-lapse radar measurements of ice thickness change (Zeising and others, 2022). Using phase-coherent data acquisition, these measurements can disentangle the observed thickness change into strain thinning and basal melt (Nicholls and others, 2015). This has provided much insights, e.g., in terms of relevant tidal (Sun and others, 2019) and seasonal timescales (Vankova and Nicholls, 2022).

Here, we investigate to what extent the radar-imaged isochronal ice stratigraphy (Eisen and others, 2004) can provide additional information for inferring mass-balance parameters. On grounded ice, radar-imaged internal reflection horizons (IRHs) have been used in multiple ways, for example, to infer the surface accumulation history (Waddington and others, 2007; MacGregor and others, 2009; Catania and others, 2010; Steen-Larsen and others, 2010; Wolovick and others, 2021; Theofilopoulos and Born, 2023), velocity patterns of the ice flow (Eisen, 2008; Holschuh and others, 2017), ice-rise evolution (Drews and others, 2015; Henry and others, 2023) or large-scale model calibration (Leysinger Vieli and others, 2011; Sutter and others, 2021). On ice shelves, surface accumulation rate can also be derived from the radar-measured shallow stratigraphy (Pratap and others, 2022), but not from intermediate depths and below where the stratigraphy is also influenced by basal melt and ice flow. The stratigraphy of ice shelves differs for various combinations of surface accumulation and basal melt rates (Visnjevic and others, 2022). This suggests that given an ice flow model of the internal stratigraphy that accounts for the surface accumulation and basal melt rates, we can use observed IRHs to recover the surface accumulation and basal melt rate histories (Fig. 1). Thus, our goal is to solve the inverse problem of inferring the surface accumulation and basal melt rates that can explain the observed IRHs under the physical constraints of the ice flow model.

Inverse problems, also known as *inversion*, *data assimilation* or *inference* problems in the literature, denote the task of finding the

model parameters that are compatible both with empirical observations and prior knowledge. This problem is widespread in the geosciences, e.g., in hydrogeology (Linde and others, 2015), seismology (Symes, 2009) or in climate science (Tebaldi and Sansó, 2008). Bayesian inference provides a powerful framework for solving inference problems, but conventional Bayesian approaches are restricted to models for which the so-called ‘likelihood function’ is tractable. A tractable likelihood function is one that can be efficiently evaluated (see Appendix C.1 for examples of tractable and intractable likelihood functions). However, this is not the case in our setup. We therefore use simulation-based inference (SBI, Papamakarios and Murray (2016); Lueckmann and others (2017); Cranmer and others (2020)) to solve the inverse problem presented in this work. In SBI, we evaluate the forward model under different values of the model’s parameters from a prior distribution. We use the resulting simulated dataset to train a neural network that performs conditional density estimation. In the neural posterior estimation (NPE) variant, the network approximates the Bayesian posterior distribution. A key advantage of NPE is the amortization of simulation cost. An amortized inference framework is one that, once trained, can be instantly applied to find the posterior distribution for any new observation without requiring more simulations or training. Importantly, SBI does not require the forward model to be differentiable and can also work with ‘blackbox’ models. Therefore, our approach can be extended to a variety of preexisting forward models. To the authors’ knowledge, this work is the first application of SBI in glaciology, but we note that it has already been applied in other geoscientific disciplines such as geothermics (Omagbon and others, 2021), hydrogeology (Allgeier and Cirpka, 2023), hydrology (Hull and others, 2022) and molecular ecology (Overcast and others, 2021).

In this study, we consider steady-state ice shelves and IRHs in the local meteoric ice (LMI) body of ice shelves (Das and others, 2020). This work is a test case for inferring atmospheric and oceanographic boundary conditions from the ice stratigraphy with a novel inference technique that provides uncertainty estimates.

Our approach can be transferred to other ice flow regimes (e.g. flank flow on grounded ice) where similar scientific questions can be explored. Our approach can similarly be adapted to ice shelves exhibiting marine ice formation. Moreover, the isochronal stratigraphy of ice shelves and ice sheets (including the neighboring ice rises) is currently the only archive of surface accumulation and basal melt over the past hundreds of years. Our approach is capable of testing this archive. Thus, this study provides one link between observational initiatives (such as AntArchitecture, Bingham and others, 2024) for Antarctica-wide internal stratigraphy datasets and the modeling community.

The paper is structured as follows: In Section 2, we describe our forward model of the internal stratigraphy of an ice shelf and introduce our inference approach. In Section 3, we detail the synthetic ice shelf construction. We also present the results of inferring the mass-balance parameters from this synthetic stratigraphy and compare the posterior distribution to a known ground truth. In Section 4, we describe the setting of the Ekström Ice Shelf (EIS) and the dataset of observed IRHs along the central flow line transect. We then provide the results of our inference framework and compare them to independent measurements of surface accumulation uniquely available in this location for the periods 1996–2005 and 2014–23. In Section 5, we interpret our results and evaluate our approach. We finally conclude and discuss future perspective in Section 6.

2. Methodology

2.1. Forward model

We denote spatially varying parameters as functions, e.g. $\dot{a}(x)$ or at times \dot{a} for brevity, while bold-faced characters denote the discretized values of this function on a specified grid, e.g. $\dot{\mathbf{a}} = [\dot{a}(x_1), \dots, \dot{a}(x_n)]^T$.

2.1.1. Ice flow model

We model ice shelves using the shallow shelf approximation (SSA) (Morland, 1984). Throughout this study, we consider ice shelves in steady state. Consequently, the ice surface s , base f , thickness $h = s - f$ and velocity v are all fixed throughout our simulations. We assume plug flow for the ice shelf regime, meaning that the horizontal velocity profile does not change in the vertical direction z . These assumptions results in the mass-balance condition

$$\nabla \cdot (hv) = \dot{m}, \quad (1)$$

where hv is the total mass flux, $\nabla \cdot$ is the divergence operator and $\dot{m} = \dot{a} - \dot{b}$ is the total mass-balance rate. Here we use the convention that the surface accumulation rate \dot{a} is positive for mass gain of the ice shelf and the basal melt rate \dot{b} is positive for mass loss. In this exploratory study, we focus on flow lines. We parameterize our domain such that x denotes the distance along the flow line, and v_x now denotes the velocity parallel to the flow line. The two-dimensional (2-D) geometry is only valid for observations located on flow lines and in the absence of lateral compression and extension. While the former is approximately true in our case, the latter is unrealistic for most Antarctic ice shelves. To account for ice flux into or out of our modeling domain, we, therefore, include the ice flux component normal to the flow line as an additional, spatially variable term to the total mass-balance rate \dot{m} (Appendix A). We test the validity of this approach in a 2-D synthetic example (Section 3) that includes a spatially variable total mass balance

and lateral compression. For the real-world scenario, we estimate the the normal ice flux component from satellite velocities.

We seek to predict the steady-state internal stratigraphy for a given flow line and possible surface accumulation and basal melting rate profiles. We define the internal stratigraphy to be a set of isochronal layer elevations $\{e_1(x), \dots, e_l(x)\}$, with $f(x) \leq e_1(x) \leq e_2(x) \leq \dots \leq e_l(x) \leq s(x)$. One approach to calculate the internal stratigraphy uses the SSA expression for the vertical component of the velocity (Greve and Blatter, 2009) to have a fully specified velocity field. This can then be used to calculate the age field $\mathcal{A}(x, z)$ of the shelf. Contours of constant age (isochrones) then define the internal stratigraphy. However, these methods suffer from numerical diffusion and can be computationally expensive (Visnjevic and others, 2022).

The computational efficiency of the forward model is crucial for our inference method, as we need to evaluate the forward model many times. As a result, we opt instead to use an implementation of the tracer method (Born, 2017; Born and Robinson, 2021). The model is seeded with vertical segments each with a thickness profile $\{h_1(x), \dots, h_l(x)\}$, such that the sum matches the ice geometry $\sum_{l=1}^L h_l(x) = h(x)$. The horizontal velocity $v_x(x)$ is used to advect mass within segments and to thin or thicken the segments as a function of the prescribed strain rates. The accumulation $\dot{a}(x)$ and melt $\dot{b}(x)$ rates are used to add new segments or take away mass from the two boundary segments at the top and bottom of the shelf respectively. The (isochronal) layer elevations are then the boundaries between our modeled segments. We use the convention that e_l corresponds to the top of segment l , which can be calculated using the cumulative thicknesses of the segments below,

$$e_l(x) = f(x) + \sum_{l'=1}^l h_{l'}(x). \quad (2)$$

In our simulations, we used a high temporal resolution of one isochronal layer per year. Despite the high resolution, the layer tracing method allows for determining the internal stratigraphy in a computationally efficient manner. For the domains and timescales considered in our study, the complete forward model can be evaluated on the order of 60 s on a single CPU core, enabling the application of SBI methods (see Appendix E for details).

To uniquely determine the layer thicknesses in such a scheme, we need to specify the boundary conditions on the layer thicknesses h_l at the inflow boundary $x = 0$ (here corresponding to the grounding line). The true boundary conditions are typically not known. However, the stratigraphy in a large part of the domain is still independent of the boundary conditions. This zone corresponds to the LMI body of ice shelves (Das and others, 2020). When inferring from observed stratigraphy data, we use only data within the LMI body. We detail our model of the LMI body in Appendix A.

2.1.2. Noise model

The ice flow model predicts isochronal layers with varying depth over spatial scales of kilometers. Observed IRHs, however, also show variability on sub-kilometer scales. This systematic model-data misfit is caused by errors in input datasets (such as surface velocity, geometry), coarse resolution of the forward model and omission of higher order processes that are not included in the forward model, such as the effect of rheology. For inference, it is important that the predicted isochrones have consistent statistical properties with the observed IRHs. This is achieved by the definition of an appropriate noise model.

The ice flow model predicts isochronal layer elevations $\{\mathbf{e}_1(\mathbf{x}), \dots, \mathbf{e}_L(\mathbf{x})\}$ on a fixed grid $\mathbf{x} \in \mathbb{R}^N$ where N is the number of grid points. Guided by empirical observations, the noise model should have the property that the errors of different modeled layers l at different depths are spatially correlated and amplified for deeper layers. We, therefore, define a layer-wise noise model as the product of an x -dependent baseline noise function and a z -dependent vertical amplification factor. More precisely, the additive noise $\delta_l \in \mathbb{R}^N$ of layer l is defined as

$$\delta_l = \epsilon \odot \mathbf{T}(\mathbf{e}_l), \quad (3)$$

where $\epsilon = [\epsilon_1, \dots, \epsilon_N]^\top$ is a \mathbf{x} -dependent noise profile, which is shared for all layers, $\mathbf{T}(\mathbf{e}_l) = [T(e_{l,1}), \dots, T(e_{l,N})]^\top$ is a deterministic function of elevation (increasing with depth), and \odot denotes an element-wise product. The vertical scaling $\mathbf{T}(\cdot)$ mimics uncertainties in the travel time-to-depth conversion which depend on the density $\rho(z)$. Here, this is done using $\rho(z)$ as in Drews and others (2016) and an empirical density-permittivity relation (Looyenga, 1965) to calculate the radio-wave speed $c(z)$. This results in the factor

$$T(z) = \int_z^s \frac{dz'}{c(z')}, \quad (4)$$

which we then discretize on the set of layer elevations.

The sub-kilometer variability of the observed IRHs are modeled with power spectral densities ϵ :

$$\epsilon = A_\epsilon \sum_{n=1}^N \sqrt{\exp^{\beta_n}} \cos(2\pi\omega_n + \chi_n), \quad (5)$$

where the log power spectral densities β_n and offsets χ_n are randomly sampled from normal and uniform distributions respectively: $\beta_n \sim \mathcal{N}(\mu_{\beta_n}, \sigma_{\beta_n}^2)$ and $\chi_n \sim U(-\pi, \pi)$. The frequencies ω_n are the corresponding Fourier frequencies of the simulation grid \mathbf{x} and A_ϵ is a global scale factor (set to 4×10^{-10}). In the synthetic ice shelf (Section 3), we define the distribution of the log power densities using $\sigma_{\beta_n}^2 = 0.5$ and

$$\mu_{\beta_n} = -8(1 - \exp^{-200\omega_n}). \quad (6)$$

For EIS, the distribution means μ_{β_n} and variances $\sigma_{\beta_n}^2$ were calibrated given the observed IRHs on a separate set of calibration simulations (full details in Appendix B). We emphasize that this representation of the noise model is a choice—we define a mathematical model of the mismatch, rather than model a physical effect directly. Thus, other choices are possible. We choose this representation of the noise model for its flexibility and interpretability.

By combining the ice flow model with the empirically guided noise model, we have arrived at a physically motivated forward model to sample a plausible observed internal stratigraphy of an ice shelf from the mass-balance rate parameters \dot{a} and \dot{b} .

2.2. Inference

Having established the forward model, we arrive at the *inverse problem* of finding the surface accumulation and basal melt rates that best explain the observed internal stratigraphy. We use Bayes theorem with model parameters θ and outcomes X :

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}. \quad (7)$$

Here, $p(\theta|X)$ is the posterior distribution of the parameters given a particular outcome X , $p(X|\theta)$ is the likelihood function of the

model, $p(\theta)$ is the prior distribution encoding our existing knowledge on the plausible values of θ and $p(X)$ is the model evidence. The goal of Bayesian inference is to find the posterior $p(\theta|X_o)$, where X_o is *observed data* which has the same form as X , but is measured, instead of simulated.

2.2.1. Simulation-based inference

It is generally not possible to analytically solve for the Bayesian posterior distribution (Eqn (7)), as the evidence term $p(X)$ cannot be computed. Approximate methods exist to solve Eqn (7) using knowledge of only the likelihood function and prior distribution. In this work, we deploy SBI, an approximate Bayesian inference and *likelihood-free* approach, using only samples from our forward model. In SBI, we use artificial neural networks to approximate conditional probability distributions. While there exist different variants of SBI which target either the likelihood $p(X|\theta)$ or the likelihood ratio (see Cranmer and others (2020) for an overview), we focus on NPE, which approximates the posterior distribution directly (Papamakarios and Murray, 2016; Greenberg and others, 2019).

In NPE, we generate a training dataset $\{(\theta_k, X_k)\}_{k=1}^K$ (Fig. 2) by sampling parameters from the prior $\theta_k \sim p(\theta)$ and sampling from the forward model $X_k \sim p(X|\theta_k)$. To approximate the posterior distribution, a variational family of distributions $q_\phi(\theta|X)$ is typically defined in terms of a neural network with learnable weights ϕ . We represent q_ϕ as a normalizing flow (Durkan and others, 2019; Kobyzev and others, 2019; Papamakarios and others, 2019a). Normalizing flows are flexible generative models, which, once trained, can be used either to sample or evaluate the (log-) probability density function of the conditional distribution $q_\phi(\theta|X)$, for any outcome X in the support of the training dataset. We provide a brief description of normalizing flows in Appendix C.2 and refer the reader to Papamakarios and others (2019a) for a review.

In NPE, the neural network is trained by minimizing the expected negative log-probability

$$\mathcal{L}(\phi) = \mathbb{E}_{\theta_k \sim p(\theta), X_k \sim p(X|\theta_k)}[-\log q_\phi(\theta_k|X_k)] \quad (8)$$

on the training dataset. More intuitively, this loss seeks to maximize the probability assigned to the training data. It can be trivially shown that minimizing this loss is equivalent to minimizing the (forward) Kullback–Leibler (KL) divergence between the variational distribution and the true posterior distribution (see C.3).

It has been shown that, if there exists a set of weights ϕ such that $q_\phi(\theta|X)$ is the true posterior distribution, and in the limit of infinite training samples $K \rightarrow \infty$, the minimum of the loss in Eqn (8) is reached when $q_\phi(\theta|X) = p(\theta|X)$ for all X —i.e. when our estimated distribution matches the true posterior (see Proposition 1 of Papamakarios and Murray 2016 for full statement and proof).

We additionally make use of an *embedding network*, which are commonly used in SBI workflows to improve performance. Embedding networks learn *summary statistics* $Y(X)$, which are lower-dimensional representations of the outcomes X . Using the embedding $Y(X)$ as an input to the normalizing flow instead of X itself reduces the model complexity. The embedding network is trained jointly with the normalizing flow. In our setting, X_k are spatially varying IRH elevations, and so we choose a 1-dimensional (1-D) convolutional neural network as our embedding net, resulting in 50-dimensional embeddings on which the posterior network is conditioned (full details in Appendix D). Throughout this work, we use the `sbi` package for Python (Tejero-Cantero and others, 2020) to perform inference.

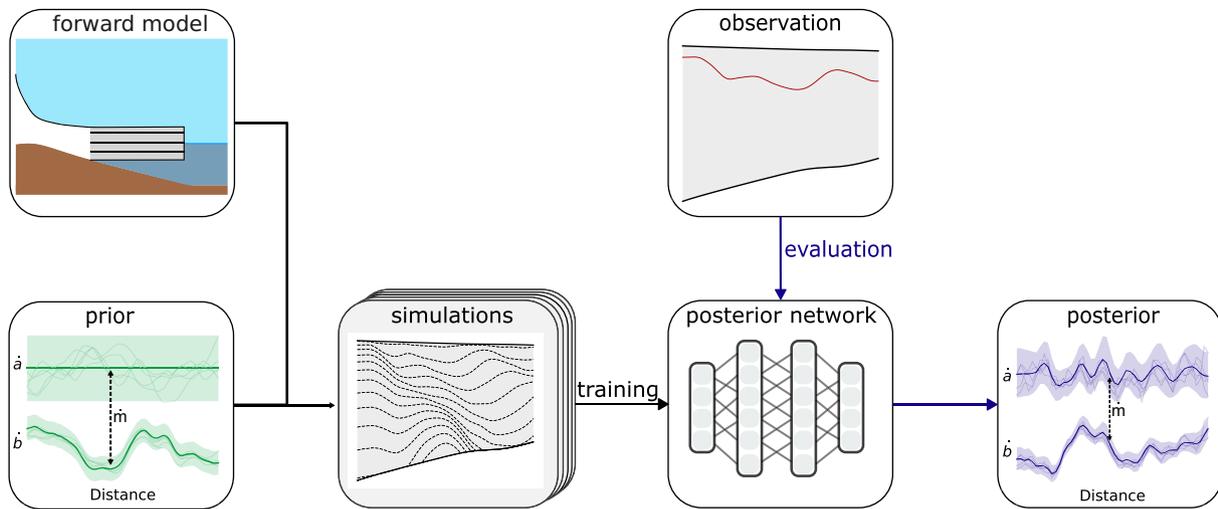


Figure 2. Simulation-based inference workflow. SBI has two primary phases: training and evaluation. In the training phase, accumulation rates are randomly sampled from a prior distribution, the corresponding basal melt rates are obtained using total mass balance, and the resulting internal stratigraphy is calculated using the forward model. These simulations from the prior are used to train a neural network which parameterizes conditional distributions. In the evaluation phase, the trained network is conditioned on the observed IRH and outputs the Bayesian posterior distribution over the parameters (without any additional calls to the forward model).

2.2.2. Definitions of model parameters, outputs and observations

We define $\theta = \hat{\mathbf{a}}_{\text{inf}} = [\hat{a}_i, \dots, \hat{a}_{i_j}]^\top$, the values of the surface accumulation rate on a discretized grid $\tilde{\mathbf{x}}$. In our experiment, we choose the number of inference grid points $J = 50$ as a compromise between computational complexity of the inference problem while still inferring accumulation rate at a high resolution of ~ 2.5 km. This is smaller than the discretized grid \mathbf{x} we use for our simulations, which has 500 gridpoints in our experiments. In practice, we take $\tilde{\mathbf{x}}$ to be a regularly spaced subset of \mathbf{x} , so that $\hat{\mathbf{a}}_{\text{inf}}$ can also be taken as a subset of \mathbf{a} . However, $\tilde{\mathbf{x}}$ can be any discretization of the flow line and need not be a subset of \mathbf{x} . Furthermore, despite defining θ to only represent the surface accumulation, any inference of the surface accumulation rate automatically extends to inference of the basal melt rates. This is because for any probability distribution $q(\hat{\mathbf{a}}_{\text{inf}})$, the total mass-balance relationship implies that $\hat{\mathbf{b}}_{\text{inf}} \sim q(\hat{\mathbf{a}}_{\text{inf}} - \hat{\mathbf{m}}_{\text{inf}})$, where $\hat{\mathbf{b}}_{\text{inf}}, \hat{\mathbf{m}}_{\text{inf}}$ are the respective discretizations of $\hat{\mathbf{b}}, \hat{\mathbf{m}}$ onto $\tilde{\mathbf{x}}$.

We now turn to describing the observation, X_o , and forward model outcomes X_k . The observed data are a set of different IRHs, $\{\mathbf{e}_m(\mathbf{x})\}_{m=1}^M$, where $e_m(x_i)$ is the elevation of the m th IRH in our dataset at grid position i . The IRH elevations need not and typically are not observed at the same locations as the simulation gridpoints; and so we first interpolate the IRH elevations onto the simulation grid \mathbf{x} using linear spline interpolation (as implemented in `scipy` (Virtanen and others, 2020)). Therefore, we assume $\{\mathbf{e}_m(\mathbf{x})\}_{m=1}^M$ is already defined on \mathbf{x} . One reasonable choice is to define the observation X_o as the entire set of all measured IRH elevations. However, in our work, we choose to separately infer the mass balance from each IRH in our observed dataset. This choice has two advantages: first, ordering IRHs by depth also corresponds to their reverse age order, with the oldest IRHs being the deepest. Thus, inferring the surface accumulation and basal melt rates for deeper IRHs corresponds to inferring the average rates over longer periods of time. By comparing the inferred mass-balance parameters obtained with different IRHs, we can reason whether or not our steady-state assumption is valid. The second advantage is practical—we seek

a consistent representation of the observations that can be applied across ice shelves. Given a different ice shelf, there will be a different number of IRHs at different depths. Therefore, the embedding net for these data will have to have a different architecture for each ice shelf. In our representation, the embedding net can always be a 1-D convolutional net, as the observations are always 1-D vectors.

Thus, given a dataset of M observed IRHs, we have M inference problems to solve, where each observation corresponds to one IRH. It is, therefore, reasonable to take one isochronal layer of the simulated stratigraphy as the output of the forward model. For the m th inference problem, we define the outcome of the forward model as the isochronal layer \mathbf{e}_l that is closest to IRH m (in the mean square sense). More precisely, for inference problem m and simulation k , we define the observation of the forward model to be $X_k^m = \mathbf{e}_l(\mathbf{x}_{i \geq i(m)})$, where

$$l^* = \arg \min_l \|\mathbf{e}_l(\mathbf{x}_{i \geq i(m)}) - \mathbf{e}_m(\mathbf{x}_{i \geq i(m)})\|_2^2. \tag{9}$$

Here, $i(m)$ is the index of the boundary of the LMI body for IRH $\mathbf{e}_m(\mathbf{x})$. For $i < i(m)$, the IRH $e_m(x_i)$ is outside the LMI body and for $i \geq i(m)$ within the LMI body (see Appendix B for details). We further define $\mathbf{x}_{i \geq i(m)} = [x_{i(m)}, x_{i(m)+1}, \dots, x_N]^\top$ as the restriction of the gridpoints \mathbf{x} to within the LMI body of $\mathbf{e}_m(\mathbf{x})$. We correspondingly set the observation for IRH m to $X_o^m = \mathbf{e}_m(\mathbf{x}_{i \geq i(m)})$. Our choice to select the simulated isochronal layer that most closely matches the IRH is due to the true age of the IRH being unknown. This introduces degeneracy into the forward model—two simulations with different surface accumulation and basal melt rate parameterizations can produce isochronal layers with a similar geometry but different ages. It is, therefore, important to define the prior distribution appropriately, which we do in the following section.

2.2.3. Choice of prior distribution

We aim to approximate the posterior distribution

$$p(\hat{\mathbf{a}}|X_o^m) \propto p(X_o^m|\hat{\mathbf{a}})p(\hat{\mathbf{a}}). \tag{10}$$

The likelihood $p(X_o^m|\hat{\mathbf{a}})$ is not tractable but can be sampled from using the forward model. To specify the prior, we use the long-term snow accumulation observations of the Neumayer stations

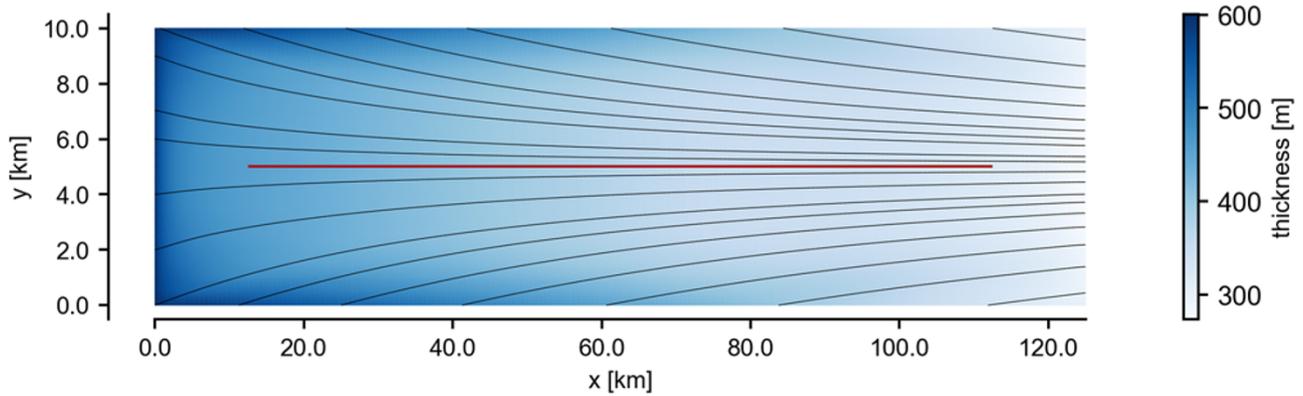


Figure 3. Two-dimensional flow tube domain setup for the synthetic example. Map view of the simulated ice shelf's surface. Flow lines (gray lines) converge to the central flow line (red). Color indicates ice thickness. The input variables for the internal stratigraphy model are evaluated on the central flow line.

(Wesche and others, 2016; Wesche and Regnery, 2022) over more than 30 years to define an empirically motivated prior for EIS, which we also use for the synthetic ice shelf. We first assume that localized surface melt ($\dot{a} < 0$) is possible, but rare. We also observe that average rate of accumulation is $\sim 0.5 \text{ m a}^{-1}$, and that the accumulation rate is almost everywhere under 2 m a^{-1} . Finally, we take the accumulation rate to vary smoothly in space. We define a prior distribution that satisfies these criteria, while still allowing for a broad range of surface accumulation rate profiles. We define the following generative process for $\dot{\mathbf{a}}$: first, we draw a sample $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$ from a Gaussian process with mean function $\mu = 0$ and a Matérn kernel with a Matérn- ν of 2.5 and a length scale of 2500 m (Rasmussen and Williams, 2005). We then independently sample an offset $\mu_{\text{off}} \sim \mathcal{N}(0.5, 0.25^2)$ and scale $\sigma_{\text{sc}} \sim U(0.1, 0.3)$ parameter. Finally, we set $\dot{\mathbf{a}} = \sigma_{\text{sc}} \boldsymbol{\alpha} + \mu_{\text{off}} \mathbf{1}$. We inspect the implicit prior this defines over the basal melt rates in Section 4.3.

Defining the prior in this way is sufficiently expressive to capture numerous accumulation rate profiles, while also restricting the samples to conform to empirical knowledge. Additionally, the prior is shared for all M inference problems we have defined, and one evaluation of the forward model provides an observation X_k^m for each of the inference problems. Thus, the same training dataset can be used for all posterior networks in our SBI approach, significantly reducing the computational costs.

3. Synthetic test case

Before we apply the presented workflow to EIS, we showcase its applicability in a synthetic test case in which all parameters are known.

3.1. Configuration of shelf and flow line

We test our workflow on a 2-D flow tube geometry from which we extract a flow line to infer the prescribed surface and basal accumulation rates as done later in the case of EIS. The flow tube is modeled using *icepack* (Shapero and others, 2021) on a grid $L_x = 125 \text{ km} \times L_y = 10 \text{ km}$, with the along-flow direction x and across-flow direction y . We prescribe a Dirichlet boundary condition at the inflow and lateral boundaries, with a constant thickness of h_0 , and a constant along-flow velocity of v_{0x} . The outflow boundary is set to be a static calving front. We initialize with a zero centered, longitudinally symmetric across-flow velocity v_{0y} on the lateral boundaries, resulting in a flow field that has convergence

(i.e. mass input) on the center flow line. We prescribe a spatially variable total mass balance \dot{m} : In our experiments, we set $v_{0x} = 100 \text{ m a}^{-1}$, $v_{0y} = \pm 20 \text{ m a}^{-1}$ at $y = 0$ and $y = L_y$, respectively,

$$\dot{m} = -0.6 - 0.05 \frac{x}{L_x} + 0.3 \exp\left(-\left(\frac{x - 0.7L_x}{0.1L_x}\right)^2\right). \quad (11)$$

We let the geometry evolve under the SSA approximation until steady state is reached. From the steady-state ice shelf, we choose a discretization of the central flow line, \mathbf{x} , and extract the relevant variables along this flow line to define the internal stratigraphy model (Fig. 3). The numerical values for additional parameters for the spin-up are given in Appendix D. The variables we need are the surface \mathbf{s} and base \mathbf{f} elevations, the along-flow velocities \mathbf{v}_x , and the along- and across-flow flux divergences $d(\mathbf{v}_x \mathbf{h})/dx$, $d(\mathbf{v}_y \mathbf{h})/dy$. These define the total mass balance, since:

$$\dot{\mathbf{a}} - \dot{\mathbf{b}} = \frac{d(\mathbf{v}_x \mathbf{h})}{dx} + \frac{d(\mathbf{v}_y \mathbf{h})}{dy}. \quad (12)$$

We then solve the inverse problem which accounts in this case for mass gain through lateral compression. We choose a random sample from the prior distribution as the ground truth, $\dot{\mathbf{a}}_{\text{GT}}$, from which $\dot{\mathbf{b}}_{\text{GT}}$ follows accordingly. The forward model is then sampled to obtain a set of ground truth layer elevations, $\mathbf{e}_0(\mathbf{x})$. From these layer elevations, we choose to perform inference for four layers of ages 50, 100 and 150, and 300 years (labeled 1–4 in ascending order of age). These ages roughly correspond to the range of ages of the IRHs that we expect to observe on ice shelves.

3.2. Inference results

We evaluate the trained neural posterior network on the ground truth isochronal layer of age 50 years. The inferred posterior mean for the surface accumulation rate parameter is close to the ground truth accumulation rate (Fig. 4a,c) with the ground truth lying within the 95% confidence intervals of the posterior distribution.

Next, we evaluate the forward model on samples from the posterior (and prior) distribution to get the respective *predictive distributions*. The prior predictive distribution (Fig. 4b, green) is the distribution over the internal layers generated by simulating the forward model with mass-balance parameters drawn from the prior distribution. The posterior predictive distribution (Fig. 4b, blue) is defined similarly by simulating with mass-balance parameters from the posterior distribution. The posterior predictive

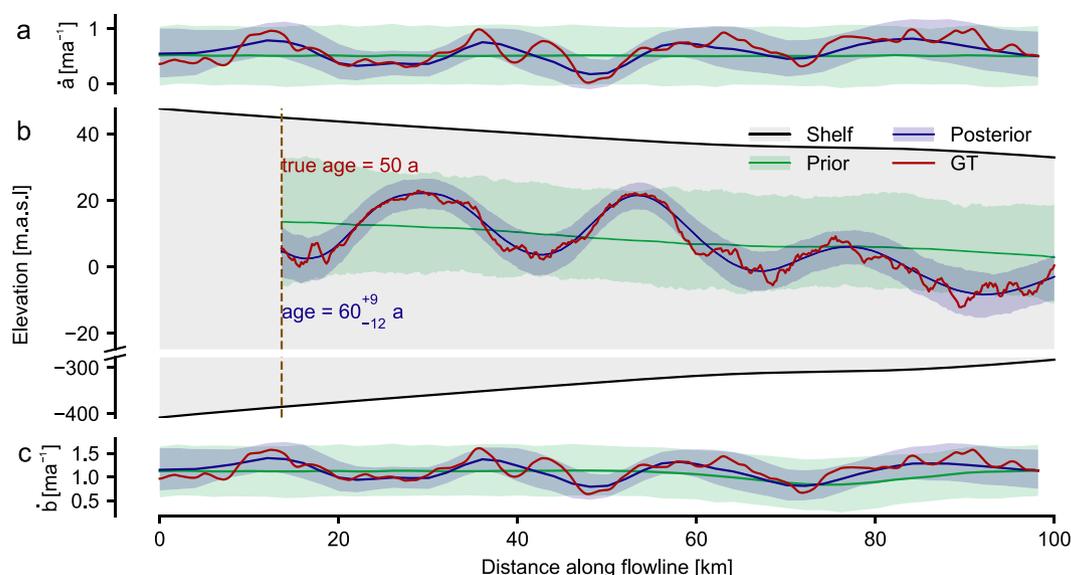


Figure 4. Prior and posterior (predictive) for the synthetic dataset. (a and c) Prior and posterior over surface accumulation and basal melt rates respectively for layer 1 of the synthetic ice shelf, of age 50 years. Solid line is the distribution mean, the shaded region represents the 5th and 95th percentiles. The ground truth (GT) parameters used to generate the reference isochronal layer are shown in red. (b) Cross section of the ice shelf. Prior and posterior predictive distributions for the layer closest matching the ground truth isochronal layer. The vertical dashed line represents the LMI boundary for this isochronal layer. The posterior predictive reconstructs the observed layer with higher accuracy and lower uncertainty. The posterior predictive distribution of the age of the isochronal layer is 60^{+9}_{-12} years (meaning a median of 60 years, and 16th and 84th percentiles of 48 and 69 years, respectively). The average root-mean-square error (RMSE) relative to the GT isochronal layer is 3.9 m for the posterior predictive distribution and 11.5 m for the prior predictive distribution.

matches the ground truth isochronal layer with high fidelity. We calculate the RMSE of the predictive simulations relative to the ground truth layer elevations for 1000 simulations using prior and posterior samples. The average RMSE for the posterior predictive distribution is 3.9 m, compared to 11.5 m for the prior predictive distribution. Uncertainties in the layer elevations are much smaller than those of the prior predictive distribution. This is in contrast to the posterior uncertainty over the mass-balance rates, which is still considerable. This showcases the importance of our uncertainty-aware approach: there is more than one parameterization of accumulation and basal melt rates that can lead to similar isochronal layers.

The posterior uncertainty is also reflected in the inferred age of the isochronal layer. We infer an age of 60^{+9}_{-12} years for this layer (meaning a median of 60 years, and 16th and 84th percentiles of 48 and 69 years, respectively). This value closely matches the age of the ground truth isochronal layer, which was not used during inference. Thus, we have produced an estimate of the age of the layer without requiring time intensive measurements such as ice cores. We report the posterior distributions for deeper synthetic layers in Appendix F.

Finally, while we do not use the isochronal layer elevations outside the LMI boundary for inference, we can still infer the surface accumulation and basal melt rates at these locations. This is because the values of surface accumulation rate and basal melt rate still affect the downstream isochronal layer elevations, and so the observed elevations in the LMI body still contain information about the mass-balance rates upstream of the LMI boundary. Thus, we are still able to infer the mass-balance rates for $x < 15$ km.

4. Ekström Ice Shelf

EIS is a medium-sized ice shelf located between the Sörasen and Halvfarryggen Ice Rises in Dronning Maud Land, East Antarctica

(Fig. 5c). EIS makes for an appropriate study site since the steady-state assumption likely holds (Drews and others, 2013; Schannwell and others, 2019). Moreover, because of the proximity of the Neumayer station III, numerous observations are available, e.g. ice thickness, surface velocities and most importantly surface accumulation rates, which we will use later for validation.

4.1. Data preprocessing

First, we used Antarctic Mapping Tools (Greene and others, 2017), BedMachine Antarctica (Morlighem and others, 2017) and ITS_LIVE (Gardner and others, 2018; 2022) to obtain the surface elevation s , thickness h and velocity \mathbf{v} for EIS. In order to define the flow tube domain for EIS, we also used the `itslive_flowline` tool to find two flow lines which formed the side-boundaries of the domain. The other two boundaries of the domain were the grounding line, and a straight line connecting the two flow lines. The straight line was chosen to ensure that the radar transect where data were measured is wholly contained within the flow tube domain.

We preprocessed the raw ice shelf geometry and velocity data prior to evaluating the model. This ensured numerical stability of the forward model. Using the `icepack` package for Python (Shapiro and others, 2021), we first smoothed the raw thickness data by solving a regularized minimization problem. We then solved for the best-fitting velocity by fitting a fluidity parameter in an SSA model to the observed velocity and smoothed thickness. The hyperparameters used for preprocessing are given in Appendix D.

4.2. Radar measurements of internal stratigraphy

Internal stratigraphy data along the central flow line of EIS (Fig. 5a) were acquired using a ground-based ground-penetrating radar

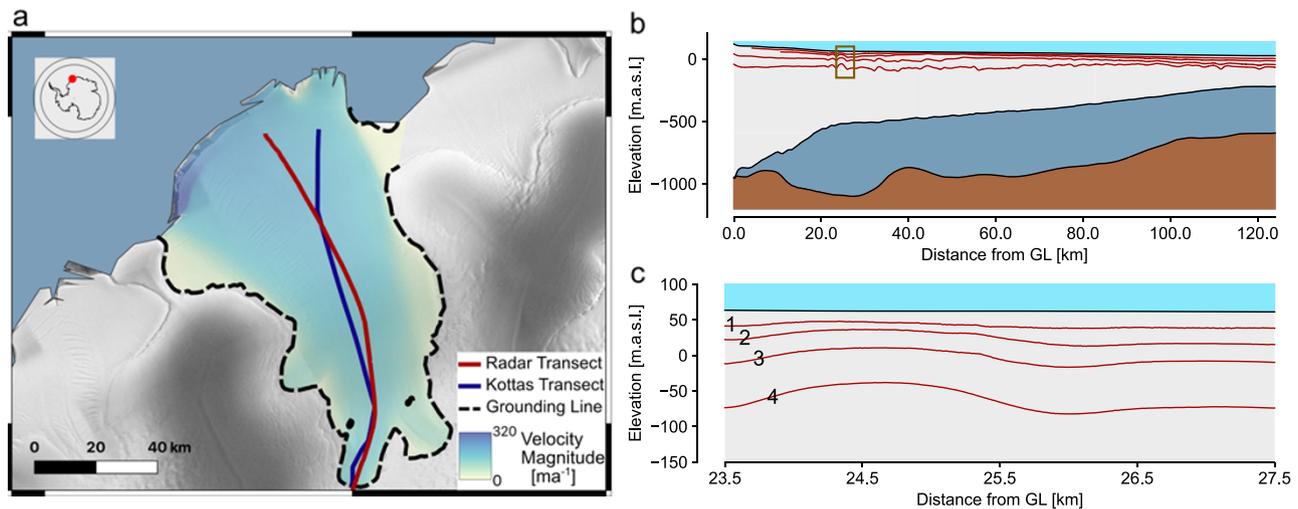


Figure 5. Overview of the Ekström Ice Shelf. (a) Satellite view of Ekström Ice Shelf along with location of the radar transect along the central flow line (red line) and the Kottas traverse (blue line). An independent estimate of surface accumulation via stake arrays is available on Kottas traverse, which we use to validate our results. In our model, we use the velocity data from ITS_LIVE (Gardner and others, 2018; 2022). (b) Vertical cross-section view of the radar transect, along with ice surface and base take from BedMachine Antarctica (Morlighem and others, 2017), starting at the grounding line (GL). Red lines indicate four picked internal reflection horizons (IRHs). (c) Zoom in on box in B. The IRHs are numbered 1–4 in order of increasing depth. This plot is shown with the radar data used to label the IRHs in Figure 11.

with a center frequency of 50 MHz (pulseEKKOTM from Sensors & Software) in two consecutive field seasons (2021/22 and 2022/23) with logistic support from the Neumayer III station (Wesche and others, 2016; Wesche and Regnery, 2022). Radar processing was done with ImpDAR (Lilien and others, 2020) and included trace averaging to equidistant spacing (10 m), bandpass filtering (with cutoff frequencies of 20 and 75 MHz), and a topographic correction using the REMA surface elevation (Howat and others, 2019). The latter provides observations consistent with the modeling setup. The radar detects the ice–ocean interface and continuous IRHs down to ~ 200 m depth (Fig. 5b and c). Four IRHs were digitized along the entire 130 km long profile using a semi-automatic maximum tracking scheme. The vertical offset of IRHs at the profile junction in the mid-shelf region between both years is much smaller than the radar system’s wavelength in ice (~ 3.4 m). Consequently, IRHs were connected without adjustments. For the travel time-to-depth conversion, we used a depth-density profile representative for ice shelves of the Dronning Maud Land Coast (Hubbard and others (2013), eqn (1)).

4.3. Inference results

We inspect the prior over the basal melt rates as a validation of our modeling choices. The implicit prior is the same as the prior defined for the surface accumulation, with the mean shifted by the total mass balance on the flow line, \bar{m} . The basal melt rate is larger (up to 4 m a^{-1}) near the grounding line and gradually stabilizes in the along-flow direction to values between 0 and 1 m a^{-1} downstream. This is in agreement with previous estimates for basal melt profiles on this particular ice shelf (Neckel and others, 2012).

We infer the surface accumulation and basal melt rates from IRH 2 in our dataset, which has an average (ice equivalent) depth of 30 m (Fig. 6). The posterior over the surface accumulation rate has uncertainty comparable to that of the prior. However, there is a shift in the overall spatial trend of the accumulation rate; particularly, there is higher surface accumulation rate at ~ 20 km from the grounding line. Accumulation rate also increases steadily

downstream the flow line. As in the synthetic case, the posterior predictive distribution reproduces the observed IRH with much higher fidelity and confidence than the prior predictive distribution. The average RMSE relative to the observed IRH is 4.6 m for 1000 posterior predictive simulations, compared to 11.8 m for 1000 prior predictive simulations. The posterior predictive produces an independent estimate of the unknown age of the IRH of 84_{-30}^{+52} years (meaning a median of 84 years, and 16th and 84th percentiles of 54 and 136 years, respectively).

Our method can use much deeper IRHs for the inference of accumulation and basal melt rates. For IRH 4 of the observed dataset (of average depth 131 m), the proportion of the IRH that is within the LMI body is smaller. This is due to the unknown boundary condition influencing the IRH elevation at much further points along the flow line. This discarding of data has visible effects on the posteriors over the mass-balance parameters (Fig. 7). These rates are now more similar to the priors for the first 60 km of the transect and only diverge at points further down the ice shelf, where the values of accumulation and basal melt rates affect the dynamics of the IRH. Regardless, the posterior predictive distribution resembles the observed IRH at higher fidelity and precision than the prior predictive. The average RMSE relative to the observed IRH is 10.0 m for the posterior predictive distribution and 16.4 m for the prior predictive distribution. The estimated age of this IRH by our method is 188_{-49}^{+96} years. The uncertainty of the age estimates reasonably increases for deeper IRHs.

5. Discussion

5.1. Posterior mass-balance rates are consistent between IRHs

We compare the four posteriors over the surface accumulation obtained from the Ekström IRH dataset (Fig. 8). The posteriors for the shallower IRHs 1–3 all show a similar qualitative relationship: a local maximum of the accumulation at a distance of ~ 20 km from the grounding line, followed by a steady increase in the accumulation downstream. The increase in accumulation at ~ 20 km is even identified in the posterior for IRH 3, despite

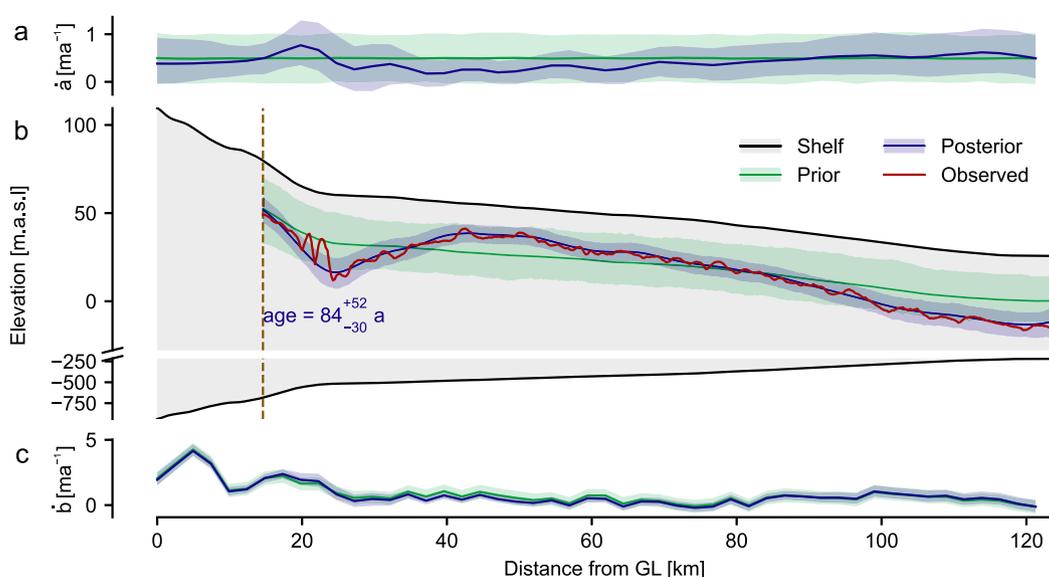


Figure 6. Prior and posterior (predictive) for the Ekström dataset, IRH 2, of average depth 30 m. (a and c) Prior and posterior over surface accumulation and basal melt rates respectively, starting at the grounding line (GL). Solid line is the distribution mean, the shaded region represents the 5th and 95th percentiles. (b) Cross section of the ice shelf. Prior and posterior predictive distributions for the layer closest matching the observed IRH. The vertical dashed line represents the LMI boundary for this IRH. The posterior predictive reconstructs the observed IRH with higher accuracy and lower uncertainty. The posterior predictive distribution of the age of the IRH is 84^{+52}_{-30} years (meaning a median of 84 years, and 16th and 84th percentiles of 54 and 136 years, respectively). The average RMSE relative to the observed IRH is 4.6 m for the posterior predictive distribution and 11.8 m for the prior predictive distribution.

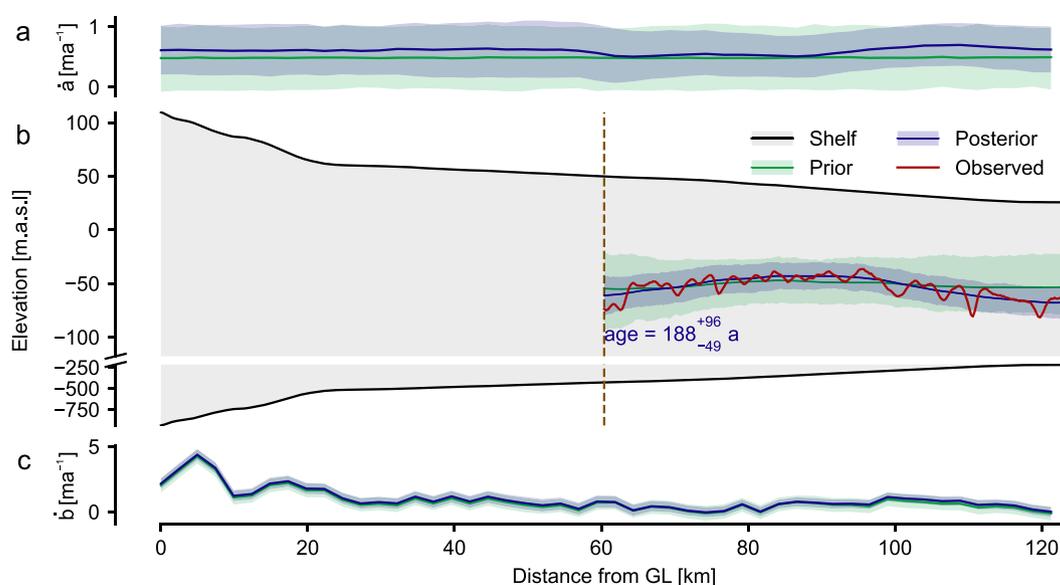


Figure 7. Prior and posterior (predictive) for the Ekström dataset, IRH 4, of average depth 113 m. Same as Figure 6 for the deeper IRH. The posterior predictive distribution of the age of the IRH is 188^{+96}_{-49} years. The average root-mean-square error relative to the observed IRH is 10.0 m for the posterior predictive distribution and 16.4 m for the prior predictive distribution.

the LMI boundary being downstream of it, at ~ 30 km from the grounding line. This is reasonable, as the mass-balance parameters at a given location affect the flow field downstream of this location, and consequently, the formation of isochronal layers. For IRH 4, the LMI boundary is much further downstream at ~ 60 km. Thus, the local surface accumulation maximum at ~ 20 km is not found; however, the overall trend of increasing surface accumulation downstream is still identified. There is a corresponding trend in the basal melt rate, as the local basal melt rate still exhibits a maximum at ~ 20 km. The reason for this is unknown, but the location

corresponds both with the seaward limit of the tidal flexure zone and with the confluence region of ice originating from the eastern tributary. One or both of these factors could alter the basal melt rates inferred at this location. As we will show later (Section 5.3), this local maximum also appears in independent remote-sensing estimates.

The inferred posteriors also allow us to estimate the age of the IRHs. By sampling from the posterior distribution, and evaluating the forward model with the resulting mass-balance parameter samples, we obtain a distribution of isochronal layers similar to

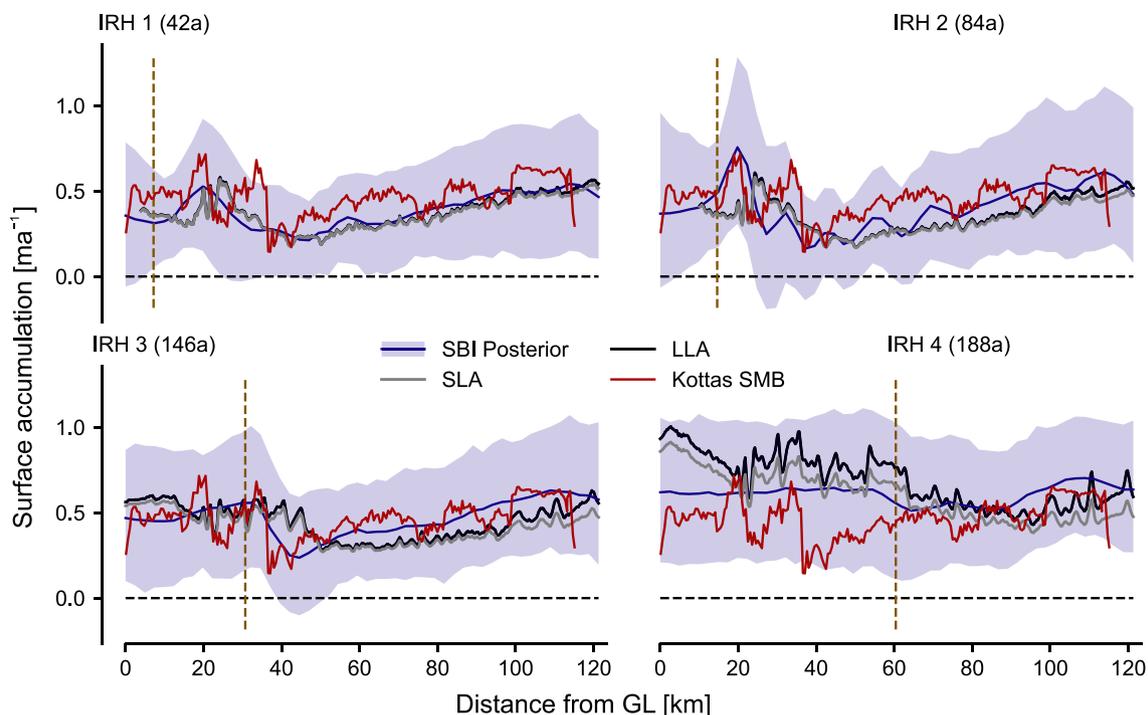


Figure 8. Ekström Ice Shelf—dependence of posterior surface accumulation rate on depth of IRH used for inference. The posteriors are compared to the shallow layer approximation (SLA) and local layer approximation (LLA) (Waddington and others, 2007), and an estimate of the distribution of the accumulation rate based on measurements along the Kottas traverse. See Figure G1 for yearly Kottas measurements. As the real age of the IRHs is not known, the SBI-derived median age is used for the SLA and LLA approximations. Median ages for IRH 1–4 are 42_{-12}^{+32} , 84_{-30}^{+52} , 146_{-38}^{+52} and 188_{-49}^{+96} years. The LMI boundary, representing where the IRH data were masked, is shown with the brown dashed lines.

the observed IRH, with known ages. Thus, we estimate the ages of the four IRHs as 42_{-12}^{+32} , 84_{-30}^{+52} , 146_{-38}^{+52} and 188_{-49}^{+96} years. This is an important finding because IRH age is otherwise only accessible through ice coring. Our results, however, depend on a realistic prior for the surface accumulation and basal melt rates, as defined in Section 2.2.3. Given a miscalibrated prior, the estimated ages would not be reliable (see Fig. H1 for an example). We hypothesize that, given an independent measurement of the IRH age, our approach could further constrain the posterior distributions over the mass-balance parameters.

The consistent spatial patterns of magnitudes of accumulation rates inferred from IRHs 1–4 are supportive of EIS being in steady state over the last hundreds of years but given that steady-state is one of our model assumptions this interpretation needs to be considered with care.

5.2. Comparison to shallow and local layer approximations

To validate our approach, we compare the inferred surface accumulation rate of our experiments with estimates from other methods. First, we computed the shallow layer approximation (SLA) and local layer approximation (LLA) as described in Waddington and others (2007). Given the depth and age of IRH m , the SLA and LLA approximations for the accumulation rate $\dot{\mathbf{a}}$ are defined as

$$\begin{aligned} \dot{\mathbf{a}}_{\text{SLA}}^m &= \frac{1}{\mathcal{A}_m} (\mathbf{s} - \mathbf{e}_m(\mathbf{x})), \\ \dot{\mathbf{a}}_{\text{LLA}}^m &= -\ln \left(1 - \frac{\mathbf{s} - \mathbf{e}_m(\mathbf{x})}{\mathbf{h}} \right) \frac{\mathbf{h}}{\mathcal{A}_m}, \end{aligned} \quad (13)$$

where \mathcal{A}_m is the age of IRH m . Intuitively, the SLA takes the ice thickness above layer m and divides it by the layer age, whereas

LLA accounts for strain thinning assuming a linear vertical velocity profile (which is often the case for ice-shelf flow). Since the age of the observed IRHs is not known, we use the median age of the posterior predictive distribution results. As expected, we observe that both SLA and LLA closely match the SBI posterior mean accumulation rate for the shallow IRHs of median estimated ages 42 and 84 years (Fig. 8). As the strain rates of the flow are small, the relatively shallow IRHs (mean ice equivalent depth of 30 m) have not notably deformed, and hence the assumptions of SLA and LLA are appropriate. However, for the deeper IRHs 3 and 4 of estimated ages 146 and 188 years, we see that both SLA and LLA estimates diverge from our posterior mean accumulation rate. This shows that more involved approaches are required when using deeper IRHs for inference. For deeper IRHs where the SLA and LLA no longer applied, Steen-Larsen and others (2010) inferred the surface accumulation rates on grounded ice using a Monte Carlo approach. By treating the age of the IRH as an additional parameter to infer, they were able to identify the age of the IRH with high confidence. Extensions of our approach could incorporate this parameterization to reduce the uncertainty of the inferred IRH.

5.3. Comparison with independent estimates of surface accumulation and basal melting

For the Ekström transect comparison data are provided by repeat readings of accumulation stakes in 500 m spacing along the nearby Kottas traverse (Fig. 8). Yearly readings are available in the period 1996–2005 and on a yearly to three-yearly interval between 2014 and 2023 (Mengert, 2018). We use this dataset to construct a direct estimate of time-averaged surface accumulation rate along the central flow line transect over these periods. For this, we project

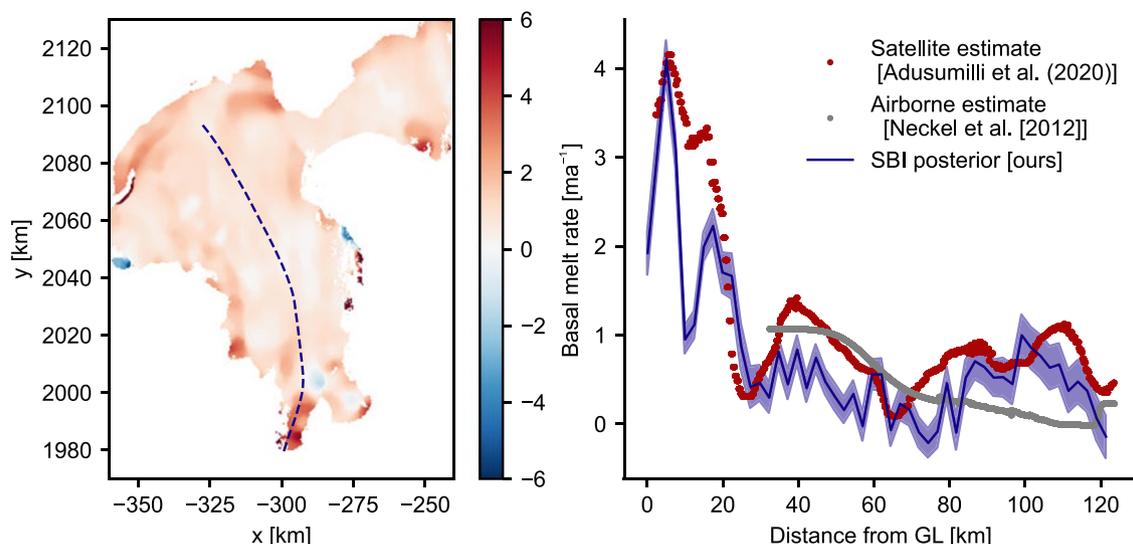


Figure 9. Basal melting rates comparison. (a) Map of basal melt rates for Ekström Ice Shelf, using data from Adusumilli and others (2020). (b) Comparison of inferred basal melt rates from IRH 1 to independent estimates of basal melt rates, calculated on the flow line transect.

the measurements from the Kottas traverse to the flow line transect, taking into account an increased uncertainty for increasing projection distance (see Appendix G for details).

The Kottas traverse accumulation measurements closely match the posterior means of our approach (Fig. 8) for IRHs 1 and 2. As the accumulation rate measurements on the Kottas traverse span the past 26 years, it may not be a good validation for the deeper IRHs. Regardless, the Kottas accumulation rate measurements lie within the posterior uncertainty for IRHs 3 and 4. These comparisons further corroborate our approach and highlight the advantages of uncertainty-aware methods, especially as the measured accumulation rates also varied considerably year-to-year (Appendix G).

We also compare our inferred basal melt rates with independent measurements of basal melt rates. In Fig. 9, we show the basal melt rates inferred from IRH 1 for the Ekström dataset, in comparison to independent estimates of basal melt rates through satellite altimetry data (Adusumilli and others, 2020), and through airborne radar measurements of the ice-shelf thickness (Neckel and others, 2012). We observe a quantitative match between our posterior basal melting rate and the estimates from Adusumilli and others (2020). The ice-shelf wide melt rate estimates show overall comparatively little spatial variability. The most notable difference is that our basal melt rates show more variability in the first 20 km of the profile, and this could be due to the proximity of the grounding zone where the SSA approximation does not hold. However, we note that also the satellite-derived estimates show this oscillation in basal melt rates albeit with a smaller magnitude. The estimates from Neckel and others (2012) excluded the grounding zone area but otherwise show a good match in magnitude but with much less spatial variability. This is because they decided to apply spatial smoothing the degree of which could be revisited given the new results derived here.

The good quantitative match with independently collected data both for surface accumulation and basal melting increases our confidence for our inferred surface accumulation and basal melt rates. However, these results are limited by some of our modeling assumptions, which we discuss next.

5.4. Limitations of modeling approach

The fidelity at which the posterior predictive distributions reproduce the observed IRHs of EIS (Figs. 6 and 7) supports our modeling choices for this ice shelf, as the combination of the forward model and accumulation rate prior distribution are sufficiently expressive to reproduce the IRHs.

However, our inferred surface accumulation and basal melt rates rely on the modeling assumption that the internal radar data are collected on a flow-line transect. This assumption is required to conclude that the ice observed in the internal stratigraphy is indeed the accumulated ice modeled in our domain, as commonly assumed in the literature (Waddington and others, 2007; Steen-Larsen and others, 2010; Theofilopoulos and Born, 2023). Similarly, in the context of ice rises, it is often assumed that the transverse velocity is negligible relative to the vertical velocity of the ice, so that the surface accumulation rate can be inferred along the same transect as the IRH data (Callens and others, 2016; Koch and others, 2024). However, many of the available IRH measurements do not align with flow lines of the ice sheet. In this case, our assumption would not be valid.

Because our observations are on a flow line transect only, it is difficult to judge to what extent unidentified three-dimensional (3-D) effects overprint our analysis. Previous approaches (Pattyn and others, 2012) have had similar limitations because ice-flow divergence and/or convergence could not be predicted by their 2-D forward model. They concluded that their inferred basal melt rates which best matched the radar stratigraphy would be a lower boundary because ice flow on the ice shelf was convergent. In our case, we do correct for the observed convergence from observed surface velocities along the flow tube. The normal component of ice-flow is always $< 20 \text{ m a}^{-1}$ and often $< 1 \text{ m a}^{-1}$. This is small compared to the along-flow velocities and also compared to the total surface mass balance accumulated along the flow tube. Together with the empirical validation with independently collected surface accumulation and basal melt rates, this increases our confidence that our modeling approach yields trustworthy results. Yet, a more rigorous quantification of 3-D effects, for example, in a

synthetic study using a 3-D forward model can provide further validation. Others have made progress in this direction, for example, Wolovick and others (2021) consider a 3-D steady-state ice sheet and jointly infer the temporally averaged accumulation rate and geothermal heat flow. This is done by using more radar attributes in addition to stratigraphy such as existence or absence of subglacial water and/or basal freeze-on.

Common to most previous studies is the steady-state assumption which is imposed because the inclusion of transient ice thickness changes increases the model parameter space to a degree which cannot easily be solved in the inverse problem, particularly with quantified uncertainties over the inferred parameters. Ways forward in this regard could be deterministic gradient descent schemes with explicitly calculating sensitivity matrices as suggested by Theofilopoulos (2022); Theofilopoulos and Born (2023) and this will be an important step forward to better exploit the growing IRH archive for ice-sheet modeling.

5.5. SBI as a tool for geoscientific inversion problems

The inverse problem tackled in this work typifies geoscientific inverse problems, as the forward model is defined in terms of a partial differential equation, and the parameters are high dimensional and vary in space. Hence, it is valuable to compare the SBI approach in this case to the wide variety of methods and algorithms that have been developed to solve geoscientific inverse problems. In the remainder of the section, we discuss NPE as used in our work. However, there exist other variants of SBI with relative advantages and disadvantages, depending on the problem setting (Cranmer and others, 2020; Lueckmann and others, 2021). In particular, we provide a brief discussion of the neural likelihood estimation (NLE) variant in Appendix C.4.

The SBI approach as presented here has two key features. First, we estimate the Bayesian posterior distribution, providing quantitative uncertainty estimates. Modeling uncertainty is important as it can highly influence and propagate to future modeling predictions. Additionally, locations of high uncertainty show areas requiring further study, helping to guide future work. This is in contrast to deterministic inversion methods, which do not estimate uncertainty, or likelihood-based inference methods which are not possible when the likelihood defined by the simulator is not known. Thus, approximate Bayesian methods and SBI in particular can be applied to a larger class of inference problems. Second, a unique advantage of *single round* SBI methods (Cranmer and others, 2020) such as NPE (as used in our study) is *amortization*. Our method as presented here is not yet fully amortized, instead amortizing the vast majority of the computational cost, as preprocessing relies on the observed value of X . In order to train the density estimator $q_{\phi}(\hat{\mathbf{a}}|X_k^m)$, we first calculate X_k^m for each simulation dependent on the value of X_o^m . Our method still amortizes the cost of simulating the forward model many times, which is by far the largest computational cost in the approach. In the Ekström example, we have evaluated the forward model a total of 190 000 times, accounting for ~99% of the total computation cost (Appendix E). This amortization is specific to the geometry and velocity of the EIS; different geometries and velocities change the dependence of the internal layers on the mass-balance parameters, which would require simulating from a new model.

On the other hand, SBI faces some limitations as an inference tool. Primarily, SBI methods are known to require a large number of simulations to be trained (Lueckmann and others, 2021). This problem suffers from the curse of dimensionality—the number

of simulations required scales exponentially with the number of parameters we are trying to infer (in this work, we limited the number of parameters to 50). This is particularly challenging for geoscientific problems, where typically the parameters of interest vary spatially (and temporally), and thus the number of parameters can grow very large. The SBI approach needs to be adapted to more efficiently represent high-dimensional, spatially varying parameters θ at high resolutions. Some potential approaches are polynomial or spectral representations. Future work should also explore variants of SBI that are better suited to high-dimensional or even continuous parameters (Ramesh and others, 2022; Geffner and others, 2023). Finally, SBI works under the assumption that the forward model is well-specified, meaning that given samples from the prior, it can generate simulations closely resembling the observation. The posteriors obtained by SBI can be strongly biased when this is not the case (Cannon and others, 2022). Work to address this concern has been done, e.g. by incorporating the model mismatch into the forward model (Ward and others, 2022), as done in our work using the calibrated noise model. However, designing and calibrating such noise models for each inference task are challenging, and a standard approach for addressing model mismatch does not yet exist.

6. Conclusions

We presented a novel approach for inferring the spatially varying surface accumulation and basal melt rates along ice-shelf flow lines from radar measurements of their internal stratigraphy. We validated the method on a synthetic ice shelf example and inferred the surface accumulation and basal melt rates along a flow line in EIS, Antarctica. We separately inferred the mass-balance parameters from four different IRHs. The inferred distributions were further validated by independent stake array measurements of surface accumulation rates uniquely available in Ekström Ice Shelf. Using our approach, we were able to estimate the otherwise unknown age of the IRHs as 42^{+32}_{-12} , 84^{+52}_{-30} , 146^{+52}_{-38} and 188^{+96}_{-49} years. The presented approach can be transferred to other Antarctic ice shelves and also to other flow regimes such as grounded ice. A strength of our approach is the principled uncertainty estimates in the inferred surface accumulation and basal melt rates. These uncertainty estimates can be integrated in future projections of the Antarctic ice sheet (Verjans and others, 2022; Ultee and others, 2024). We identified avenues for future work as more can be learned by relaxing the steady-state assumption on the ice shelf. The forward model and inference framework should be adapted to account for potential transient signals in the mass-balance parameters.

This work was an example use case of SBI for a geoscientific inverse problem. We showcased the strengths of SBI as a likelihood-free approach to approximate the Bayesian posterior, amortizing the cost of simulating the forward model many times. SBI can become more applicable to such inverse problems involving spatially (and temporally) varying parameters if it can be extended to deal with the challenge of high-dimensional parameter inference.

Finally, our approach highlights the value of internal stratigraphy measurements. Initiatives to map the Antarctic-wide internal stratigraphy (e.g. Bingham and others, 2024) can provide invaluable data toward uncovering the history of the Antarctic ice sheet. Sophisticated inference methods could be combined with such a dataset to provide a new, independent, Antarctica-wide parameterization of accumulation and basal melt rate histories.

Data availability statement. The extracted IRH elevations along the Ekström transect are available in Oraschewski and others (2024a). The processed radar data are additionally available in Oraschewski and others (2024b). Simulation data available in Moss and others (2023).

Software availability. Code for preprocessing Ekström Ice Shelf data and generating synthetic ice shelf data is available in Moss and others (2024a). Code for layer tracing forward model and simulation-based inference workflow is available in Moss and others (2024b).

Acknowledgements. The authors would like to thank Daniel Shapero for his inputs on use of icepack. The authors would also like to thank Andreas Born and Therese Riekch for insightful discussions on the implementation of the layer tracing solver for calculating internal stratigraphy. We acknowledge excellent logistic support from staff at Neumayer Station III and the GrouZe team on-site.

This work was funded by the German Research Foundation (DFG) under Germany's Excellence Strategy—EXC number 2064/1—390727645 and SFB 1233 'Robust Vision' (276693517) and the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A. Reinhard Drews and Vjeran Višnjić were supported by an Emmy Noether grant of the Deutsche Forschungsgemeinschaft (DR 822/3-1). We acknowledge the support by the German Academic Scholarship Foundation to Falk M. Oraschewski. Field observations were supported by the Alfred Wegener Institute through logistic grants AWL_ANT_23 (Drews) and AWL_ANT_8 (Eisen). We acknowledge support from the Open Access Publication Fund of the University of Tübingen. Guy Moss is a member of the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

References

- Adusumilli S, Fricker HA, Medley B, Padman L and Siegfried MR (2020) Interannual variations in meltwater input to the Southern Ocean from Antarctic ice shelves. *Nature Geoscience* 2020 13:9 13, 616–620. doi: [10.1038/s41561-020-0616-z](https://doi.org/10.1038/s41561-020-0616-z)
- Agosta C and 10 others (2019) Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes. *The Cryosphere* 13, 281–296. doi: [10.5194/tc-13-281-2019](https://doi.org/10.5194/tc-13-281-2019)
- Allgeier J and Cirpka OA (2023) Surrogate-model assisted plausibility-check, calibration, and posterior-distribution evaluation of subsurface-flow models. *Water Resources Research* 59, 1–18. doi: [10.1029/2023WR034453](https://doi.org/10.1029/2023WR034453)
- Asay-Davis XS and 13 others (2016) Experimental design for three interrelated marine ice sheet and ocean model intercomparison projects: MISMIP v. 3 (MISMIP+), ISOMIP v. 2 (ISOMIP+) and MISOMIP v. 1 (MISOMIP1). *Geoscientific Model Development* 9, 2471–2497. doi: [10.5194/GMD-9-2471-2016](https://doi.org/10.5194/GMD-9-2471-2016)
- Berger S, Drews R, Helm V, Sun S and Pattyn F (2017) Detecting high spatial variability of ice shelf basal mass balance, Roi Baudouin Ice Shelf, Antarctica. *The Cryosphere* 11, 2675–2690. doi: [10.5194/tc-11-2675-2017](https://doi.org/10.5194/tc-11-2675-2017)
- Bindschadler R and 17 others (2011) Getting around Antarctica: New high-resolution mappings of the grounded and freely-floating boundaries of the Antarctic ice sheet created for the International Polar Year. *Cryosphere* 5, 569–588. doi: [10.5194/TC-5-569-2011](https://doi.org/10.5194/TC-5-569-2011)
- Bingham RG and 53 others (2024) Review Article: Antarctica's internal architecture: Towards a radiostratigraphically-informed age–depth model of the Antarctic ice sheets. *EGU Sphere* 2024, 1–66. doi: [10.5194/egusphere-2024-2593](https://doi.org/10.5194/egusphere-2024-2593)
- Blei DM, Kucukelbir A and McAuliffe JD (2017) Variational inference: A review for statisticians. *Journal of the American Statistical Association* 112, 859–877. doi: [10.1080/01621459.2017.1285773](https://doi.org/10.1080/01621459.2017.1285773)
- Born A (2017) Tracer transport in an isochronal ice-sheet model. *Journal of Glaciology* 63(237), 22–38. doi: [10.1017/JOG.2016.111](https://doi.org/10.1017/JOG.2016.111)
- Born A and Robinson A (2021) Modeling the greenland englacial stratigraphy. *Cryosphere* 15, 4539–4556. doi: [10.5194/TC-15-4539-2021](https://doi.org/10.5194/TC-15-4539-2021)
- Brinkerhoff D, Aschwanden A and Fahnestock M (2021) Constraining subglacial processes from surface velocity observations using surrogate-based Bayesian inference. *Journal of Glaciology* 67(263), 385–403. doi: [10.1017/jog.2020.112](https://doi.org/10.1017/jog.2020.112)
- Burgard C, Jourdain NC, Reese R, Jenkins A and Mathiot P (2022) An assessment of basal melt parameterisations for Antarctic ice shelves. *Cryosphere* 16, 4931–4975. doi: [10.5194/TC-16-4931-2022](https://doi.org/10.5194/TC-16-4931-2022)
- Callens D, Drews R, Witrant E, Philipp M and Pattyn F (2016) Temporally stable surface mass balance asymmetry across an ice rise derived from radar internal reflection horizons through inverse modeling. *Journal of Glaciology* 62(233), 525–534. doi: [10.1017/jog.2016.41](https://doi.org/10.1017/jog.2016.41)
- Cannon B, Ward D and Schmon SM (2022) Investigating the impact of model misspecification in neural simulation-based inference. *ArXiv e-prints, arXiv:2209.01845*. doi: [10.48550/arXiv.2209.01845](https://doi.org/10.48550/arXiv.2209.01845)
- Catania G, Hulbe C and Conway H (2010) Grounding-line basal melt rates determined using radar-derived internal stratigraphy. *Journal of Glaciology* 56(197), 545–554. doi: [10.3189/002214310792447842](https://doi.org/10.3189/002214310792447842)
- Cranmer K, Brehmer J and Louppe G (2020) The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences* 117, 30055–30062. doi: [10.1073/pnas.1912789117](https://doi.org/10.1073/pnas.1912789117)
- Das I and 11 others (2020) Multidecadal basal melt rates and structure of the Ross Ice Shelf, Antarctica, using airborne ice penetrating radar. *Journal of Geophysical Research: Earth Surface* 125, e2019JF005241. doi: [10.1029/2019JF005241](https://doi.org/10.1029/2019JF005241)
- Depoorter MA and 6 others (2013) Calving fluxes and basal melt rates of Antarctic ice shelves. *Nature* 502, 89–92. doi: [10.1038/nature12567](https://doi.org/10.1038/nature12567)
- Drews R (2015) Evolution of ice-shelf channels in Antarctic ice shelves. *The Cryosphere* 9, 1169–1181. doi: [10.5194/TC-9-1169-2015](https://doi.org/10.5194/TC-9-1169-2015)
- Drews R and 6 others (2016) Constraining variable density of ice shelves using wide-angle radar measurements. *The Cryosphere* 10, 811–823. doi: [10.5194/tc-10-811-2016](https://doi.org/10.5194/tc-10-811-2016)
- Drews R, Martín C, Steinhage D and Eisen O (2013) Characterizing the glaciological conditions at Halvfarryggen ice dome, Dronning Maud Land, Antarctica. *Journal of Glaciology* 59(213), 9–20. doi: [10.3189/2013JoG12J134](https://doi.org/10.3189/2013JoG12J134)
- Drews R, Matsuoka K, Martín C, Callens D, Bergeot N and Pattyn F (2015) Evolution of Derwael ice rise in Dronning Maud Land, Antarctica, over the last millennia. *Journal of Geophysical Research: Earth Surface* 120, 564–579. doi: [10.1002/2014JF003246](https://doi.org/10.1002/2014JF003246)
- Durkan C, Bekasov A, Murray I and Papamakarios G (2019) Neural Spline Flows. *Advances in Neural Information Processing Systems*. In: Curran Associates Inc., Volume 32.
- Dutrioux P and 6 others (2014) Basal terraces on melting ice shelves. *Geophysical Research Letters* 41, 5506–5513. doi: [10.1002/2014GL060618](https://doi.org/10.1002/2014GL060618)
- Eisen O and 15 others (2008) Ground-based measurements of spatial and temporal variability of snow accumulation in East Antarctica. *Reviews of Geophysics* 46, 2. doi: [10.1029/2006RG000218](https://doi.org/10.1029/2006RG000218)
- Eisen O (2008) Inference of velocity pattern from isochronous layers in firn, using an inverse method. *Journal of Glaciology* 54(187), 613–630. doi: [10.3189/002214308786570818](https://doi.org/10.3189/002214308786570818)
- Eisen O, Nixdorf U, Wilhelms F and Miller H (2004) Age estimates of isochronous reflection horizons by combining ice core, survey, and synthetic radar data. *Journal of Geophysical Research: Solid Earth* 109, B4. doi: [10.1029/2003JB002858](https://doi.org/10.1029/2003JB002858)
- Gallée H and Schayes G (1994) Development of a three-dimensional meso- γ primitive equation model: Katabatic winds simulation in the area of Terra Nova Bay, Antarctica. *Monthly Weather Review* 122, 671–685.
- Gardner A and 6 others (2018) Increased West Antarctic and unchanged East Antarctic ice discharge over the last 7 years. *The Cryosphere* 12, 521–547. doi: [10.5194/tc-12-521-2018](https://doi.org/10.5194/tc-12-521-2018)
- Gardner A, Fahnestock M and Scambos T (2022) MEaSURES ITS_LIVE Landsat Image-Pair Glacier and Ice Sheet Surface Velocities, version 1. doi: [10.5067/IMR9D3PEI28U](https://doi.org/10.5067/IMR9D3PEI28U)
- Geffner T, Papamakarios G and Mnih A (2023) Compositional score modeling for simulation-based inference. *Proceedings of Machine Learning Research*. PMLR, pp. 11098–11116, Volume 202.
- Gladstone R and 12 others (2021) The framework for ice sheet-ocean coupling (fisoc) v1.1. *Geoscientific Model Development* 14, 889–905. doi: [10.5194/GMD-14-889-2021](https://doi.org/10.5194/GMD-14-889-2021)

- Goelzer H, Huybrechts P, Loutre MF and Fichet T (2016) Last interglacial climate and sea-level evolution from a coupled ice sheet-climate model. *Climate of the Past* **12**, 2195–2213. doi: [10.5194/cp-12-2195-2016](https://doi.org/10.5194/cp-12-2195-2016)
- Goldberg DN, Gourmelen N, Kimura S, Millan R and Snow K (2019) How accurately should we model ice shelf melt rates? *Geophysical Research Letters* **46**, 189–199. doi: [10.1029/2018GL080383](https://doi.org/10.1029/2018GL080383)
- Goldberg DN and Holland PR (2022) The relative impacts of initialization and climate forcing in coupled ice sheet-ocean modeling: Application to Pope, Smith, and Kohler Glaciers. *Journal of Geophysical Research: Earth Surface* **127**, e2021JF006570. doi: [10.1029/2021JF006570](https://doi.org/10.1029/2021JF006570)
- Greenberg D, Nonnenmacher M and Macke J (2019) Automatic posterior transformation for likelihood-free inference. *Proceedings of Machine Learning Research*. PMLR, pp. 2404–2414, Volume 97.
- Greene CA, Gwyther DE and Blankenship DD (2017) Antarctic mapping tools for MATLAB. *Computers & Geosciences* **104**, 151–157. doi: [10.1016/j.cageo.2016.08.003](https://doi.org/10.1016/j.cageo.2016.08.003)
- Greve R and Blatter H (2009) *Dynamics of Ice Sheets and Glaciers*. Berlin Heidelberg: Springer. doi: [10.1007/978-3-642-03415-2](https://doi.org/10.1007/978-3-642-03415-2)
- Gudmundsson GH, Paolo FS, Adusumilli S and Fricker HA (2019) Instantaneous Antarctic ice sheet mass loss driven by thinning ice shelves. *Geophysical Research Letters* **46**, 13903–13909. doi: [10.1029/2019GL085027](https://doi.org/10.1029/2019GL085027)
- Henry ACJ and 6 others (2023) Predicting the three-dimensional age-depth field of an ice rise. *Authorea*. doi: [10.22541/essoar.169230234.44865946/v1](https://doi.org/10.22541/essoar.169230234.44865946/v1)
- Holschuh N, Parizek BR, Alley RB and Anandakrishnan S (2017) Decoding ice sheet behavior using englacial layer slopes. *Geophysical Research Letters* **44**, 5561–5570. doi: [10.1002/2017GL073417](https://doi.org/10.1002/2017GL073417)
- Howat IM, Porter C, Smith BE, Noh MJ and Morin P (2019) The reference elevation model of Antarctica. *The Cryosphere* **13**, 665–674. doi: [10.5194/tc-13-665-2019](https://doi.org/10.5194/tc-13-665-2019)
- Hubbard B and 6 others (2013) Ice shelf density reconstructed from optical televiewer borehole logging. *Geophysical Research Letters* **40**, 5882–5887. doi: [10.1002/2013GL058023](https://doi.org/10.1002/2013GL058023)
- Hull R and 7 others (2022) Using simulation-based inference to determine the parameters of an integrated hydrologic model: A case study from the upper Colorado River basin. *Hydrology and Earth System Sciences Discussions* **2022**, 1–38. doi: [10.5194/hess-2022-345](https://doi.org/10.5194/hess-2022-345)
- Kingma DP and Ba J (2015) Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.
- Kobyzev I, Prince SJ and Brubaker MA (2019) Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**, 3964–3979. doi: [10.1109/TPAMI.2020.2992934](https://doi.org/10.1109/TPAMI.2020.2992934)
- Koch I and 9 others (2024) Radar internal reflection horizons from multi-system data reflect ice dynamic and surface accumulation history along the Princess Ragnhild Coast, Dronning Maud Land, East Antarctica. *Journal of Glaciology* **70**, e18. doi: [10.1017/jog.2023.93](https://doi.org/10.1017/jog.2023.93)
- Lenaerts JTM, Medley B, van den Broeke MR and Wouters B (2019) Observing and modeling ice sheet surface mass balance. *Reviews of Geophysics* **57**, 376–420. doi: [10.1029/2018RG000622](https://doi.org/10.1029/2018RG000622)
- Leysinger Vieli GJMC, Hindmarsh RCA, Siegert MJ and Bo S (2011) Time-dependence of the spatial pattern of accumulation rate in East Antarctica deduced from isochronic radar layers using a 3-D numerical ice flow model. *Journal of Geophysical Research: Earth Surface* **116**, F2. doi: [10.1029/2010JF001785](https://doi.org/10.1029/2010JF001785)
- Lilien DA, Hills BH, Driscoll J, Jacobel R and Christianson K (2020) ImpDAR: An open-source impulse radar processor. *Annals of Glaciology* **61**(81), 114–123. doi: [10.1017/aog.2020.44](https://doi.org/10.1017/aog.2020.44)
- Linde N, Renard P, Mukerji T and Caers J (2015) Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources* **86**, 86–101. doi: [10.1016/j.advwatres.2015.09.019](https://doi.org/10.1016/j.advwatres.2015.09.019)
- Looyenga H (1965) Dielectric constants of heterogeneous mixtures. *Physica* **31**, 401–406. doi: [10.1016/0031-8914\(65\)90045-5](https://doi.org/10.1016/0031-8914(65)90045-5)
- Lueckmann JM, Boelts J, Greenberg D, Goncalves P and Macke J (2021) Benchmarking simulation-based inference. *Proceedings of Machine Learning Research*. PMLR, 130, 343–351.
- Lueckmann JM, Goncalves PJ, Bassetto G, Öcal K, Nonnenmacher M and Macke JH (2017) Flexible statistical inference for mechanistic models of neural dynamics. *Advances in Neural Information Processing Systems*. In Curran Associates Inc., Volume 30.
- MacGregor JA, Matsuoka K, Koutnik MR, Waddington ED, Studinger M and Winebrenner DP (2009) Millennially averaged accumulation rates for the Vostok Subglacial Lake region inferred from deep internal layers. *Annals of Glaciology* **50**(51), 25–34. doi: [10.3189/172756409789097441](https://doi.org/10.3189/172756409789097441)
- Marsh OJ and 6 others (2016) High basal melting forming a channel at the grounding line of Ross Ice Shelf, Antarctica. *Geophysical Research Letters* **43**, 250–255. doi: [10.1002/2015GL066612](https://doi.org/10.1002/2015GL066612)
- Mengert M (2018) Spatial and Temporal Variation of Snow Accumulation Along Kottas Traverse, Dronning Maud land, East Antarctica. doi: [10.1013/epic.29be472c-08be-495a-ba18-8c8e295101a5](https://doi.org/10.1013/epic.29be472c-08be-495a-ba18-8c8e295101a5)
- Morland LW (1984) Thermomechanical balances of ice sheet flows. *Geophysical & Astrophysical Fluid Dynamics* **29**, 237–266. doi: [10.1080/03091928408248191](https://doi.org/10.1080/03091928408248191)
- Morlighem M and 31 others (2017) BedMachine v3: Complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters* **44**, 11,051–11,061. doi: [10.1002/2017gl074954](https://doi.org/10.1002/2017gl074954)
- Moss G and 6 others (2023) Assets for “Simulation-Based Inference of Surface Accumulation and Basal Melt Rates of an Antarctic Ice shelf from Isochronal Layers”. doi: [10.5281/zenodo.10245153](https://doi.org/10.5281/zenodo.10245153)
- Moss G and 6 others (2024a) preprocessing-ice-data. doi: [10.5281/zenodo.11440869](https://doi.org/10.5281/zenodo.11440869)
- Moss G and 6 others (2024b) sbi-ice. doi: [10.5281/zenodo.11440807](https://doi.org/10.5281/zenodo.11440807)
- Neckel N, Drews R, Rack W and Steinhage D (2012) Basal melting at the Ekström Ice Shelf, Antarctica, estimated from mass flux divergence. *Annals of Glaciology* **53**(60), 294–302. doi: [10.3189/2012AoG60A167](https://doi.org/10.3189/2012AoG60A167)
- Nicholls KW, Corr HF, Stewart CL, Lok LB, Brennan PV and Vaughan DG (2015) A ground-based radar for measuring vertical strain rates and time-varying basal melt rates in ice sheets and shelves. *Journal of Glaciology* **61**(230), 1079–1087. doi: [10.3189/2015JOG15J073](https://doi.org/10.3189/2015JOG15J073)
- Omagbon J and 8 others (2021) Case studies of predictive uncertainty quantification for geothermal models. *Geothermics* **97**, 102263. doi: [10.1016/j.geothermics.2021.102263](https://doi.org/10.1016/j.geothermics.2021.102263)
- Oraschewski FM, Moss G, Koch I, Ershadi MR, Eisen O and Drews R (2024a) Ground-Penetrating Radar Data (50MHz) and Internal Reflection Horizons Along the Central Flowline of Ekström Ice Shelf, Dronning Maud Land, East Antarctica. doi: [10.1594/PANGAEA.965143](https://doi.org/10.1594/PANGAEA.965143)
- Oraschewski FM, Moss G, Koch I, Ershadi MR, Eisen O and Drews R (2024b) Ground-Penetrating Radar Data (50MHz) and Internal Reflection Horizons Along the Central Flowline of Ekström Ice Shelf, Dronning Maud Land, East Antarctica - NetCDF File. doi: [10.1594/PANGAEA.965141](https://doi.org/10.1594/PANGAEA.965141)
- Overcast I and 17 others (2021) A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. *Molecular Ecology Resources* **21**, 2782–2800. doi: [10.1111/1755-0998.13514](https://doi.org/10.1111/1755-0998.13514)
- Papamakarios G and Murray I (2016) Fast ϵ -free Inference of Simulation Models with Bayesian Conditional Density Estimation. *Advances in Neural Information Processing Systems*. In Curran Associates Inc., Volume 29.
- Papamakarios G, Nalisnick ET, Rezende DJ, Mohamed S and Lakshminarayanan B (2019a) Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research* **22**, 1–64.
- Papamakarios G, Sterratt D and Murray I (2019b) Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* **89**. In PMLR
- Pattyn F and 8 others (2012) Melting and refreezing beneath Roi Baudouin Ice Shelf (East Antarctica) inferred from radar, GPS, and ice core data. *Journal of Geophysical Research: Earth Surface* **117**, F4. doi: [10.1029/2011JF002154](https://doi.org/10.1029/2011JF002154)
- Pattyn F, Favier L, Sun S and Durand G (2017) Progress in numerical modeling of Antarctic ice-sheet dynamics. *Current Climate Change Reports* **3**, 174–184. doi: [10.1007/s40641-017-0069-7](https://doi.org/10.1007/s40641-017-0069-7)
- Pratap B and 7 others (2022) Three-decade spatial patterns in surface mass balance of the Nivlisen Ice Shelf, central Dronning Maud Land, East Antarctica. *Journal of Glaciology* **68**(267), 174–186. doi: [10.1017/JOG.2021.93](https://doi.org/10.1017/JOG.2021.93)

- Ramesh P and 6 others** (2022) GATSBI: Generative adversarial training for simulation-based inference. In *International Conference on Learning Representations*.
- Rasmussen CE and Williams CKI** (2005) *Gaussian Processes for Machine Learning*. The MIT Press.
- Reese R, Gudmundsson GH, Levermann A and Winkelmann R** (2017) The far reach of ice-shelf thinning in Antarctica. *Nature Climate Change* **8**, 53–57. doi: [10.1038/s41558-017-0020-x](https://doi.org/10.1038/s41558-017-0020-x)
- Schannwell C, Drews R, Ehlers TA, Eisen O, Mayer C and Gillet-Chaulet F** (2019) Kinematic response of ice-rise divides to changes in ocean and atmosphere forcing. *The Cryosphere* **13**, 2673–2691. doi: [10.5194/tc-13-2673-2019](https://doi.org/10.5194/tc-13-2673-2019)
- Shapiro DR, Badgeley JA, Hoffman AO and Joughin IR** (2021) icepack: a new glacier flow modeling package in Python, version 1.0. *Geoscientific Model Development* **14**, 4593–4616. doi: [10.5194/gmd-14-4593-2021](https://doi.org/10.5194/gmd-14-4593-2021)
- Steen-Larsen HC, Waddington ED and Koutnik MR** (2010) Formulating an inverse problem to infer the accumulation-rate pattern from deep internal layering in an ice sheet using a Monte Carlo approach. *Journal of Glaciology* **56**(196), 318–332. doi: [10.3189/002214310791968476](https://doi.org/10.3189/002214310791968476)
- Sun S, Hattermann T, Pattyn F, Nicholls KW, Drews R and Berger S** (2019) Topographic shelf waves control seasonal melting near Antarctic ice shelf grounding lines. *Geophysical Research Letters* **46**, 9824–9832. doi: [10.1029/2019GL083881](https://doi.org/10.1029/2019GL083881)
- Sutter J, Fischer H and Eisen O** (2021) Investigating the internal structure of the Antarctic ice sheet: The utility of isochrones for spatiotemporal ice-sheet model calibration. *Cryosphere* **15**, 3839–3860. doi: [10.5194/tc-15-3839-2021](https://doi.org/10.5194/tc-15-3839-2021)
- Symes WW** (2009) The seismic reflection inverse problem. *Inverse Problems* **25**, 123008. doi: [10.1088/0266-5611/25/12/123008](https://doi.org/10.1088/0266-5611/25/12/123008)
- Tarasov L, Dyke AS, Neal RM and Peltier W** (2012) A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. *Earth and Planetary Science Letters* **315–316**, 30–40. doi: [10.1016/j.epsl.2011.09.010](https://doi.org/10.1016/j.epsl.2011.09.010)
- Tebaldi C and Sansó B** (2008) Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *Journal of the Royal Statistical Society Series A: Statistics in Society* **172**, 83–106. doi: [10.1111/j.1467-985X.2008.00545.x](https://doi.org/10.1111/j.1467-985X.2008.00545.x)
- Tejero-Cantero A and 7 others** (2020) SBI: A toolkit for simulation-based inference. *Journal of Open Source Software* **5**, 2505. doi: [10.21105/joss.02505](https://doi.org/10.21105/joss.02505)
- Theofilopoulos A** (2022) *Reconstructing surface mass balance from the englacial stratigraphy of the Greenland Ice Sheet*. Ph.D. thesis, The University of Bergen.
- Theofilopoulos A and Born A** (2023) Sensitivity of isochrones to surface mass balance and dynamics. *Journal of Glaciology* **69**(274), 311–323. doi: [10.1017/jog.2022.62](https://doi.org/10.1017/jog.2022.62)
- Ultee L, Robel AA and Castruccio S** (2024) A stochastic parameterization of ice sheet surface mass balance for the Stochastic Ice-Sheet and Sea-Level System Model (StISSM v1.0). *Geoscientific Model Development* **17**, 1041–1057. doi: [10.5194/gmd-17-1041-2024](https://doi.org/10.5194/gmd-17-1041-2024)
- Vankova I and Nicholls KW** (2022) Ocean variability beneath the Filchner-Ronne ice shelf inferred from basal melt rate time series. *Journal of Geophysical Research: Oceans* **127**, e2022JC018879. doi: [10.1029/2022JC018879](https://doi.org/10.1029/2022JC018879)
- van Wessem JM and 18 others** (2018) Modelling the climate and surface mass balance of polar ice sheets using RACMO2 – Part 2: Antarctica (1979–2016). *The Cryosphere* **12**, 1479–1498. doi: [10.5194/tc-12-1479-2018](https://doi.org/10.5194/tc-12-1479-2018)
- Verjans V, Robel AA, Seroussi H, Ultee L and Thompson AF** (2022) The stochastic ice-sheet and sea-level system model v1.0 (STISSM v1.0). *Geoscientific Model Development* **15**, 8269–8293. doi: [10.5194/gmd-15-8269-2022](https://doi.org/10.5194/gmd-15-8269-2022)
- Virtanen P and 34 others** (2020) SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272. doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Visnjec V and 6 others** (2022) Predicting the steady-state isochronal stratigraphy of ice shelves using observations and modeling. *The Cryosphere* **16**, 4763–4777. doi: [10.5194/tc-16-4763-2022](https://doi.org/10.5194/tc-16-4763-2022)
- Waddington ED, Neumann TA, Koutnik MR, Marshall HP and Morse DL** (2007) Inference of accumulation-rate patterns from deep layers in glaciers and ice sheets. *Journal of Glaciology* **53**(183), 694–712. doi: [10.3189/002214307784409351](https://doi.org/10.3189/002214307784409351)
- Ward D, Cannon P, Beaumont M, Fasiolo M and Schmon S** (2022) Robust Neural Posterior Estimation and Statistical Model Criticism. *Advances in Neural Information Processing Systems*. In Curran Associates Inc., pp. 33845–33859, Volume 35.
- Wesche C and 6 others** (2016) Neumayer III and Kohlen station in Antarctica operated by the Alfred Wegener Institute. *Journal of Large-Scale Research Facilities JLSRF* **2**, A85–A85. doi: [10.17815/jlsrf-2-152](https://doi.org/10.17815/jlsrf-2-152)
- Wesche C and Regnery J** (2022) Expeditions to Antarctica: Ant-Land 2021/22 Neumayer Station III, Kohlen Station, Flight Operations and Field Campaigns. *Berichte zur Polar- und Meeresforschung = Reports on polar and marine research*. doi: [10.57738/bzpm_0767_2022](https://doi.org/10.57738/bzpm_0767_2022)
- Winkelmann R, Levermann A, Martin MA and Frieler K** (2012) Increased future ice discharge from Antarctica owing to higher snowfall. *Nature* **492**, 239–242. doi: [10.1038/nature11616](https://doi.org/10.1038/nature11616)
- Wolovick MJ, Moore JC and Zhao L** (2021) Joint inversion for surface accumulation rate and geothermal heat flow from ice-penetrating radar observations at Dome A, East Antarctica. Part I: Model description, data constraints, and inversion results. *Journal of Geophysical Research: Earth Surface* **126**, e2020JF005937. doi: [10.1029/2020JF005937](https://doi.org/10.1029/2020JF005937)
- Wood SN** (2010) Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**, 1102–1104. doi: [10.1038/nature09319](https://doi.org/10.1038/nature09319)
- Zeising O, Steinhage D, Nicholls KW, Corr HFJ, Stewart CL and Humbert A** (2022) Basal melt of the southern Filchner Ice Shelf, Antarctica. *The Cryosphere* **16**, 1469–1482. doi: [10.5194/tc-16-1469-2022](https://doi.org/10.5194/tc-16-1469-2022)

Appendix A. Forward model details

The layer tracing scheme described in Section 2.1 is equivalent to solving a set of advection equations for a set of layers. Here, we explicitly write down the advection equations solved and the boundary conditions defined. We correct for 2-D effects in the advection equations by adding a term for the normal flow into the flow line. We then account for the inflow boundary condition at the grounding line $x = 0$.

A 1-D advection equation for a layer on a flow line reads

$$\frac{\partial h_l}{\partial t} = v_x \frac{\partial h_l}{\partial x}. \quad (\text{A1})$$

In practice, real flow lines have some incoming or outgoing (normal) flux, q_y , where y denotes the horizontal direction perpendicular to x . We account for this normal ice flux and instead solve

$$\frac{\partial h_l}{\partial t} = v_x \frac{\partial h_l}{\partial x} + r_l(x) \frac{\partial q_y(x)}{\partial y}, \quad (\text{A2})$$

where $r_l(x) = h_l(x)/h(x)$ is the ratio of thickness of layer l to the total thickness of the shelf. This equation holds due to the plug flow assumption, in which the flux divergence $\partial q_y/\partial y$ is independent of the depth z . The quantity $\partial q_y/\partial y$ is constant for all layers and independent of the layer thickness. This normal flux component accounts for lateral compression or extension of the flowtube centered on the flow line, and in our case we estimate it from satellite inferred velocities. For EIS, the normal flow component is small compared to the total mass balance along the flow lines, but for other cases this correction might be much more significant.

In order to define the inflow boundary condition, we would need to know the relative thickness of the incoming layers (or alternatively, the vertical age distribution at x_0). This is typically not available in radar measurements of the stratigraphy. It is, therefore, important to use only the IRH elevation information within the LMI body of the flow line. This is the region of our domain which is independent of the boundary condition we chose at x_0 (Fig. A1). This region can be found by tracking the trajectory traced by a particle initially at $(x = x_0, z = s(x_0))$. The LMI body is the region of the domain above this path. As a consequence of this consideration, we need to discard more of the IRH elevation data for the deeper IRHs in the dataset.

For a complete definition of the simulator, we still need to define some boundary condition at x_0 . Since the true boundary condition is not known, we

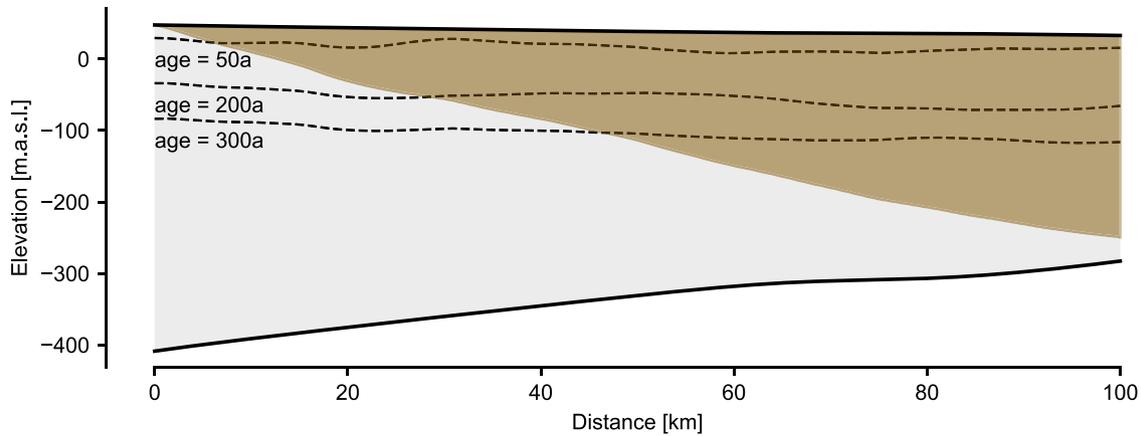


Figure A1. Local meteoric ice (LMI) body. The layer elevations in the LMI body (shaded region) are independent of the inflow boundary conditions. Outside the LMI body, the layer elevations are dependent on this boundary condition, and so using IRH observations within this region would require assuming the internal stratigraphy of the incoming ice.

choose an inflow boundary condition which improves the numerical stability of the layer tracing model,

$$\frac{\partial h_i}{\partial t} = \frac{\partial q_x}{\partial x} \Big|_{x=0} + r_i(x) \frac{\partial q_y(x)}{\partial y}. \quad (\text{A3})$$

This boundary condition makes the implicit assumption that near the inflow boundary, the layer thickness profile is a scalar multiple of the total thickness,

$$h(x) \frac{\partial h_i(x)}{\partial x} = h_i(x) \frac{\partial h(x)}{\partial x}. \quad (\text{A4})$$

We define a similar boundary condition at $x = L_x$; however, this boundary condition does not have a similar effect on the applicability of the IRH data.

Appendix B. Calibrating the simulator

We calibrate hyperparameters of the noise model using a set of 1000 simulations for the same ice shelf geometry, velocity and mass-balance parameters prior. This set of calibration simulations is not used again to train NPE (Section 2.2.1) to avoid overfitting. We use the same set of calibration simulations to calibrate the per-layer LMI boundary mask (Section 2.2.2). Algorithm 1 defines our calibration procedure for the parameters of the noise distribution for EIS.

In Section 2.2.3, we used the LMI boundary $i(m)$ for each IRH m in order to define the domain $\mathbf{x}_{i \geq i(m)}$ on which we compare the observed IRH data to the simulated isochronal layers. The physical interpretation of the LMI boundary is found in Appendix A, and here we specify the exact definition of $i(m)$.

For each simulation k in the calibration set, we define the trajectory of a particle starting at the surface of the inflow boundary ($x = 0, z = s(0)$) by $p^k(t) = (x^k(t), z^k(t))$. In practice, since we are on a flow line, $v_x > 0$ everywhere, and thus x^k increases monotonically with t . Therefore, the trajectory traces out a unique curve $z^k(x)$ in the domain.

Using this path, we define the LMI boundary $i(m, k)$ for simulation k and IRH m to be the first point in the x domain such that the path is below the IRH elevation,

$$i(m, k) = \min_i \{ i : z^k(x_i) < e_m(x_i) \}. \quad (\text{B1})$$

In the case that the path stays above the IRH for the entire domain, we define $i(m, k) = L_x$, meaning that the IRH is entirely outside the LMI body.

This definition gives a different LMI body for each simulation. In order to perform inference, we require one a fixed LMI body boundary to use across all simulations. We define $i(m)$ ‘pessimistically’ as the 75th percentile of the calculated $i(m, k)$ boundaries in the calibration set. Therefore, some of the simulated layer elevations we use will be dependent on the unknown boundary conditions but a small amount that should not affect the inferred results.

Appendix C. Additional information on simulation-based inference

C.1. Tractable and intractable likelihood functions

Here, we provide intuition about the difference of forward models which have intractable and tractable likelihood functions. Forward models which have tractable likelihood functions are ones where the density $p(X|\theta)$ can be explicitly evaluated for a given (X, θ) pair, where X refers to the observed data and θ are the parameters of the model. As an example, a common setting in glaciology is a deterministic forward model with additive observational noise. Such forward models are of the form

$$X = f(\theta) + \epsilon, \quad (\text{C1})$$

where $f(\theta)$ is a deterministic function, and ϵ follows some distribution, $\epsilon \sim p_{\epsilon}(\epsilon)$. In such models, the likelihood $p(X|\theta)$ can be evaluated as

$$p(X|\theta) = p_{\epsilon}(X - f(\theta)). \quad (\text{C2})$$

A more general setting considers nondeterministic forward models, where latent variables z are sampled, and then affect further computations in the forward model. As a motivational example, consider a stochastic differential equation solved with an Euler–Maruyama scheme, that is, an iterative update

$$X_{t+1} = X_t + f(X_t, \theta, t)dt + g(X_t, \theta, t)z_t\sqrt{dt}, \quad (\text{C3})$$

where f is now a deterministic function dependent on the state X_t and parameters θ and t , g is a deterministic function, and $z_t \sim \mathcal{N}(0, I)$ is random noise. Suppose the outcome from this model is the state at the end of the simulation, $X = X_T$. In order to compute the likelihood of this model, we need to integrate over all the noise samples $z = [z_1, \dots, z_{T-1}]$,

$$p(X|\theta) = \int p(X, z|\theta)dz. \quad (\text{C4})$$

In general, approximating this integral is significantly more expensive than generating a sample from $p(X|\theta)$ by simulating, and so we call this likelihood intractable

For our forward model, the intractability of the likelihood follows from the fact that we select the closest layer to the IRH after sampling the noise. Therefore, the noise model cannot be efficiently calculated as in Eqn (C2). Note that, even in the case of tractable likelihood functions, it might be beneficial to use ‘likelihood-free’ inference methods, for example in the case that the forward model is computationally expensive to compute.

C.2. Normalizing flows

While NPE can be performed with any valid parameterization of the variational distribution $q_{\phi}(\theta|X)$, in this work, we make the common choice to use a normalizing flow model. In the following, we adapt the notation of Papamakarios and others (2019a).

Algorithm 1: Noise model calibration

Inputs: (Noiseless) simulated layer elevations $\mathbf{e}_l^{(k)}(\mathbf{x})$ for $l = 1, \dots, L$, $k = 1, \dots, K_{\text{cal}}$ and simulation grid \mathbf{x} , IRH interpolated on the same simulation grid, $\mathbf{e}_m(\mathbf{x})$

Outputs: Mean and standard deviation $\mu_{\beta_n}, \sigma_{\beta_n}$ defining noise model

```

for  $k \leftarrow 1$  to  $K_{\text{cal}}$  do
     $l^* \leftarrow \arg \min_l \|\mathbf{e}_l^{(k)}(\mathbf{x}) - \mathbf{e}_m(\mathbf{x})\|_2^2$ ; # best fitting layer to IRH
     $\tilde{\mathbf{e}}_{l^*}^{(k)}(\mathbf{x}) \leftarrow \mathbf{e}_{l^*}^{(k)}(\mathbf{x}) - \text{lowpass\_filter}(\mathbf{e}_{l^*}^{(k)}(\mathbf{x}))$ ; # detrend best layer
     $\tilde{\mathbf{e}}_m(\mathbf{x}) \leftarrow \mathbf{e}_m(\mathbf{x}) - \text{lowpass\_filter}(\mathbf{e}_m(\mathbf{x}))$ ; # detrend IRH
     $\gamma_1^{(k)}, \dots, \gamma_n^{(k)} \leftarrow \text{PSD} \left( [\mathbf{e}_m(\mathbf{x}) - \tilde{\mathbf{e}}_m(\mathbf{x})] - [\mathbf{e}_{l^*}^{(k)}(\mathbf{x}) - \tilde{\mathbf{e}}_{l^*}^{(k)}(\mathbf{x})] \right)$ ; # power spectral density of difference
     $\beta_1^{(k)}, \dots, \beta_n^{(k)} \leftarrow \log(\gamma_1^{(k)}), \dots, \log(\gamma_n^{(k)})$ 
for  $n \leftarrow 1$  to  $N$  do
     $\mu_{\beta_n} \leftarrow \frac{1}{K_{\text{cal}}} \sum_{k=1}^{K_{\text{cal}}} \beta_n^{(k)}$ ; # empirical mean of coefficients
     $\sigma_{\beta_n} \leftarrow \sqrt{\frac{1}{K_{\text{cal}}-1} \sum_{k=1}^{K_{\text{cal}}} (\beta_n^{(k)} - \mu_{\beta_n})^2}$ ; # empirical standard deviation of coefficients
return  $\mu_{\beta_1}, \sigma_{\beta_1}, \dots, \mu_{\beta_N}, \sigma_{\beta_N}$ 

```

The goal of normalizing flows is to learn a map from a distribution $u \sim p_u(u)$ to a distribution of the same dimensionality $x \sim p_x(x)$. Here, $p_x(x)$ is the target distribution, and $p_u(u)$ is a simple distribution that can easily be evaluated and sampled from (e.g. a multivariate Gaussian). If we can express x as a differentiable, invertible transformation of u , i.e. $x = T(u)$, then we can also evaluate the probability

$$p_x(x) = p_u(u) |\det J_T(u)|^{-1}, \tag{C5}$$

where $u = T^{-1}(x)$ and J_T is the Jacobian matrix of the transformation T .

Normalizing flows are then defined by a sequence of transformations T_1, \dots, T_K which transform $z_0 = u$ to $z_K = T_K \circ \dots \circ T_1(u) = x$. Each of the transformations is learnable. The composition of differentiable, invertible transformations is also invertible and differentiable, with the determinant of the Jacobian satisfying

$$\det J_{T_2 \circ T_1}(u) = \det J_{T_2}(T_1(u)) \cdot \det J_{T_1}(u). \tag{C6}$$

One common way to parameterize the individual transformations T_i is through *coupling transforms*, where the input z is split into two parts, $z = [z_{1:d}, z_{d+1:D}]$, where D is the dimensionality of $z \in \mathbb{R}^D$. The first part, $z_{1:d}$, remains unchanged. The second part is transformed elementwise, $z'_i = \tau(z_i; h_i)$, where τ is some monotonic function of z_i conditioned on $h_i = F(z_{1:d})$, for some learnable function F of the first d components of z . The vector is permuted between each layer, so that not the same components are mapped through the identity with each transformation. An advantage of this parameterization is that the Jacobian matrix can be calculated as the product of its diagonal components, making the evaluation of the log-probabilities significantly faster. In this work, we use neural spline flows (NSFs, Durkan and others 2019). In NSFs, τ are monotonic spline functions, which are analytically invertible, yet highly flexible.

C.3. Connection between NPE loss and Kullback-Leibler divergence

We note that the expected forward KL divergence between the true posterior $p(\theta|X)$ and the variational distribution $q_\phi(\theta|X)$ can be decomposed as

$$\begin{aligned} \mathbb{E}_{p(x)} [D_{\text{KL}}(p(\theta|X) \| q_\phi(\theta|X))] &= \mathbb{E}_{p(x)p(\theta|X)} \left[\log \frac{p(\theta|X)}{q_\phi(\theta|X)} \right] \\ &= -\mathbb{E}_{p(x)p(\theta|X)} [\log q_\phi(\theta|X)] + \mathbb{E}_{p(x)p(\theta|X)} [\log p(\theta|X)], \end{aligned} \tag{C7}$$

where the first term on the right hand side is the negative of the loss in Eqn (8), and the second term is independent of the variational parameters ϕ and is thus a constant with respect to the variational parameters ϕ which we optimize. Therefore, minimizing the NPE loss (Eqn (8)) is equivalent to minimizing the expected forward KL divergence between the variational and true posterior distributions.

C.4. Neural Likelihood Estimation

In our work, we use NPE to directly approximate the posterior distribution. However, other variants of SBI exist. In particular, a prominent method of SBI is NLE (Wood, 2010; Papamakarios and others, 2019b). In NLE, the goal is to approximate the likelihood function $p(X|\theta)$ from a dataset of simulations $\{\theta_k, X_k\}_{k=1}^K$. This likelihood can then be used to perform Bayesian inference using the existing set of Bayesian inference methods, such as Markov chain Monte Carlo (MCMC), or variational inference (Blei and others, 2017).

Similarly to NPE, we define a variational distribution (e.g. a normalizing flow), $q_\phi(X|\theta)$, which can be trained by minimizing the negative log-probabilities of the model outcomes X_k given model parameters θ_k

$$\mathcal{L}(\phi) = \mathbb{E}_{\theta_k \sim p(\theta), X_k \sim p(X|\theta_k)} [-\log q_\phi(X_k|\theta_k)]. \tag{C8}$$

Notice that the conditioning of θ_k, X_k is reversed relative to Eqn (8).

The likelihood model learned in NLE can also be thought of as an emulator of the forward model, as it allows us to draw samples $X \sim p(X|\theta)$. This makes NLE methods advantageous also for problems where the likelihood is tractable, as the learned emulators can typically be evaluated faster than the original forward model. Related ideas have been explored for glaciological problems. For example, Tarasov and others (2012) use Bayesian neural networks as a surrogate model for particular parameter-output pairs in a glacial systems model, which they then use to perform inference with MCMC. Similarly, Brinkerhoff and others (2021) use a deterministic residual neural network as a surrogate model to predict a low-dimensional representation of outcomes from a sub-glacial hydrology model. They then derive an approximation of the uncertainty in this model to arrive at a nondeterministic likelihood which can then be used for inference.

Appendix D. Further implementation details

For completeness, we give the values of the important hyperparameters involved in the workflow. We provide the details of the spatial and temporal resolutions of our various simulations, along with the regularization strengths of the smoothing of the EIS data (Tables D1, D2 and D3).

During inference, we applied NPE as implemented in the `sbi` package (Tejero-Cantero and others, 2020) to obtain the results for both the synthetic (Section 3) and Ekström (Section 4) ice shelves. We used the NSF (Durkan and others, 2019) as implemented in Lueckmann and others (2021), and with the same architecture of five flow transformations, two residual blocks of 50 hidden units each, ReLU nonlinearity and 10 bins. We also embedded the 500-dimensional observation of layer elevations to a 50-dimensional summary statistic used as the condition for the NSF. The embedding network consisted of two convolutional layers with kernel size 5, each followed by ReLU activations and max pooling with kernel size 2. The number of output channels for the two convolutional layers were 6 and 12, respectively. The output channels of the second convolutional layer were then concatenated and fed through two

Table D1. Hyperparameters for synthetic ice shelf spin-up modeling

Parameter	Value
Mesh resolution (x, y)	(310 m, 250 m)
Spin-up duration	1000 years
Spin-up time step	1.0 years
Boundary conditions	Dirichlet inflow and side boundaries

Table D2. Hyperparameters for the preprocessing of the data for Ekström Ice Shelf

Parameter	Value
Mesh resolution	300 m
Thickness smoothness reg. penalty	1000
Log fluidity reg. penalty	1000
Boundary conditions	Dirichlet inflow and side boundaries

Table D3. Layer tracer forward model simulation configuration

Parameter	Value
Simulation time	1000 years
Time step	0.5 years

fully connected linear layers, each followed by ReLU activations. The number of hidden units was set to 50. Training was done as in Lueckmann and others (2021), with the exception that the batch size was set to 1000 (default is 50). For each NPE run, we train five networks initialized with different random seeds and report in our results the run with the best validation loss.

Following the work of Lueckmann and others (2021), we split the simulation dataset into training and validation datasets with a 90–10 split. We optimize the loss in Eqn (8) using an Adam Optimizer (Kingma and Ba, 2015) with a batch size of 50, a learning rate of 0.0005, with the maximum gradient norm clipped to 5.0. For each training epoch, we calculate the validation loss on the entire validation dataset. If the validation loss has not surpassed its best value for 20 training epochs, we assume convergence and stop training.

Appendix E. Computational costs

We provide a breakdown of the approximate computational costs of the different stages in our workflow for both synthetic and Ekström Ice Shelves in Tables E1 and E2, respectively. These are dependent the hardware used and vary stochastically as a result of random number generators. This section is intended to provide intuition into the relative scales of the different stages of the workflow, rather than exact measurements. We had access to 16-core Intel Xeon Gold 2.9 GHz CPU nodes and Nvidia RTX 2080ti GPU nodes. While the large number evaluations of the forward layer tracing model were by far the most computationally intensive section of the workflow in both cases, these simulations were trivially performed in parallel across CPU cores, thus reducing the wall-clock time of the workflow.

This analysis highlights the advantages of our amortized approach to inference. For EIS, the total computational time of the noiseless simulations accounted for $\approx 99.8\%$ of the total computational time of the workflow. The simulations need not be repeated when we infer from other IRHs, greatly benefiting the computational efficiency of inference as the number of IRHs in the dataset increases. This advantage is slightly reduced when considering the parallelization we have used (Table E2), as the training of the probability density

Table E1. Synthetic ice shelf approximate computational cost breakdown. Some tasks are embarrassingly parallelizable—parallel resources and times are shown in square brackets. All times reported in minutes

Task	Node [parallel]	Time [parallel]
Spin-up of ice shelf	CPU	120
1000 calibration simulations	CPU [100 cores]	1000 [10]
189 000 noiseless simulations	CPU [100 cores]	2×10^5 [2000]
Noise and layer selection	CPU [100 cores]	400 [20]
Training NPE for 1 IRH	GPU	30

Note: Only the last two tasks need to be repeated for each IRH measurement.

Table E2. Ekström Ice Shelf approximate computational cost breakdown. Some tasks are embarrassingly parallelizable—parallel resources and times are shown in square brackets. All times reported in minutes

Task	Node [parallel]	Time [parallel]
Generating mesh	CPU	< 1
Data preprocessing	CPU	10
1000 calibration simulations	CPU [100 cores]	1000 [10]
189 000 noiseless simulations	CPU [100 cores]	2×10^5 [2000]
Noise and layer selection	CPU [100 cores]	400 [20]
Training NPE for 1 IRH	GPU	30

Note: Only the last two tasks need to be repeated for each IRH measurement.

estimator is not easily parallelizable across GPU nodes. Accounting for this still results in $\approx 97.6\%$ of the computational cost being amortized.

Appendix F. Additional results

We show the posterior and posterior predictive distributions when inferring from isochronal layer 3 in the synthetic ice shelf dataset, of age 150 years (Fig. F1). This isochronal layer has average depth of 120 m, comparable to the deepest IRH in the Ekström dataset. While the uncertainty is much higher than for the shallow layer, the posterior over the surface accumulation still shows a higher mean than the prior in the downstream section of the ice shelf. Additionally, the posterior predictive reconstructs the ground truth isochronal layer better than the prior predictive. The mean RMSE of the posterior predictive layers relative to the ground truth is 13.6 m, compared to 19.8 m for the prior. Therefore, we are still able to reconstruct additional information about the mass balance parameters, even from much deeper layers.

We also explore the dependence of the inferred posterior over surface accumulation rate on the depth of the layer used for inference in the synthetic case (Fig. F2), similar to the analysis done in Section 5.2 for EIS. For the synthetic ice shelf, the surface accumulation and basal melt rates were held constant for the entire simulation time, and hence the increased uncertainty with depth seen in Fig. F2 highlights that information about the mass balance parameters is gradually lost with time as a result of the action of the simulator. Indeed, for the deepest ground truth isochronal layer of average depth 183 m, the posterior distribution is almost identical to the prior distribution.

Finally, we report the RMSE in the predicted isochronal layer elevations, relative to the true IRH (for EIS) or the ground truth isochronal layer (for the synthetic ice shelf). This is done across simulations from 1000 simulations for each of the prior and posterior distributions, for each IRH. The RMSE is consistently lower for the posterior predictive distribution for all depths, for both the synthetic ice shelf (Table F1) and EIS (Table F2).

Appendix G. Kottas traverse data

Here we describe the mapping of the surface accumulation measurements on Kottas traverse to the flow line transect. For each measurement year, and each location \tilde{x}_i on the Kottas traverse, we find the nearest point x_i on the flow line transect. We assume the accumulation rate at this point to be normally distributed, with mean \tilde{a}_i (the Kottas traverse measurement at \tilde{x}_i), and variance $\sigma_a^2 \|\tilde{x}_i - x_i\|_2^2 / l_a$. We set the length scale $l_a = 2.5$ km and the accumulation rate variance $\sigma_a^2 = 0.25^2 \text{ m}^2 \text{ m}^{-2}$. These values are chosen in accordance to the definition of the surface accumulation rate prior distribution (Section 2.2.3).

The yearly variations of the estimated surface accumulation as measured using the stake line along the Kottas traverse (Fig. G1) reasonably agree with the posterior inferred using IRH 2. These show that there is a high year-to-year variability in the surface accumulation (Mengert 2018, Fig. 13) even in this steady-state region in East Antarctica. However, the mean of these yearly measurements matches the inferred posterior mean closely.

Appendix H. Synthetic results with miscalibrated prior

An additional outcome of our approach is the estimation of the age of isochronal layers. However, the validity of this estimate depends on the prior distribution containing the true mass-balance parameters. When the true mass-balance

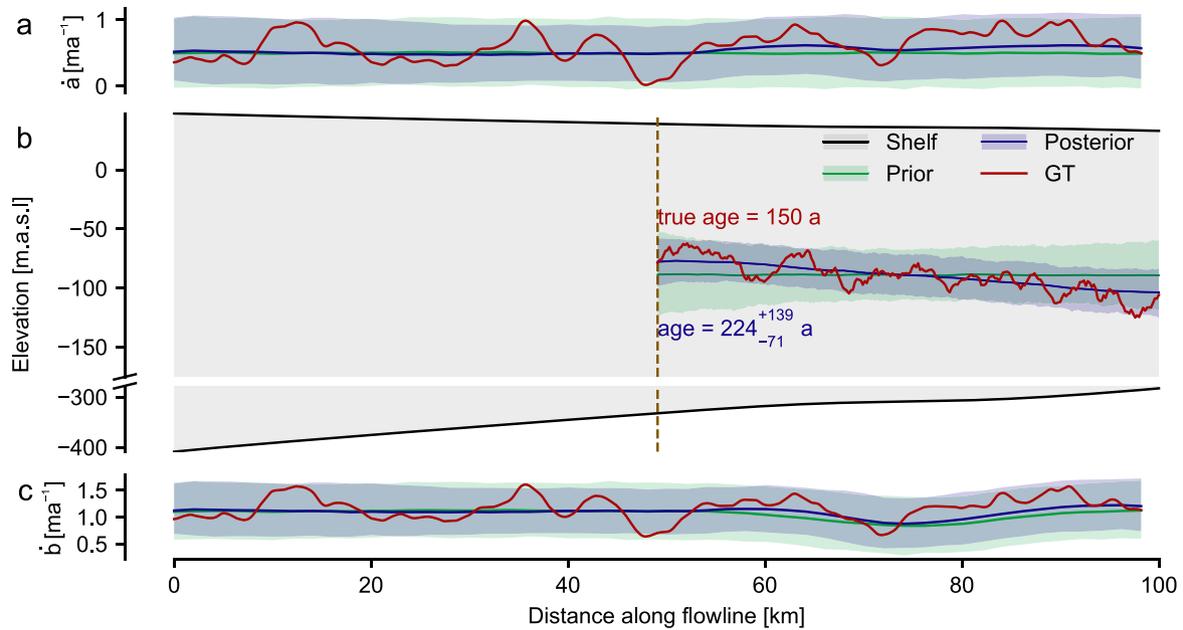


Figure F1. Prior and posterior (predictive) for the synthetic dataset. (a and c) Prior and posterior over surface accumulation and basal melt rates respectively for layer 3 of the synthetic ice shelf, of age 150 years. Solid line is the distribution mean, the shaded region represents the 5th and 95th percentiles. The ground truth (GT) parameters used to generate the reference isochronal layer are also shown. (b) Cross section of the ice shelf. Prior and posterior predictive distributions for the layer closest matching the ground truth isochronal layer. The vertical dashed line represents the LMI boundary for this isochronal layer. The posterior predictive reconstructs the observed layer with higher accuracy and lower uncertainty. The posterior predictive distribution of the age of the isochronal layer is 224⁺¹³⁹₋₇₁ years.

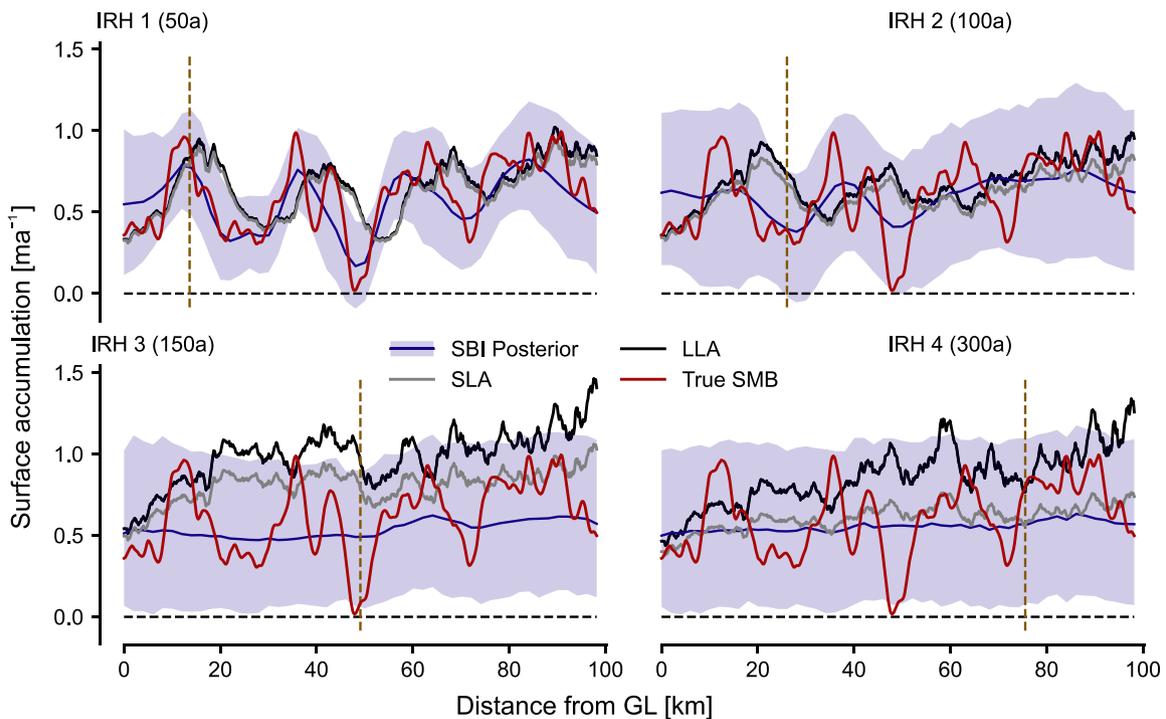


Figure F2. Synthetic shelf: dependence of posterior surface accumulation rate on depth of layer used for inference.

parameters have low probability under the prior distribution, the resulting estimate for the age of the isochronal layer can be wrong. We test this on the synthetic ice shelf by choosing a ground truth surface accumulation \dot{a}_{mis} that has low probability under the prior distribution. We calculate the isochronal layer of age 100 years under this ground truth and obtain the posterior distribution as before, using the same set of simulations as in the main text. The posterior

does not capture that the mean surface accumulation rate should be higher than what is defined in the prior (Fig. H1a). However, this is not a failure of the inference method, as we can see that the posterior predictive still reconstructs the ground truth isochronal layer at higher fidelity than the prior (Fig. H1b). The predicted age of the isochronal layer 164⁺¹⁰¹₋₄₄ years, which greatly overestimates the true age of 100 years.

Table F1. Synthetic ice shelf—prior and posterior predictive distribution root-mean-square error (RMSE) relative to ground truth IRH, estimated from 1000 samples. The mean and standard deviations (SDs) in the RMSE are reported. All values are in meters

IRH number	Prior		Posterior	
	RMSE mean	RMSE SD	RMSE mean	RMSE SD
1	11.5	3.9	3.9	0.5
2	16.0	7.6	7.3	1.2
3	19.8	8.4	13.6	3.5
4	22.1	7.9	19.8	6.8

Table F2. Ekström Ice Shelf—prior and posterior predictive distribution root-mean-square error (RMSE) relative to ground truth IRH, estimated from 1000 samples. The mean and standard deviations (SDs) in the RMSE are reported. All values are in meters

IRH number	Prior		Posterior	
	RMSE mean	RMSE SD	RMSE mean	RMSE SD
1	6.8	2.1	3.0	0.9
2	11.8	3.4	4.6	1.3
3	17.0	6.6	6.8	1.4
4	16.4	7.6	10.0	2.1

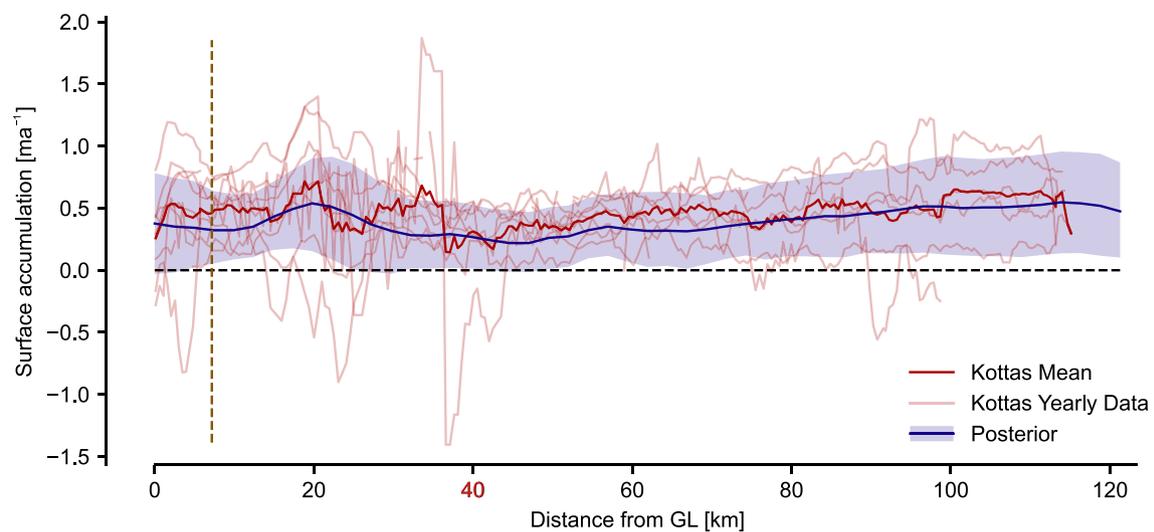


Figure G1. Yearly variations of the Kottas surface accumulation stakes measurement dataset. Years shown are 1995–2005 and 2017–19. These are compared with the posterior distribution inferred using IRH 1 of the Ekström IRH dataset.

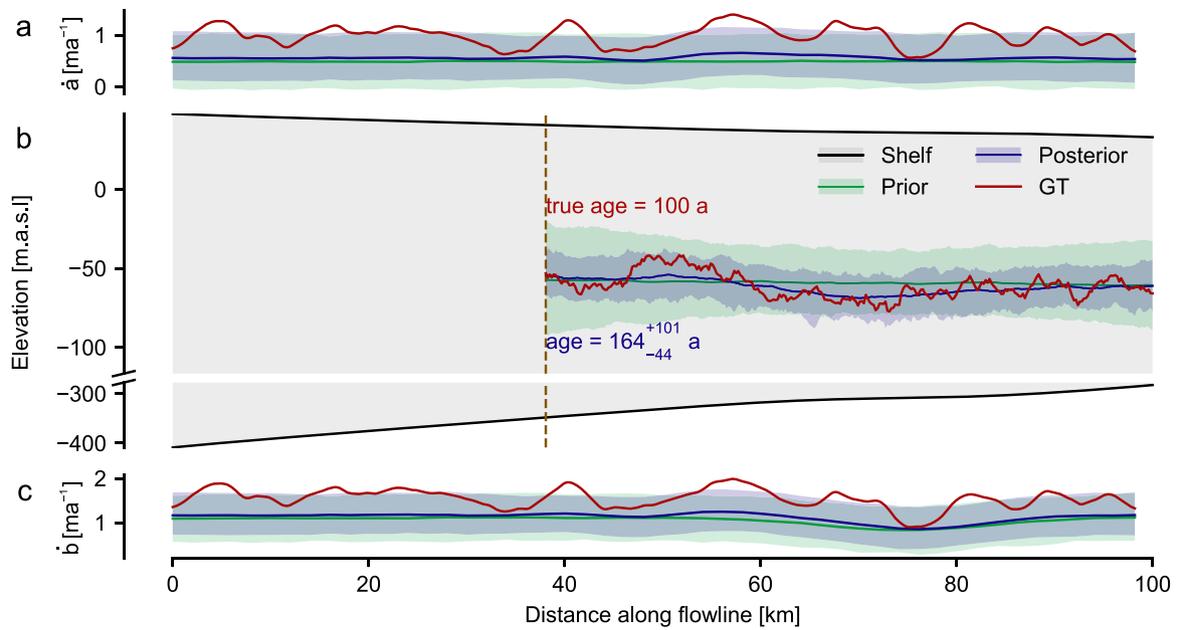


Figure H1. Prior and posterior (predictive) for the synthetic ice shelf with the low-probability ground truth. (a and c) Prior and posterior over surface accumulation and basal melt rates respectively for an isochronal layer of age 100 years. Solid line is the distribution mean, the shaded region represents the 5th and 95th percentiles. The ground truth (GT) parameters used to generate the reference isochronal layer are also shown. (b) Cross section of the ice shelf. Prior and posterior predictive distributions for the layer closest matching the ground truth isochronal layer. The vertical dashed line represents the LMI boundary for this isochronal layer. The posterior predictive reconstructs the observed layer with higher accuracy and lower uncertainty. The posterior predictive distribution of the age of the isochronal layer is 164^{+101}_{-44} years.

Appendix I. Radar data

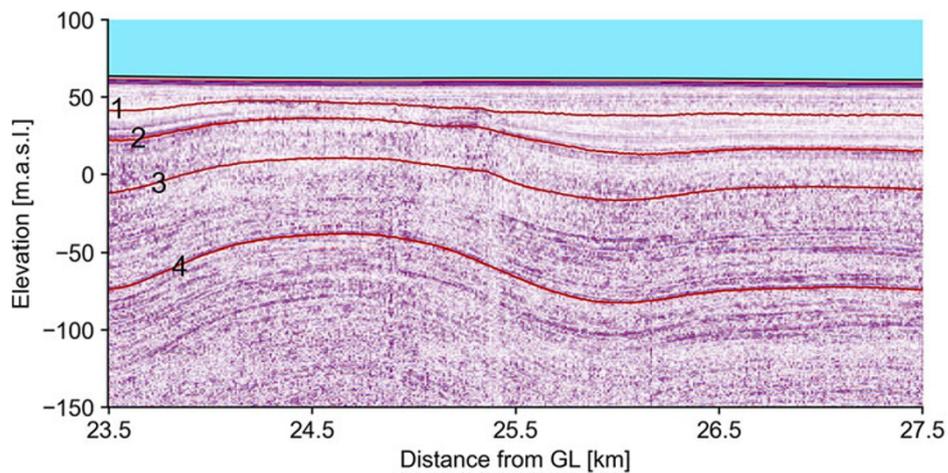


Figure I1. Radargram along transect. Zoom in on section of vertical cross-section view of the radar transect (Figure 5c). Color gradient indicates radargram data from radar survey of transect. The four labeled IRHs picked from this radargram are labeled in order of depth.