ORIGINAL PAPER

Multifidelity uncertainty quantification for ice sheet simulations

Nicole Aretz[1] · **Max Gunzburger[1]** · **Mathieu Morlighem[2]** · **Karen Willcox[1]**

Abstract

Ice sheet simulations suffer from vast parametric uncertainties, such as the basal sliding boundary condition or geothermal heat flux. Quantifying the resulting uncertainties in predictions is of utmost importance to support judicious decision-making, but high-fidelity simulations are too expensive to embed within uncertainty quantification (UQ) computations. UQ methods typically employ Monte Carlo simulation to estimate statistics of interest, which requires hundreds (or more) of ice sheet simulations. Cheaper low-fidelity models are readily available (e.g., approximated physics, coarser meshes), but replacing the high-fidelity model with a lower fidelity surrogate introduces bias, which means that UQ results generated with a low-fidelity model cannot be rigorously trusted. Multifidelity UQ retains the high-fidelity model but expands the estimator to shift computations to low-fidelity models, while still guaranteeing an unbiased estimate. Through this exploitation of multiple models, multifidelity estimators guarantee a target accuracy at reduced computational cost. This paper presents a comprehensive multifidelity UQ framework for ice sheet simulations. We present three multifidelity UQ approaches—Multifidelity Monte Carlo, Multilevel Monte Carlo, and the Best Linear Unbiased Estimator—that enable tractable UQ for continental-scale ice sheet simulations. We demonstrate the techniques on a model of the Greenland ice sheet to estimate the 2015-2050 ice mass loss, verify their estimates through comparison with Monte Carlo simulations, and give a comparative performance analysis. For a target accuracy equivalent to 1 mm sea level rise contribution at 95 % confidence, the multifidelity estimators achieve computational speedups of two orders of magnitude.

Keywords Uncertainty quantification · Multi-fidelity · Ice sheet simulation · Surrogate modeling

1 Introduction

Sea level rise is impacting both coastal ecosystems and our societies. Mass loss from ice sheets is becoming a major driver of sea level change, and large-scale simulations of the Greenland and Antarctic ice sheets have a central role when evaluating policies to combat climate change. Since these simulations must be over time horizons spanning decades and centuries, modeling decisions can significantly influence the predicted ice mass change and thus the predicted contribution to sea level rise. Even among state-of-the-art high-fidelity ice sheet models, as illustrated by model intercomparison studies [1–4], expert opinions on modeling parameterizations and datasets differ, and there is high variance in ice mass loss projections. To provide effective support to decision-making, it is paramount to quantify the uncertainty associated with these simulation-based projections, yet doing so is computationally prohibitive because it would require an ensemble of high-fidelity simulations to be run over many different modeling choices. In this paper we lay out a multifidelity uncertainty quantification (UQ) framework that addresses this challenge.

Model intercomparison studies have a long tradition in glaciology, including in model validation [5–7], verification based on historical data [2, 4], and for high-fidelity projections [1, 3, 8]. Moreover, parametric uncertainties in ice

✉ Nicole Aretz
  nicole.aretz@austin.utexas.edu

  Max Gunzburger
  max.gunzburger@austin.utexas.edu

  Mathieu Morlighem
  Mathieu.Morlighem@dartmouth.edu

  Karen Willcox
  kwillcox@oden.utexas.edu

[1] Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, 201 E 24th St, Austin 78712, Texas, USA

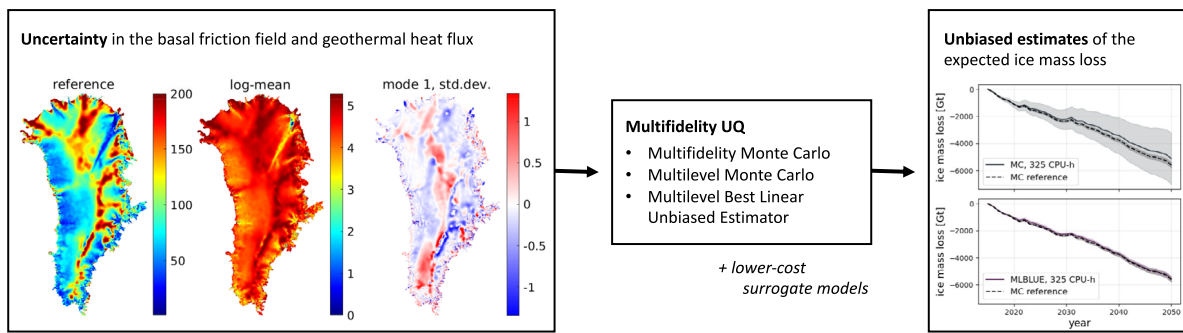[2] Department of Earth Sciences, Dartmouth College, Hanover 03755, NH, USA

Fig. 1 Multifidelity UQ framework

sheets are studied extensively and explored often through sensitivity analyses, e.g., [9–11] for the geothermal heat flux, or [12–14] for the basal friction. Acknowledging parametric uncertainty and its importance for high-fidelity projections, most ice sheet models infer the basal friction field from observational data of the surface ice velocity in a deterministic [15, 16] or more recently Bayesian [17–19] inverse problem; other parameters like the geothermal heat flux [20, 21] or the basal topography [22–24] are inferred once and then distributed throughout the community through datasets. While these are examples of *inverse* UQ, where data is used to reduce the uncertainty in a parameter, in this paper we are primarily concerned with *forward* UQ, where the goal is to quantify the influence of parametric uncertainty on an output of interest (OoI) such as the predicted ice mass loss (see Fig. 1).

Forward UQ is challenging for computationally expensive ice sheet simulations since typically a large ensemble of projections is required to approximate the statistics (e.g., the mean) of the OoI under a given probabilistic description of parametric uncertainties. In [25, 26], expert judgements were pooled and formalized to quantify the uncertainty in existing ice sheet projections. In [27], a mass-balance model of so-called "intermediate complexity" was used in a Monte Carlo (MC) estimation with 5,000 samples, but the projection was limited to a time horizon of 400 days and a glacier of approximately $16\,km^2$. Other studies have balanced required ensemble sizes and computational costs by accepting a $10\,km$ resolution [28, 29], using hybrid or approximated physics models [30, 31], or built probabilistic emulators [32–36] trained on high-fidelity simulations. A drawback in using these modeling simplifications is the model bias introduced by replacing the high-fidelity ice sheet model with a cheaper surrogate model. In doing so, there is no guarantee that the UQ results will reflect what would have been obtained using the most trusted high-fidelity model. Multi-model ensemble studies have also been employed to compare the deterministic projections of several high-fidelity models, each built to reflect their modeller's best domain expertise. However,

because these high-fidelity models are so expensive, only small ensembles are possible. Moreover, most ensemble studies do not account for the uncertainties in each individual model. Multifidelity UQ methods can overcome these challenges: Formal UQ at a desired target accuracy is achieved by leveraging surrogate models to reduce the computational cost, but in a structured way that guarantees a statistically unbiased estimate of the high-fidelity OoI statistic.

Multifidelity UQ methods exploit the correlation between the high-fidelity model and less accurate but computationally cheaper surrogate models to construct an unbiased estimator of the high-fidelity expectation. Compared to MC sampling, multifidelity estimators achieve an improved accuracy for any prescribed computational budget (or equivalently, they achieve a given target accuracy with a reduced computational budget). In this paper we present three predominant multifidelity UQ methods, and highlight their applicability for ice sheet simulations: the Multifidelity Monte Carlo (MFMC) method [37, 38], the Multilevel Monte Carlo (MLMC) method [39, 40], and the Multilevel Best Linear Unbiased Estimator (MLBLUE) [41]. Our paper has three objectives: (1) To establish the necessity of unbiased UQ for ice mass loss projections by showing the errors that can be incurred by using approximate models without a formal multifidelity framework; (2) to demonstrate the efficiency of multifidelity UQ methods for computationally expensive glaciology simulations; and (3) to provide a simple but flexible algorithmic framework to implement, use, and interpret MFMC, MLMC, and MLBLUE on any ice sheet code without prior experience in UQ or surrogate modeling. To emphasize the relevance and applicability of our work, we employ a community ice sheet code — the Ice-sheet and Sea-level System Model (ISSM, [42]) — and follow the projection protocols [43] of the Ice Sheet Model Intercomparison Project (ISMIP6, [44]) contribution to Coupled Model Intercomparison Project Phase 6 (CMIP6, [45]), with the goal to make the transfer of techniques to similar applications as easy as possible.

In Section 2, we describe our high-fidelity model of the Greenland ice sheet, describe uncertainties in the basal friction and geothermal heat flux fields introduced by different data sets, and show the large variance these cause in the high-fidelity projections. We also discuss different types of surrogate models, and illustrate their model bias. In Section 3 we provide primers to the MFMC, MLMC, and MLBLUE methods, each with a detailed algorithmic description to facilitate their implementation. In Section 4, we demonstrate these methods for estimating the expected ice mass loss for the year 2050, and discuss the benefits of multifidelity UQ for ice sheet simulations.

2 Model of the Greenland Ice Sheet

This section sets up the high-fidelity model of the Greenland ice sheet, which we use to demonstrate the necessity and challenges of UQ for ice sheet simulations. Our high-fidelity model comprises governing equations for ice temperature, velocity and thickness (Section 2.1). We describe uncertainties in the basal friction field and the geothermal heat flux that stem from the variety of available datasets in the literature (Section 2.2). The predictive uncertainty thus reflects the influence that datasets have on model simulations. This section also provides brief descriptions of some of the many surrogate models readily available in ice sheet codes (Section 2.3), including coarse-grid approximations, simplified-physics models, and statistical emulators.

2.1 High-Fidelity Model

We consider a model of the Greenland ice sheet that consists of three coupled nonlinear partial differential equations (PDEs), namely

- a thermal model, which governs changes in the temperature T within the ice sheet, based on the conservation of energy;
- a dynamical model with nonlinear rheology for the velocity vector $\mathbf{v} = (v_x, v_y, v_z)^\top$, which governs the motion of the ice sheet, based on the conservation of momentum;
- a mass transport model, which governs changes in the thickness h of the ice sheet, based on the conservation of mass.

Out of these variables, the temperature T and the velocity components v_x, v_y, and v_z in x-, y-, and z-direction are each three-dimensional fields, defined on the ice sheet domain $\Omega(t)$; the ice thickness h is two-dimensional, defined on the horizontal extent Ω_{2D} of the ice sheet, and determines how the ice geometry $\Omega(t)$ evolves in time t vertically.

The specific three-part system we consider is a simplification of more sophisticated ice sheet models[1] that incorporate additional ice sheet features such as ice–ocean interactions and basal hydrology. Despite these simplifications, the system we consider is already at a scale where UQ tasks are computationally demanding, and suffices to demonstrate how multifidelity UQ techniques can obtain trusted estimates of statistical OoIs at a much lower computational cost than MC estimation.

2.1.1 Geometry and initialization

We model the main part of the Greenland ice sheet under the shared economic pathway scenario SSP1-2.6 ([46, 47], chapter 1.6), which is a low emission scenario. For our atmospheric forcing, we follow the experimental protocol [43] of the ISMIP6 Greenland study [1]. Consequently, the projections for our multifidelity UQ study start in the year 2015, and we perform a spin-up run from our model initialization in $t_0 := 1989$ to obtain the 2015 initial conditions (described below). For facilitating comparisons with Monte Carlo sampling, it suffices that we limit our projections to the years 2015 to $t_f := 2050$, though the methodology applies analogously to longer projection regimes.

For our domain outline, we traced any ice of thickness greater than 5 m reported in [22, 48], removed all disconnected parts (e.g., islands, numerical artifacts) and major bottlenecks, and smoothed the obtained outline. Our final outline encompasses approximately 86.64 % of the ice-covered area in [49] (version 2, with coastline from Jeremie Mouginot) and approximately 99.59 % of the total ice mass in [22, 48]. Using the 1995–2015 averaged surface ice velocity from [50] and the ice thickness [22, 48] for reference, we create our high-fidelity mesh with a resolution varying from 100 m to 15 km. We denote this domain by Ω_{2D}, and keep it fixed throughout all simulations,

To obtain the basal topography Γ_b, we interpolate the ice base from [22, 48] onto our mesh. Similarly, for the surface topography $\Gamma_s(t_0)$, we interpolate the ice altitude data from [49] (version 2). Following [22], these initializations give us a geometry consistent with the year 2009. For any $t_0 \leq t \leq 2009$, we thus keep the geometry fixed, with a constant ice thickness $h(t) = \Gamma_s(t_0) - \Gamma_b$. Starting from the year $t = 2009$, $h(t)$ obeys the ice thickness equations defined in Section 2.1.4, and the ice surface topography $\Gamma_s(t) = \Gamma_b + h(t)$ evolves accordingly. For 3D variables, the 2D domain Ω_{2D} and its mesh are extruded with five layers, starting from

---

[1] Removing the simplifications would result in even better performance of the multifidelity UQ estimation approaches discussed in this paper, but then the reference high-fidelity UQ results would become computationally prohibitive, preventing us from analyzing the performance of the multifidelity approaches.

the ice base Γ_b upwards to the ice surface $\Gamma_s(t)$. We denote the obtained 3D domain by $\Omega(t)$, using the variable t to stress its variability in time.

To initialize our model in $t_0 = 1989$, we compute the steady-state equilibrium of the ice temperature T and velocity $\mathbf{v}$ with CNRM-CM6-1 1960–1989 reference forcing [51]. This initialization mimics that the Greenland ice sheet is believed to have been in steady state in the 1960–1989 time period [52, 53]. We then run our model using the equations described in the following subsections, until the start year $t = 2015$ prescribed in [43] to obtain the initial conditions of temperature T, velocity $\mathbf{v}$, and ice thickness h used in our predictive runs.

2.1.2 Stress balance model

There exists a hierarchy of ice sheet dynamical models that can be ordered according to decreasing physical fidelity and which, for the most part, also possess decreasing computational simulation costs. At the top of the hierarchy is the most generally accepted (with respect to physical fidelity) *full-Stokes model* (FS). However, for a given ice geometry $\Omega(t)$ and a given temperature field $T(t)$, the numerical solution of the FS model is challenging and expensive for several reasons, which include the usage of fine grids and long time horizons; stable and sufficiently accurate spatial and temporal discretization choices (e.g., finite element and time-stepping schemes, respectively); and the need to solve for four field variables over the 3D domain (three velocity components and the pressure). As a result, few ice sheet codes support the FS equations. In significantly greater use by practitioners is the *higher-order model*[2] (HO), which is an ice sheet dynamical model representing the next lower level of physical fidelity from FS. For reference, in the ISMIP6 Greenland study [1], none of the 21 participating models solved the FS system, but seven solved the HO equations. The remaining 14 groups opted for further physical approximations in the form of the Shallow-Shelf or Shallow-Ice Approximations, or their hybrid combination ([1], Table A1).

The HO model is derived as a simplification of the FS model by taking advantage of the small vertical-to-horizontal aspect ratio of the ice sheet. The net result is that the horizontal gradients of vertical velocities are neglected compared to vertical gradients of horizontal velocities, and bridging effects are also neglected. The resulting HO model is then

given by the system of PDEs

$$\nabla \cdot (2\mu\dot{\varepsilon}_{\mathrm{HO},1}) = \rho_{\mathrm{ice}} g \frac{\partial s(t)}{\partial x} \tag{1}$$

$$\nabla \cdot (2\mu\dot{\varepsilon}_{\mathrm{HO},2}) = \rho_{\mathrm{ice}} g \frac{\partial s(t)}{\partial y} \tag{2}$$

for $\mathbf{x} = (x, y, z) \in \Omega(t)$ and $t_0 < t \leq t_f$, and where $\rho_{\mathrm{ice}} = 917\,\mathrm{kg/m^3}$ is the ice density, $g = 9.81\,\mathrm{m/s^2}$ is the gravitational acceleration, $s(x, y; t)$ is the altitude of the surface $\Gamma_s(t)$ vertically above the point $\mathbf{x} = (x, y, z)$, and the strain rates $\dot{\varepsilon}_{\mathrm{HO},1}, \dot{\varepsilon}_{\mathrm{HO},2}$ take the form

$$\dot{\varepsilon}_{\mathrm{HO},1} = \begin{pmatrix} 2\dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_y}{\partial y} \\ \dfrac{1}{2}\dfrac{\partial v_x}{\partial y} + \dfrac{1}{2}\dfrac{\partial v_y}{\partial x} \\ \dfrac{1}{2}\dfrac{\partial v_x}{\partial z} \end{pmatrix},$$

$$\dot{\varepsilon}_{\mathrm{HO},2} = \begin{pmatrix} \dfrac{1}{2}\dfrac{\partial v_x}{\partial y} + \dfrac{1}{2}\dfrac{\partial v_y}{\partial x} \\ \dfrac{\partial v_x}{\partial x} + 2\dfrac{\partial v_y}{\partial y} \\ \dfrac{1}{2}\dfrac{\partial v_y}{\partial z} \end{pmatrix}.$$

The effective ice viscosity μ follows the generalized Glen's flow law

$$\mu = \frac{B(T)}{\sqrt{2}} \left\| (\dot{\varepsilon}_{\mathrm{HO},1}, \dot{\varepsilon}_{\mathrm{HO},2}) \right\|^{-\frac{n-1}{n}} \tag{3}$$

where the Glen's flow law exponent n is typically taken as $n = 3$. To determine the local ice rigidity $B(T)$ for the ice temperature $T(t)$, we apply the temperature relationship provided by [55, p. 97].

The system Eqs. 1 and 2 is subject to homogeneous zero-Neumann boundary conditions on the surfaces $\Gamma_{\mathrm{air}}(t)$ exposed to air, i.e.,

$$\dot{\varepsilon}_{\mathrm{HO},1} \cdot \mathbf{n} = 0, \qquad \dot{\varepsilon}_{\mathrm{HO},2} \cdot \mathbf{n} = 0$$

for $\mathbf{x} = (x, y, z) \in \Gamma_{\mathrm{air}}(t)$, and where $\mathbf{n} = (n_x, n_y, n_z)$ is the outward pointing unit normal of $\Omega(t)$. The stress balance with the atmospheric pressure has been neglected because that pressure is negligible in the setting we consider. In contrast, at the ice-water interface Γ_w the water pressure gets applied through

$$2\mu\dot{\varepsilon}_{\mathrm{HO},1} \cdot \mathbf{n} = f_w \mathbf{n}_x, \quad 2\mu\dot{\varepsilon}_{\mathrm{HO},2} \cdot \mathbf{n} = f_w \mathbf{n}_y$$

for $\mathbf{x} = (x, y, z) \in \Gamma_w$, and with

$$f_w(\mathbf{x}) = \rho_{\mathrm{ice}} g (s(x, y; t) - z) + \rho_w g \min\{z, 0\}$$

---

[2] Here, the nomenclature "higher order" is relative to even more simplified ice sheet models such as the shallow-shelf and shallow-ice models; see e.g., Section 2.3, and [54].

where $\rho_{\mathrm{w}} = 1023\,\mathrm{km/m^3}$ is the ocean water density. At the basal boundary Γ_{b}, we prescribe sliding boundary conditions

$$2\mu\dot{\varepsilon}_{\mathrm{HO,1}} \cdot \mathbf{n} = -\alpha^2 N\, v_x, \quad 2\mu\dot{\varepsilon}_{\mathrm{HO,2}} \cdot \mathbf{n} = -\alpha^2 N\, v_y$$

for $\mathbf{x} = (x, y, z) \in \Gamma_{\mathrm{b}}$, and where α is the basal friction field, and $N = \rho_{\mathrm{ice}}gh + \rho_{\mathrm{water}}gb$, with b denoting the altitude of the basal topography Γ_{b}, is the effective pressure, following Budd's friction law [56]. We make the simplifying assumption that all ice is grounded such that α is constant in time. Note that floating ice is not essential to our considered scenario as this assumption only affects 0.22 % of the domain.

Since the basal friction field α is only indirectly observable, we model it as a random variable, see Section 2.2. However, in our reference parameterization used for the spin-up and control runs, we choose α_{ref} by minimizing the misfit between observed and simulated surface velocity data, described by the cost function

$$\begin{aligned}J(\alpha; \mathbf{v}_{\mathrm{obs}}) &= 180 \int_{\Gamma_{\mathrm{s}}} \|\mathbf{v}(\alpha) - \mathbf{v}_{\mathrm{obs}}\|^2 dS \\ &+ 0.6 \int_{\Gamma_{\mathrm{s}}} \left(\log\left(\frac{\|\mathbf{v}(\alpha)\| + \varepsilon}{\|\mathbf{v}_{\mathrm{obs}}\| + \varepsilon} \right) \right)^2 dS \\ &+ 8 \times 10^{-6} \int_{\Gamma_b} \|\nabla\alpha\|^2 dS.\end{aligned} \quad (4)$$

Here, $\mathbf{v}(\alpha)$ is the HO solution at $t = t_0$ for a given basal friction field α, $\Gamma_{\mathrm{s}} = \Gamma_{\mathrm{s}}(t_0)$ is the domain's initial surface boundary interpolated from [49] (version 2), and $\mathbf{v}_{\mathrm{obs}}$ is the observed 1995-2015 averaged surface velocity field from [50]. The variable $\varepsilon = 2.22 \times 10^{-16}\,\mathrm{m/s}$ is added to avoid division by zero. The weights for the first two cost functions are normalization parameters chosen to balance the influence of the two terms; the weight for the regularization was chosen via an L-curve analysis.

The HO system Eqs. 1 and 2 is solved for the 3-dimensional velocity fields v_x and v_y. After they have been determined, the vertical velocity field v_z can be constructed by from the constraint $\mathrm{div}\,\mathbf{v} = 0$, which is required by the conservation of mass, assuming that ice is an incompressible material. Note that although both the FS and HO models are posed on the same three-dimensional domain $\Omega(t)$, the HO model involves solving for only two unknowns (v_x and v_y) instead of four (v_x, v_y, v_z, and pressure p). Moreover, without the additional stability requirements posed by the pressure term [57, 58], the HO system is simpler to solve numerically. Hence, the computational costs associated with the HO model are smaller than that for the FS model. For a detailed introduction and comparison of both equations, we refer to [54].

2.1.3 Thermal model

The thermal model is derived from the balance equation of internal energy and Fourier's law of heat transfer under the assumption of constant heat conductivity $\kappa_{\mathrm{ice}} = 2.4\,\mathrm{WK/m}$, constant heat capacity $c_{\mathrm{ice}} = 2093\,\mathrm{JK/kg}$, and constant ice density $\rho_{\mathrm{ice}} = 917\,\mathrm{kg/m^3}$. Given the velocity vector $\mathbf{v}(t) = (v_x(t), v_y(t), v_z(t))^\top$ at time $t_0 \le t \le t_f$, the thermal model is described by the PDE

$$\rho_{\mathrm{ice}}c_{\mathrm{ice}}\left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T \right) = -\kappa_{\mathrm{ice}}c_{\mathrm{ice}}\Delta T + \mathrm{tr}(\sigma \cdot \dot{\varepsilon}) \quad (5)$$

for $\mathbf{x} \in \Omega(t)$ and $t_0 < t \le t_f$, and where σ is the Cauchy stress tensor, $\dot{\varepsilon}$ the strain rate tensor, and tr denotes the trace operator. The PDE Eq. 5 accounts for the transfer of energy (due to dissipation via the Fourier heat law), convection (due to ice movement), friction and viscous heating (due to ice deformation and sliding).

At the surfaces Γ_{air} of the ice sheet exposed to air, i.e., the upper surface and the part of the lateral surface above sea level, we locally prescribe the mean annual air temperature $T_{\mathrm{air}}(\mathbf{x})$ predicted by CNRM-CM6-1 [51] as Dirichlet boundary condition. As the geometry $\Omega(t)$ changes, T_{air} is adjusted to account for changes in altitude as described in [43]. To preserve continuity of T between years, at the beginning of each year we linearly interpolate between the old and new air temperature over a duration of 9.125 days. At the basal boundary Γ_{b} of the ice sheet touching the bedrock, we prescribe the Neumann boundary condition

$$-\kappa_{\mathrm{ice}}c_{\mathrm{ice}}\nabla T \cdot \mathbf{n} = q_{\mathrm{geo}}(x, y)$$

for $\mathbf{x} = (x, y, z) \in \Gamma_b$ and $t_0 \le t \le t_f$, and where q_{geo} denotes the geothermal heat flux. Since the latter is only sparsely observable through ice core measurements, we model it as a random variable (see Section 2.2); for our reference parameterization, we choose q_{geo} as the average of the fields in [20] and [21].

2.1.4 Ice thickness equation

The mass and geometry of the ice sheet are determined by the ice thickness h, whose evolution is governed by the PDE

$$\frac{\partial h}{\partial t} = -\nabla \cdot (h\bar{\mathbf{v}}) + \dot{M}_{\mathrm{s}} - \dot{M}_{\mathrm{b}} \quad (6)$$

for $(x, y) \in \Omega_{\mathrm{2D}}$, and $t_0 \le t \le t_f$. The PDE Eq. 6 is connected to the velocity $\mathbf{v}$ of the ice sheet through the depth-averaged horizontal velocity vector $\bar{\mathbf{v}}$. In addition, Eq. 6 depends the imposed climate forcing and the ice temperature T through the basal melt rate $\dot{M}_{\mathrm{b}}$ and the surface mass balance $\dot{M}_{\mathrm{s}}$. The basal melt $\dot{M}_{\mathrm{b}}$ is caused by ice-movement induced
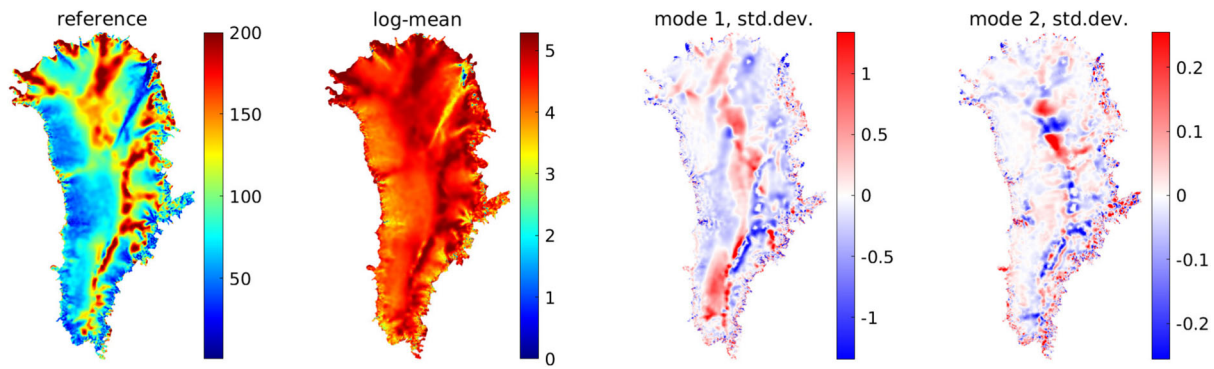
Fig. 2 Far left: Reference basal friction field at the parametric mean (measured in $(s/m)^{1/2}$); Center left to right: mean and first two modes (scaled by their standard deviation) of the log-normal distribution (measured in $\ln(s/m)/2$)

friction and geothermal heating; it is updated throughout the computations using the geothermal heat flux q_{geo}, the ice velocity $\mathbf{v}$, and the ice temperature T. The surface mass balance $\dot{M}_s$ is an input function that is positive wherever accumulation occurs, e.g., from snowfall. For our purposes here, we impose the surface mass balance predicted by the CNRM-CM6-1 climate model [51] with the altitude dependent adjustment described in [43].

Our OoI is the ice mass loss in the modelled domain of Greenland compared to a control run over the 2015–2050 time period. The purpose of the control run is to negate model drift, and is computed using the CNRM-CM6-1 [51] SSP1-2.6 1960–1989 reference forcing, i.e., the same forcing as used during the initialization at equilibrium. The ice mass is computed by integrating the ice thickness h over Ω_{2D}, and multiplying with the ice density $\rho_{ice} = 917\,kg/m^3$. Consequently, ice mass can be lost not only through melting, but also through ice discharge as ice is advected outside of the modelled domain along the periphery of the ice sheet.

2.2 Model uncertainties

Ice sheet simulations are subject to a number of uncertainties because many of the physical properties cannot be directly observed through measurements and their indirect inference from data leads to model parameter uncertainties. Moreover, projections are further uncertain because future forcing conditions are unknown. Reports to inform policy making treat this issue by defining socioeconomic pathway scenarios that climate models can simulate. Ice sheet simulations use these climate model projections to define the future forcing, while parameter uncertainties are treated by comparing projections of models from different expert groups with each other for the same scenarios (e.g., ISMIP6 [1, 3]). However, these approaches do not result in formal forward propagation of parameter uncertainties to determine quantitative effects on projections; rather, the ensemble approach computes a mean projection by averaging across the projections of different

models (typically a small number). We address that gap here through the introduction of UQ methods that formally characterize parameter uncertainties and compute the impact of those uncertainties on projections.

We first characterize uncertainty in the basal friction field α, which governs the sliding of the ice at the bedrock and thus has a strong influence on the ice velocity $\mathbf{v}$. However, the basal friction field can only be indirectly observed through velocity data at the surface, and is thus uncertain. Estimates of the basal friction field are commonly obtained by solving an inverse problem such as Eq. 4. We model the uncertainty in the basal friction field induced by the choice for $\mathbf{v}_{obs}$: Using the same weights as in Eq. 4, we compute basal friction fields α_i as the minimizers of the cost function $J(\alpha; \mathbf{v}_i^{obs})$ with $\mathbf{v}_i^{obs}$ being the annual average surface velocity data for the year i ($i = 2015, \ldots, 2021$) from [59]. In these inverse problems we use the geometry of our 2015 initial condition, and start the minimization from α_{ref} as initial guess, such that the data $\mathbf{v}_i^{obs}$ can loosely be interpreted as measurements obtained for a basal friction α drawn from a log-normal distribution

$$\alpha = \exp\left(\ln(\alpha_{ref}) + \sum_{i=1}^{7} X_{\alpha,i}\phi_i\right), \qquad (7)$$

with random variable $X_\alpha \in \mathbb{R}^7$, $X_\alpha \sim \mathcal{N}(0, \Sigma_\alpha)$, and modes $\phi_1, \ldots, \phi_7$ defined on the basal surface Γ_b. Note that by modeling α through a log-normal distribution, we are guaranteed that α is positive everywhere. To determine $\Sigma_\alpha \in \mathbb{R}^{7\times7}$ and $\phi_1, \ldots, \phi_7$, we form the covariance matrix between $\ln(\alpha_i) - \ln(\alpha_{ref})$, $i = 2015, \ldots, 2021$; we then choose ϕ_i as the i-th principal component and Σ_α as the associated diagonal covariance matrix (c.f., [60]). Figure 2 shows the log-mean and the two dominant principal components scaled by their respective standard deviation. The first mode identifies large areas with strong uncertainty, in the second the uncertainties are smaller and slightly more local, but still affect large areas of the domain. The remaining modes (not

Fig. 3 Geothermal heat flux fields from [21] (left) and [20] (right) measured in W/m^2
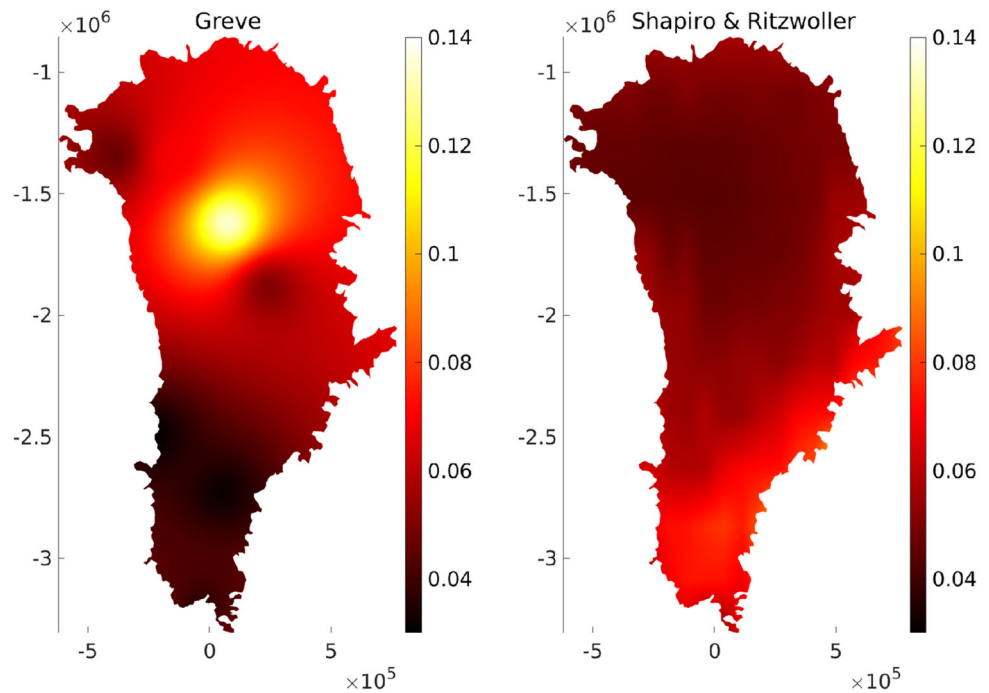
shown) follow this trend in reflecting increasingly more local fluctuations in the basal friction field α.

Next, we characterize the uncertainty in the geothermal heat flux, which describes how much heat moves outward from the interior of the earth and is needed for the basal boundary condition of the thermal model. Since the basal boundary is underneath the ice, the geothermal heat flux is difficult to observe and therefore has large uncertainty [10]. Sparse measurements are obtained through ice cores. There are many approaches using these sparse geothermal heat flux measurements to infer the complete geothermal heat flux field, including machine learning [61, 62], topographic corrections [63], magnetic fields [64, 65], and others [10, 66]. Different approaches lead to flux fields that are vastly different (see [61], Figure 13, or [10]), and there generally appears to be disagreement on how to best quantitatively describe the flux field. For instance, in ISMIP6, 13 groups used the heat flux from [20], where seismic data is used to extrapolate heat flow measurements, five groups used [21], where ice-core measurements are used for local adjustments of the global heat flux map [67], and the remaining groups used different fluxes ([1], Table A1). Yet, as shown in Fig. 3, the geothermal heat flux fields from [21] and [20] are considerably different.

We build an uncertainty model for the geothermal heat flux using the fields that were predominantly chosen in the ISMIP6 models: the field q_{geo}^{G} with superscript "G" for "Greve" from [21], and the field q_{geo}^{SR} with superscript "SR" for "Shapiro-Ritzwoller" from [20], both depicted in Fig. 3. Following the example at the NASA Sea Level Change

Portal,[3] we model the uncertainty of the geothermal heat flux through

$$q_{\text{geo}} = X_{\text{geo}} q_{geo}^{\text{G}} + (1 - X_{\text{geo}}) q_{geo}^{\text{SR}}$$

where $X_{\text{geo}} \sim \mathcal{U}(0, 1)$ is a uniformly distributed random variable taking values between 0 and 1.

We now show quantitatively the effects of these uncertainties in basal friction field and geothermal heat flux on predicted ice mass loss, treating X_{geo} and X_α as independent variables. This choice neglects how the inversion for the basal friction field would absorb increased/decreased ice velocities caused by stronger/weaker geothermal heating (see [42]), and therefore leads to higher/lower local ice velocity. To additionally model this effect, the uncertainties in q_{geo} need already be included when generating the training data for the model Eq. 7 such that $(X_{\text{geo}}, X_\alpha)$ can be modelled as a correlated random vector. Here, we used the mean $\mathbb{E}(q_{\text{geo}})$ when building the model Eq. 7.

Figure 4 shows the high-fidelity projections of the 2015–2050 ice mass loss in Greenland for 32 samples of $(X_{\text{geo}}, X_\alpha)$ chosen via Latin-Hypercube sampling. The colors are ordered according to the magnitude of the projections for the year 2050. It can be seen that the modelled uncertainty in the geothermal heat flux and the basal friction field lead to large variations in the predicted ice mass change. These variations grow increasingly as time progresses. Comparing the mean

---

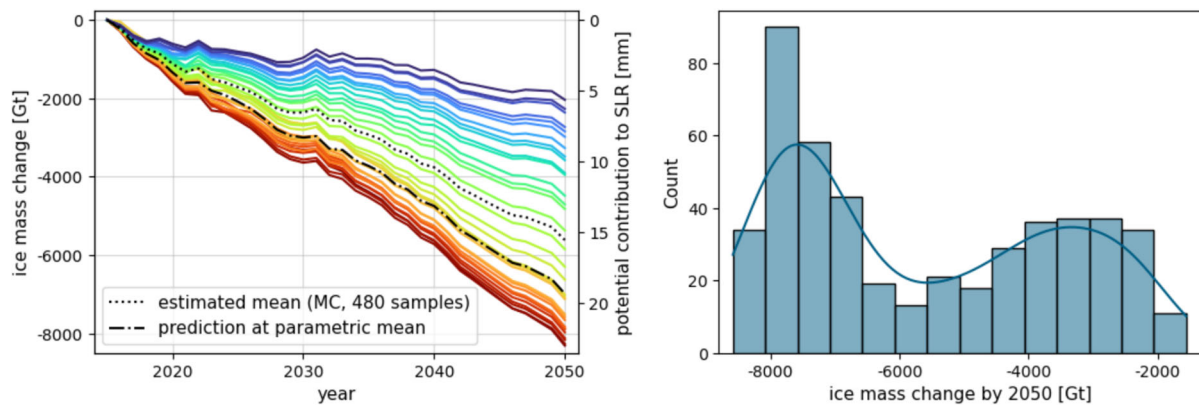[3] https://sealevel.nasa.gov/, accessed Aug. 2024

Fig. 4 Left: High-fidelity ice mass change projections for 32 samples (colored lines) of the geothermal heat flux and basal friction fields; right: Histogram and kernel density estimate of the high-fidelity ice mass change projections for the year 2050 computed over 480 samples of the geothermal heat flux and basal friction fields

ice mass change over 480 projections of i.i.d. samples (dotted line) with the predicted ice mass change at the mean parameter $\mathbb{E}[(X_{\text{geo}}, X_\alpha)] = (0.5, 0, \ldots, 0)$ (dashed-dotted line), we see a discrepancy of 1310 Gt by the year 2050. This discrepancy is perhaps not surprising considering that our high-fidelity model is highly nonlinear with respect to the uncertain parameters. Moreover, the histogram of 480 samples plotted in Fig. 4 illustrates that the model nonlinearity leads to a non-Gaussian distribution in ice mass change outputs, likely due to the presence of tipping points in ice retreat. This highlights the critical importance of conducting a full quantitative assessment of uncertainty.

We remark that the multifidelity UQ framework introduced in Section 3 is agnostic to the specific choice of modeled uncertainties; thus, other characterizations of the uncertainty in the geothermal heat flux and basal friction fields are permitted, e.g., through Gauss Markov random fields (see [68]), posterior distributions (see [17, 18]), or learned distributions (see [19, 62]). Similarly, other uncertainties may be considered, such as the basal topography (see [22, 69]), or hyperparameters associated with data-fit surrogate models.

2.3 Surrogate modeling

The multifidelity UQ framework relies on the availability of surrogate lower-fidelity/lower-cost models. In this section we provide brief descriptions of some of the types of surrogate models readily available in many ice sheet codes.

A straightforward means for obtaining lower-fidelity/lower-cost surrogates is to use coarser grids and/or larger time steps when discretizing the continuous models. This approach was first used for MLMC estimation in [39] and further developed in [40]. Because a coarsened computational model has fewer degrees of freedom, its solution can be expected to

require less computational effort. Note that when coarsening a discretization, it is important to still adhere to stability conditions (e.g., the CFL condition for time step size).

Another common form of surrogate model in computational glaciology is a simplified-physics model, especially for the stress balance equations. The development of simplified-physics approximations is primarily motivated by the high computational cost of the FS equations, which necessitates surrogate modeling for many applications. Consequently, there exists a vast variety of approximated physics models, primarily the HO equations (if the high-fidelity model is the Full Stokes system), SSA equations [70], the Shallow Ice Approximation [71], mono-layer models [72, 73], and hybrid models that impose different equations on different parts of the domain [74, 75]. The prevalence of approximated physics models in the ISMIP6 studies [1–4] and throughout the ice sheet literature shows the existing trust in these methods. It can thus be expected that approximated physics models have strong correlations with the high-fidelity model, and are already implemented in many ice sheet codes.

In the SSA model, vertical shear is neglected such that only the depth-averaged velocity field $\bar{\mathbf{v}}$ on the 2D domain Ω_{2D} is solved for. The governing equations become:

$$\nabla \cdot (2\bar{\mu} h \dot{\varepsilon}_{\text{SSA},1}) - \alpha^2 N \bar{\mathbf{v}}_x = \rho g h \frac{\partial s}{\partial x} \tag{8}$$

$$\nabla \cdot (2\bar{\mu} h \dot{\varepsilon}_{\text{SSA},2}) - \alpha^2 N \bar{\mathbf{v}}_y = \rho g h \frac{\partial s}{\partial y} \tag{9}$$

for $\mathbf{x} = (x, y) \in \Omega_{\text{2D}}$ and $t_0 < t \leq t_f$, and where the mean ice viscosity $\bar{\mu}$ is computed from Glen's flow law Eq. 3 with strain rate

$$\dot{\varepsilon}_{\text{SSA},1} = \begin{pmatrix} 2\dfrac{\partial \bar{\mathbf{v}}_x}{\partial x} + \dfrac{\partial \bar{\mathbf{v}}_y}{\partial y} \\ \dfrac{1}{2}\dfrac{\partial \bar{\mathbf{v}}_x}{\partial y} + \dfrac{1}{2}\dfrac{\partial \bar{\mathbf{v}}_y}{\partial x} \end{pmatrix}$$

Table 1 Discretization and runtime of our ISSM models

mesh name	resolution min.	max.	DoFs/variable 2D	3D	cost [CPUh] SSA	HO
fine	100 m	15 km	20,455	102,275	0.455	64.888
medium	100 m	30 km	6,554	32,770	0.105	22.889
coarse	100 m	50 km	3,600	18,000	0.059	4.689

and

$$\dot{\varepsilon}_{\text{SSA},2} = \begin{pmatrix} \dfrac{1}{2}\dfrac{\partial \bar{\mathbf{v}}_x}{\partial y} + \dfrac{1}{2}\dfrac{\partial \bar{\mathbf{v}}_y}{\partial x} \\ \dfrac{\partial \bar{\mathbf{v}}_x}{\partial x} + 2\dfrac{\partial \bar{\mathbf{v}}_y}{\partial y} \end{pmatrix}.$$

The equations Eqs. 8 and 9 are solved with boundary conditions

$$\dot{\varepsilon}_{\text{SSA},1} \cdot \mathbf{n}_{\text{2D}} = 0 \qquad \dot{\varepsilon}_{\text{SSA},2} \cdot \mathbf{n}_{\text{2D}} = 0 \qquad \text{on } \partial\Omega_{\text{2D}},$$

on the interface of Ω_{2D} with air, and with

$$2\bar{\mu}h\dot{\varepsilon}_{\text{SSA},1} \cdot \mathbf{n}_{\text{2D}} = f_{\text{w,SSA}}n_x$$
$$2\bar{\mu}h\dot{\varepsilon}_{\text{SSA},2} \cdot \mathbf{n}_{\text{2D}} = f_{\text{w,SSA}}n_x$$

on the ice-water interface. Here, $\mathbf{n}_{\text{2D}} = (n_x, n_y)^\top$ denotes the outward pointing unit normal of the domain Ω_{2D}. The water pressure is applied through the Neumann flux function

$$f_{\text{w,SSA}} = \frac{1}{2}(\rho_{\text{ice}}gh^2 - \rho_{\text{w}}gb^2).$$

Since Eqs. 8 and 9 are solved for two variables defined over a 2D surface instead of two variables defined over a 3D domain, SSA models offer vast computational savings compared to HO models: For the meshes considered in this study, predictions using an SSA model were between 142 and 1,099 times faster than their HO counterparts, see Table 1.

Another tool for building surrogate models is training on available data of the high-fidelity model or its approximations. Most prevalent are emulators of the model's OoIs, for instance through Gaussian process regression; this approach was taken in the forward UQ studies [32–36]. Other forms of data-driven models include physics-based learning keeping the connection to governing equations [76] and more general machine learning approaches [77, 78]. Throughout these approaches, training-based surrogate models hinge on the availability of training data, require some amount of expertise from the modeller, and can have large generalisation errors outside the training regime, especially if the uncertain parameters are high-dimensional. On the other hand, they can achieve significantly larger speed-ups than physical approximations.

In general, for multifidelity UQ any approximation of the high-fidelity input/output map can be considered as a surrogate model as long as we can characterize its cost and its correlation with the high-fidelity model. However, despite the computational gains offered by surrogate models, the high-fidelity model should generally not be replaced when estimating its OoI as the model bias introduced by the surrogate can be significant. For example, Fig. 5 illustrates the effects of both mesh coarsening and physical approximation via SSA on ice mass change predictions for 32 samples of
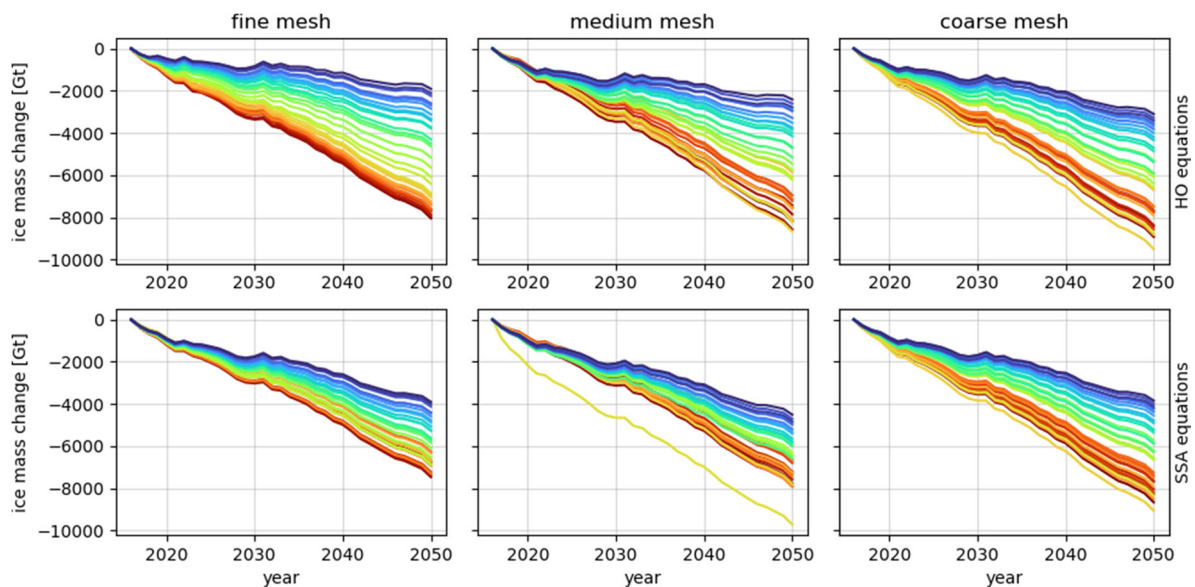

Fig. 5 Effect of mesh coarsening (left to right) and physical approximation (top to bottom) on the predicted 2015-2050 ice mass change for 32 samples of the geothermal heat flux and basal friction fields relative to control run on the same mesh with the same physical approximation

the geothermal heat flux and the basal friction fields chosen by Latin Hypercube sampling, with colors identifying predictions for the same samples in different plots. Considering the fine mesh with the HO equations to be the highest fidelity model, for the five surrogate models in Fig. 5, the mean 2050 ice mass change differs by 407 Gt to 2286 Gt (13.7 % to 77.0 %) from that estimated from the high-fidelity samples. To avoid such an introduction of model bias, multifidelity estimators do *not* replace the high-fidelity model with surrogates, but incorporate multiple models into their structure in a way that guarantees an *unbiased* estimate of the statistics of the high-fidelity OoI.

3 Multifidelity uncertainty quantification

This section presents multifidelity UQ methods that incur lower costs than Monte Carlo estimation by exploiting the statistical correlation between the high-fidelity model and its surrogates. We present preliminaries (Section 3.1) and then discuss three different methods with complementary strengths that make them appropriate in different settings:

- MFMC (Section 3.2): Handles a large variety of model types and computational costs, e.g., achieved by physical approximations and data-driven surrogates, including situations where there is no clear a priori model hierarchy;
- MLMC (Section 3.3): Is particularly appropriate when there is a hierarchy of similar lower-fidelity models, e.g., achieved by mesh coarsening;
- MLBLUE (Section 3.4): By construction is always at least as good as MLMC and MFMC, but implementation and handling can be more involved.

We also note that other multifidelity UQ methods, such as the Approximate Control Variate approach [79], introduce alternative algorithmic formulations that may be beneficial in some applications settings. For MFMC, MLMC, and MLBLUE we provide an algorithmic description for choosing an optimized subset of surrogate models with optimal weights and sample sizes, and the computation of the associated multifidelity estimator. The sample sizes are chosen such that the estimators' total costs remain within a given computational budget $c > 0$.[4] The differences between the methods and their applicability for UQ in ice sheet simulations are discussed in Section 4.4.

---

[4] We note that each algorithm can be re-written to work with a target accuracy instead.

3.1 Preliminaries

In the following, we denote with s_1 a high-fidelity output of interest (OoI), and by $s_1(\theta) \in \mathbb{R}$ its scalar[5] evaluation at a parameter sample $\theta \sim \nu$. In addition, we have at hand $L - 1 \in \mathbb{N}$ surrogate OoIs $s_2, \ldots, s_L$. Evaluating an output s_i for a parameter $\theta \sim \nu$ incurs the computational cost $c_i > 0$, e.g., measured in CPU-hours. We assume without loss of generality that the models are ordered by cost: $c_1 \geq c_2 \geq \cdots \geq c_L$. In our setting here, s_1 is the 2050 ice mass change (in Gt) computed by the high-fidelity model described in Section 2.1, and $\nu = \mathcal{U}(0, 1) \times \mathcal{N}(\mathbf{0}, \Sigma)$ is the probability distribution of the uncertainties in the geothermal heat flux and basal friction field described in Section 2.2.

Furthermore, for $i, j = 1, \ldots, L$, we define the model covariance matrix $\Sigma \in \mathbb{R}^{L \times L}$:

$$\Sigma_{i,j} := \mathbb{C}\text{ov}(s_i, s_j)$$
$$= \mathbb{E}_\nu[(s_i - \mathbb{E}_\nu[s_i])(s_j - \mathbb{E}_\nu[s_j])].$$

If this matrix is not available from theoretical estimates, it can be approximated using model convergence rates or from n_{pilot} pilot samples:

$$\Sigma_{i,j} \approx \frac{1}{n_{\text{pilot}} - 1} \sum_{k=1}^{n_{\text{pilot}}} \left(s_i(\theta_k) - \bar{s}_i^{\text{pilot}} \right) \left(s_j(\theta_k) - \bar{s}_j^{\text{pilot}} \right)$$

with $\bar{s}_i^{\text{pilot}} := n_{\text{pilot}}^{-1} \sum_{k=1}^{n_{\text{pilot}}} s_i(\theta_k)$. We expect that for most ice sheet applications, sufficient data is available from validation and verification procedures, as well as from historic, control, and spin-up runs. However, should additional sampling indeed be necessary, the model evaluations may be re-used when evaluating the multifidelity estimators. We therefore do not count the estimation of Σ towards computational costs.

3.2 Multifidelity Monte Carlo

The MFMC method employs surrogate models of arbitrary structure [37, 38]. In the following we focus on the MFMC method from [38] with the adaptation for restrictive computational budgets from [81] and an integrated model selection step. The method is summarized in Algorithm 1. We note that MFMC has been extended to estimation of sensitivities and covariances [82], is compatible with goal-oriented training of surrogate models [83, 84], and has been applied to large-scale applications, in particular in climate modeling [85].

---

[5] We restrict our exposition to scalar OoIs solely for the purpose of simpler notation; each of the presented multifidelity UQ methods can be posed similarly for vector-valued quantities. For MLBLUE, it is even possible to consider surrogate models that have additional or less OoIs than the high-fidelity model (c.f., [80]).

The MFMC estimator is based on writing the target statistic $\mathbb{E}[s_1]$ in the form

$$\mathbb{E}[s_1] = \mathbb{E}[s_1] + \alpha_2 \mathbb{E}[s_2] - \alpha_2 \mathbb{E}[s_2]$$
$$= \mathbb{E}[s_1] + \sum_{j=2}^{L} \alpha_j \big(\mathbb{E}[s_j] - \mathbb{E}[s_j]\big)$$

for arbitrary weights $\alpha_2, \ldots, \alpha_L$ (optimal values are defined below). Each expectation $\mathbb{E}[s_j]$ is then approximated using the Monte Carlo method with shared samples such that the MFMC estimator $\bar{s}_{\mathrm{MFMC}}$ has the form

$$\bar{s}_{\mathrm{MFMC}} := \frac{1}{n_1} \sum_{i=1}^{n_1} s_1(\theta_i)$$
$$+ \sum_{j=2}^{L} \alpha_j \left(\frac{1}{n_j} \sum_{i=1}^{n_j} s_j(\theta_i) - \frac{1}{n_{j-1}} \sum_{i=1}^{n_{j-1}} s_j(\theta_i) \right) \tag{10}$$

where, for each $1 \leq j \leq L$, $n_j \in \mathbb{N}$ is the number of samples for which the OoI s_j needs to be evaluated using model j, and $\theta_1, \ldots, \theta_{n_L} \sim \nu$ are i.i.d. parameter samples shared between the models. By imposing the order $1 \leq n_1 \leq n_2 \leq \cdots \leq n_L$, the cheapest model s_L is sampled the most while the expensive high-fidelity model, s_1, is sampled the least. By construction, the MFMC estimator $\bar{s}_{\mathrm{MFMC}}$ is unbiased: $\mathbb{E}[\bar{s}_{\mathrm{MFMC}}] = \mathbb{E}[s_1]$. Note that, in constrast to MLMC, the samples within the individual sums in Eq. 10 are shared. Exploiting the consequent correlation between the sums, the mean squared error (MSE) of the MFMC estimator Eq. 10 is then given by

$$\mathrm{MSE}(\bar{s}_{\mathrm{MFMC}})$$
$$= \mathbb{E}[(\bar{s}_{\mathrm{MFMC}} - \mathbb{E}[s_1])^2]$$
$$= \frac{\sigma_1^2}{n_1} + \sum_{j=2}^{L} \left(\frac{1}{n_{j-1}} - \frac{1}{n_j} \right) (\alpha_j^2 \sigma_j^2 - 2\alpha_j \rho_j \sigma_1 \sigma_j) \tag{11}$$

where, for $1 \leq j \leq L$, $\sigma_j = \sqrt{\Sigma_{j,j}}$ is the standard deviation for model s_j, and $\rho_j := \Sigma_{1,j}/(\sigma_1 \sigma_j)$ is the correlation of model j with the high-fidelity model s_1. Since the MFMC error is unbiased, the MSE is the expected squared error between the MFMC estimator and the target statistic $\mathbb{E}[s_1]$. From the formula Eq. 11 we can immediately deduce that $\alpha_j = \rho_j \sigma_1/\sigma_j = \Sigma_{1,j}/\Sigma_{j,j}$ for $j = 2, \ldots, L$ are optimal for achieving the smallest MSE in Eq. 10.

To obtain optimal sample sizes $1 \leq n_1 \leq n_2 \leq \cdots \leq n_L$ for a given total computational budget $c \geq c_1$, we solve the relaxed optimization problem

$$\min_{n_1, \ldots, n_L \in \mathbb{R}} \mathrm{MSE}(\bar{s}_{\mathrm{MFMC}})$$
$$\text{s.t.} \quad \begin{cases} 0 < n_1 \leq n_2 \leq \cdots \leq n_L \\ \sum_{j=1}^{L} n_j c_j \leq c. \end{cases} \tag{12}$$

The first constraints on the sample sizes ensure that at least as many samples are taken of the OoI s_j as are taken for the more costly OoI s_{j-1}, with the constraint $0 < n_1$ ensuring that all sample sizes are positive. The last constraint on the model evaluation costs ensures that the given computational budget c is not exceeded.

It has been shown in [38] that if $1 = |\rho_1| > |\rho_2| > \cdots > |\rho_L|$ and

$$\frac{c_{j-1}}{c_j} > \frac{\rho_{j-1}^2 - \rho_j^2}{\rho_j^2 - \rho_{j+1}^2} \tag{13}$$

hold, then the global minimum of Eq. 12 is obtained by

$$n_1 = c \left(\sum_{j=1}^{L} c_j r_j \right)^{-1}, \quad n_j = r_j n_1 \tag{14}$$

for $j = 2, \ldots, L$, where we are using the auxiliary variable

$$r_j := \sqrt{\frac{c_1(\rho_j^2 - \rho_{j+1}^2)}{c_j(\rho_1^2 - \rho_2^2)}}. \tag{15}$$

Condition Eq. 13 describes how much faster any lower-fidelity model s_j must be, compared to the more expensive model s_{j-1}, in order to make up for its worse correlation $|\rho_j| < |\rho_{j-1}|$.[6] After computing $n_1, \ldots, n_L \in \mathbb{R}$ using the closed-form solution Eq. 14, they need to be rounded to integers larger or equal than one. If possible, sample sizes are rounded down to ensure the computational budget is not exceeded. Any remaining budget can be redistributed to further decrease the MFMC estimator's variance, e.g., using the optimized rounding strategy in [81].

Algorithm 1 summarizes the computation of the sample sizes $n_1, \ldots, n_L$ and weights $\alpha_1, \ldots, \alpha_L$ for given models $s_1, \ldots, s_L$. Since the smallest MSE of the MFMC estimator may be achieved with a subset of the available surrogate models, Algorithm 1 should always be placed within an outer loop over all subsets of the available surrogate models to determine when the MSE is the smallest. In addition, the obtained value for MSE should be compared with the MSE of Monte Carlo sampling, i.e., the value $\sigma_1^2/\lfloor c/c_1 \rfloor$, to ensure

---

[6] Note that if there exists a model $j + 1$ such that $\rho_{j+1} \geq \rho_j$, then it is clearly superior to model s_j both in cost and correlation, and model s_j should be removed.

Algorithm 1 Multifidelity Monte Carlo

Input: high-fidelity model s_1, surrogate models $s_2, \ldots, s_L$, model evaluation costs $c_1 \geq \cdots \geq c_L$, model correlations $\rho_1, \ldots, \rho_L$, model variances $\sigma_1^2, \ldots, \sigma_L^2$, computational budget c
Output: sample sizes $1 \leq n_1 \leq \cdots \leq n_L$, optimal weights $\alpha_1, \ldots, \alpha_L$, estimator MSE $\text{MSE} = \mathbb{E}[(\bar{s}_{\text{MFMC}} - \mathbb{E}[s_1])^2]$

1. Assert that $1 = |\rho_1| > \cdots > |\rho_L|$. If not, then any model s_j with $\rho_j \leq \rho_{j+1}$ needs to be removed; return $\text{MSE} = \infty$.
2. Define the auxiliary variable $\rho_{L+1} = 0$.
3. Assert that the budget the budget is large enough, i.e., $\sum_{j=1}^{L} c_j \leq c$, and that Eq. 13
   holds for $j = 2, \ldots, L$. If not, then the provided combination of surrogate models is not suitable for MFMC; return $\text{MSE} = \infty$.
4. Compute the ratios r_j for $j = 1, \ldots, L$ using Eq. 15, and use them to compute $n_1, \ldots, n_L \in \mathbb{R}$ with Eq. 14.
5. For $j = 1, \ldots, L$, round all $n_j < 1$ up to 1, and round all $n_j \geq 1$ down to the next smallest integer.
6. Compute the optimal weights $\alpha_j = \rho_j \sigma_1 / \sigma_j$ for $j = 1, \ldots, L$
7. Compute the MSE of the MFMC estimator with Eq. 11 and save in variable MSE
8. Return $(n_1, \ldots, n_L)$, $(\alpha_1, \ldots, \alpha_L)$, MSE

that MFMC is indeed applicable. After the model subset has been selected, the MFMC estimator $\bar{s}_{\text{MFMC}}$ can be computed using Eq. 10.

3.3 Multilevel Monte Carlo

The MLMC method has its roots in [39, 40], where it was first formulated for grid-refinement based surrogate models and their convergence rates. The method has since been expanded structurally (e.g., Multi-Index Monte Carlo [86]), refined for stronger convergence (e.g. randomized and adaptive MLMC methods [87–89]), and combined with other sampling schemes (e.g., Multilevel Quasi Monte Carlo [90], Multilevel Markov Chain Monte Carlo [91]); we refer to [92] for an extensive introduction. In our exposition here we focus on the *non-geometric* MLMC method, which is posed for arbitrary surrogate models and based on model correlations instead of convergence rates, following the exposition in [92]. Algorithmic instructions are provided in Algorithm 2.

The MLMC estimator exploits that the expectation is a linear operator to expand the high-fidelity expectation $\mathbb{E}[s_1]$ in a telescoping sum of the form[7]

$$
\mathbb{E}[s_1] = \mathbb{E}[s_2] + \mathbb{E}[s_1 - s_2] \\
= \mathbb{E}[s_L] + \sum_{j=1}^{L-1} \mathbb{E}[s_j - s_{j+1}].
$$

The difference $s_j - s_{j+1}$ between models s_j and s_{j+1} is typically referred to as the j-th level, with $s_L = s_L - 0$ the

[7] Contrary to typical MLMC notation but consistent with all other parts of this paper, a larger index here indicates a model of *lower* computational cost, i.e. $c_L \leq c_{L-1} \leq \cdots \leq c_1$.

L-th level. To obtain the MLMC estimator, the expectation of each level is approximated *independently* via Monte Carlo sampling

$$
\bar{s}_{\text{MLMC}} := \frac{1}{n_L} \sum_{i=1}^{n_L} s_L(\theta_i^{(L)}) \\
+ \sum_{j=1}^{L-1} \frac{1}{n_j} \sum_{i=1}^{n_j} \left(s_j(\theta_i^{(j)}) - s_{j+1}(\theta_i^{(j)}) \right) \tag{16}
$$

where, for each level $1 \leq j \leq L$, the $\theta_i^{(j)} \sim \nu$, $1 \leq i \leq n_j$, are i.i.d. samples, and $n_j \in \mathbb{N}$ is the sample size for the j-th level Monte Carlo approximation. For implementation it is important to note that the samples are not shared beyond each level. The cost of evaluating Eq. 16 is thus $n_L c_L + \sum_{j=1}^{L-1} n_j (c_j + c_{j+1})$.

By construction, $\bar{s}_{\text{MLMC}}$ is an unbiased estimator of the high-fidelity model s_1, i.e., $\mathbb{E}[\bar{s}_{\text{MLMC}}] = \mathbb{E}[s_1]$. Its MSE is given by the formula

$$
\text{MSE}(\bar{s}_{\text{MLMC}}) = \mathbb{E}[(\bar{s}_{\text{MLMC}} - \mathbb{E}[s_1])^2] \\
= \frac{\sigma_L^2}{n_L} + \sum_{j=1}^{L-1} \frac{\overline{\sigma}_j^2}{n_j} \tag{17}
$$

where we are using the auxiliary variable

$$
\overline{\sigma}_j^2 := \mathbb{V}\text{ar}(s_j - s_{j+1}) = \Sigma_{j,j} - 2\Sigma_{j,j+1} + \Sigma_{j+1,j+1}
$$

to abbreviate the variance of each level. The underlying premise for the MLMC estimator to be effective is the observation that if models s_j and s_{j+1} yield similar OoIs, then the level variance $\overline{\sigma}_j^2$ is small; consequently only a small sample size n_j is required to balance out the contribution of level j to the MSE of the MLMC estimator in Eq. 17. For level L, the surrogate s_L is sampled alone without an additional model to decrease that level's variance, but since s_L is the cheapest model it can be expected that the associated cost $c_L n_L$ remains reasonable, even if n_L is large.

To balance and minimize the contributions of each level to the MLMC estimator's MSE, it suggested in [92] to choose

$$
n_L = \tau \sqrt{\frac{\sigma_L^2}{c_L}}, \quad n_j = \tau \sqrt{\frac{\overline{\sigma}_j^2}{c_j + c_{j+1}}} \tag{18}
$$

for $j = 1, \ldots, L-1$. The scaling factor $\tau > 0$ can be chosen to adhere to budget constraints: For $\bar{s}_{\text{MLMC}}$ to have a cost of c, i.e.,

$$
n_L c_L + \sum_{j=1}^{L-1} n_j (c_j + c_{j+1}) = c,
$$

is equivalent to choosing

$$\tau = c \left(\sqrt{\sigma_L^2 c_L} + \sum_{j=1}^{L-1} \sqrt{\overline{\sigma}_j^2 (c_j + c_{j+1})} \right)^{-1} . \qquad (19)$$

To obtain integers, sample sizes $n_j \leq 1$ are rounded up to 1, and down otherwise. Note that the rounding may cause the computational budget to be violated[8]; the budget constraint therefore needs to be checked before accepting a combination of surrogate models (see Step 2.4 in Algorithm 2).

Algorithm 2 Multilevel Monte Carlo

Input: high-fidelity model s_1, surrogate models $s_2, \ldots, s_L$, model evaluation costs $c_1 \geq \cdots \geq c_L$, level variances $\overline{\sigma}_1^2, \ldots, \overline{\sigma}_{L-1}^2$, model variance σ_L^2 of lowest fidelity model, computational budget c
Output: sample sizes $1 \leq n_1 \leq \cdots \leq n_L$, estimator MSE $\text{MSE} = \mathbb{E}[(\bar{s}_{\text{MFMC}} - \mathbb{E}[s_1])^2]$

1. Compute the scaling ratio τ using Eq. 19.
2. Compute $n_1, \ldots, n_L \in \mathbb{R}$ with Eq. 18.
3. For $k = 1, \ldots, L$, if $n_k \leq 1$ round it up to 1, otherwise round down.
4. Compute the MSE of the MLMC estimator with Eq. 17 and save in variable MSE
5. Return $(n_1, \ldots, n_L)$, MSE

The computation of the MLMC sample sizes and corresponding MSE is summarized in Algorithm 2. Similar to MFMC, the smallest MSE of the MLMC method may be realized by a subset of the available surrogate models. Algorithm 2 should hence be placed within an outer loop over all combinations of available surrogate models to determine the optimal model selection for which the returned value MSE is minimal. If the identified MSE is smaller than $\sigma_1^2 / \lfloor c/c_1 \rfloor$ — the MSE of MC sampling — then the MLMC estimator can be computed using the chosen models and identified sample sizes using Eq. 16.

3.4 Multilevel best linear unbiased estimator

The MLBLUE method was introduced in [41] and has since been extended with theoretical results in [93] and with algorithms for sample size optimization and multiple OoIs in [80]. In the following, we focus on the main concept and algorithmic steps, and refer to the cited literature for details. A full description of the MLBLUE method is provided in Algorithm 3.

To explain the structure of the MLBLUE estimator, we first define the vector of expectations of all available models

$$\hat{\mathbf{s}} = (\mathbb{E}[s_1], \ldots, \mathbb{E}[s_L])^\top \in \mathbb{R}^L .$$

Estimating the high-fidelity expectation $\mathbb{E}[s_1]$ is then equivalent to estimating $\mathbf{e}_1^\top \hat{\mathbf{s}}$ with $\mathbf{e}_1^\top = (1, 0, \ldots, 0) \in \mathbb{R}^L$. Let $\mathcal{S}_1, \ldots, \mathcal{S}_{\hat{L}}$ be an enumeration of all $\hat{L} := 2^L$ non-empty subsets of the model indices $\{1, \ldots, L\}$.

Focusing on any one index group $\mathcal{S}_i = \{1 \leq j_1 \leq \cdots \leq j_{L'} \leq L\}$ with $L' := |\mathcal{S}_i|$, we define the random vector $\mathbf{s}_i := (s_{j_1}, \ldots, s_{j_{\hat{L}}})^\top$ by stacking the models s_{j_k} with indices $j_k \in \mathcal{S}$ together. The evaluation $\mathbf{s}_i(\theta)$ at a sample $\theta \sim \nu$ can then be interpreted as a noisy observation of $\hat{\mathbf{s}}$:

$$\begin{aligned} \mathbf{s}_i(\theta) &= (s_{j_1}(\theta), \ldots, s_{j_{\hat{L}}}(\theta))^\top \\ &= \left(\mathbb{E}[s_{j_1}], \ldots, \mathbb{E}[s_{j_{\hat{L}}}] \right)^\top \\ &\quad + \left(s_{j_1}(\theta) - \mathbb{E}[s_{j_1}], \ldots, s_{j_{\hat{L}}} - \mathbb{E}[s_{j_{\hat{L}}}] \right)^\top \\ &=: \mathbf{R_i} \hat{\mathbf{s}} + \varepsilon_i(\theta) \end{aligned} \qquad (20)$$

where the matrix $\mathbf{R_i} \in \{0, 1\}^{L' \times L}$ removes all entries in $\hat{\mathbf{s}}$ that are not in the index set $\mathcal{S}_i$; it is defined through $(\mathbf{R_i})_{k,\ell} = \delta_{\ell, j_k}$ using the Kronecker delta. By construction, $\varepsilon_{\mathcal{S}} \in \mathbb{R}^{L'}$ is a random variable with mean zero, and covariance matrix $\mathbf{C}_i \in \mathbb{R}^{L' \times L'}$ defined by $(\mathbf{C}_i)_{k,\ell} := \Gamma_{j_k, j_\ell}$. Since $\mathbf{C}_i$ is a submatrix of the symmetric positive definite covariance matrix Γ it is itself symmetric positive definite and thus invertible.

For a given vector $\mathbf{n} \in \mathbb{N}_{\geq 0}^{\hat{L}}$ of sample sizes for each model index group, we evaluate the random vector $\mathbf{s}_i$ at $\mathbf{n}_i$ i.i.d. samples $\left\{ \theta_k^{(i)} \right\}_{k=1}^{\mathbf{n}_i}$. An estimate of $\hat{\mathbf{s}}$ is then obtained by solving the regression problem

$$\min_{\mathbf{s} \in \mathbb{R}^L} \sum_{i=1}^{\hat{L}} \sum_{k=1}^{n_i} \left(\mathbf{R}_i \mathbf{s} - \mathbf{s}_i(\theta_k^{(i)}) \right)^\top \mathbf{C}_i^{-1} \left(\mathbf{R}_i \mathbf{s} - \mathbf{s}_i(\theta_k^{(i)}) \right)$$

The minimum norm solution to this problem is obtained at

$$\hat{\mathbf{s}}^* = \Psi(\mathbf{n})^\dagger \mathbf{y}(\mathbf{n}) \qquad (21)$$

with the vector

$$\mathbf{y}(\mathbf{n}) := \sum_{i=1}^{\hat{L}} \mathbf{R}_i \mathbf{C}_i^{-1} \sum_{k=1}^{n_i} \mathbf{s}_i(\theta_k^{(i)}) \in \mathbb{R}^L \qquad (22)$$

and using the Moore-Penrose pseudoinverse of the likelihood matrix

$$\Psi(\mathbf{n}) := \mathbf{R}(\mathbf{n}) \mathbf{C}_\varepsilon(\mathbf{n})^{-1} \mathbf{R}(\mathbf{n})^\top \in \mathbb{R}^{L \times L} . \qquad (23)$$

In Eq. 23, the matrices $\mathbf{R}(\mathbf{n}) \in \{0, 1\}^{L \times m}$ and $\mathbf{C}_\varepsilon(\mathbf{n}) \in \mathbb{R}^{m \times m}$, with $m = m(\mathbf{n}) := \sum_{i=1}^{\hat{L}} \mathbf{n}_i |\mathcal{S}_i|$, contain $\mathbf{n}_i$ copies

of each matrix $\mathbf{R}_i$ and $\mathbf{C}_i$:

$$\begin{aligned} \mathbf{R}(\mathbf{n}) &:= (\mathbf{R}_1, \ldots, \mathbf{R}_1, R_2, \ldots, \mathbf{R}_{\hat{L}}), \\ \mathbf{C}_\varepsilon(\mathbf{n}) &:= \mathrm{diag}(\mathbf{C}_1, \ldots, \mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_{\hat{L}}). \end{aligned} \tag{24}$$

After solving Eq. 21, we obtain the MLBLUE estimator

$$\bar{s}_{\mathrm{MLBLUE}} := \mathbf{e}_1^\top \hat{\mathbf{s}}^* \approx \mathbb{E}[s_1]. \tag{25}$$

It was shown in [41, 94], that $\bar{s}_{\mathrm{MLBLUE}}$ is an unbiased estimator, i.e. $\mathbb{E}[\bar{s}_{\mathrm{MLBLUE}}] = \mathbb{E}[s_1]$, if the high-fidelity model s_1 is sampled at least once. This condition can be written equivalently as $\mathbf{n}^\top \mathbf{h} \geq 1$ where $\mathbf{h} \in \{0,1\}^{\hat{L}}$ is defined by $\mathbf{h}_i = \delta_{1 \in \mathcal{S}_i}$. In this case, the MLBLUE estimator's MSE is given by

$$\begin{aligned} \mathrm{MSE}(\bar{s}_{\mathrm{MLBLUE}}) &= \mathbb{E}[(\bar{s}_{\mathrm{MLBLUE}} - \mathbb{E}[s_1])^2] \\ &= \mathbf{e}_1^\top \Psi(\mathbf{n})^\dagger \mathbf{e}_1. \end{aligned} \tag{26}$$

The sample size $\mathbf{n}_i$ for each model vector $\mathbf{s}_i$ should be chosen to minimize the MSE of the MLBLUE estimator $\bar{s}_{\mathrm{MLBLUE}}$. It was shown in [80] that the optimal $\mathbf{n}$ for a given budget c solves the semi-definite programming problem

$$\begin{aligned} &\min_{\mathbf{n} \geq 0, t \in \mathbb{R}} \ t \\ &\text{s.t.} \ \begin{pmatrix} \Psi(\mathbf{n}) & \mathbf{e}_1 \\ \mathbf{e}_1^\top & t \end{pmatrix} \text{ is s.p.d.,} \\ &\text{and } \mathbf{n}^\top \mathbf{c} \leq c, \ \mathbf{n}^\top \mathbf{h} \geq 1, \end{aligned} \tag{27}$$

where the acronym s.p.d. stands for symmetric positive definite, and the vector $\mathbf{c} \in \mathbb{R}^{\hat{L}}$ contains the costs for evaluating the model group i, i.e., $\mathbf{c}_i = \sum_{j \in \mathcal{S}_i} c_j \geq 0$. The obtained sample size vector $\mathbf{n}$ can then be used to compute the optimal MLBLUE estimator $\bar{\mathbf{s}}_{\mathrm{MLBLUE}}$. The procedure is summarized in Algorithm 3.

Both the MFMC and the MLMC estimator can be written in the MLBLUE index group structure, albeit with predetermined weights. Recall that MLMC optimizes the sample size and MFMC optimizes both the sample size and the weights. The MLBLUE method goes further to optimize sample sizes jointly with the estimator's structure in the form of selected model groups and their weights. This means that by construction, the MSE of the MLBLUE estimator is at least as small as that of the MFMC and the MLMC method, meaning that the MLBLUE estimator is guaranteed to be at least as good as MFMC's and MLMC's, and in many cases will be better. However, since the number $\hat{L} = 2^L$ of model index groups grows exponentially fast in the number $L - 1$ of available surrogate models and because the covariance matrix Γ may be arbitrarily badly conditioned and

---

Algorithm 3 Multilevel Best Linear Unbiased Estimator

Input: high-fidelity model s_1, surrogate models $s_2, \ldots, s_L$, model evaluation costs $c_1 \geq \cdots \geq c_L$, model covariance matrix $\Sigma \in \mathbb{R}^{L \times L}$, computational budget c

Output: MLBLUE estimator $\bar{s}_{\mathrm{MLBLUE}} \approx \mathbb{E}[s_1]$, estimator variance $\mathrm{MSE} = \mathbb{V}\mathrm{ar}(\bar{s}_{\mathrm{MLBLUE}})$

1. Set $\hat{L} := 2^L$. Choose an enumeration $\mathcal{S}_1, \ldots, \mathcal{S}_{\hat{L}}$ of all $\hat{L}$ nonempty subsets of the model indices $\{1, \ldots, L\}$.
2. For $i = 1, \ldots, \hat{L}$, compute the matrices $\mathbf{C}_i$ and $\mathbf{R}_i$ (defined after Eq. 20), and the cost entry $\mathbf{c}_i = \sum_{k=1}^{|\mathcal{S}_i|} c_{j_k}$ for the model index set $\mathcal{S}_i = \{j_1 \leq \cdots \leq j_{|\mathcal{S}_i|}\}$
3. Using the definitions in Eqs. 23 and 24, solve the semi-definite programming problem Eq. 27 to obtain the sample sizes $\mathbf{n}$. We suggest using an off-the-shelf solver.
4. To compute the MLBLUE MSE, first compute $\mathbf{R}(\mathbf{n})$ and $\mathbf{C}_\varepsilon(\mathbf{n})$ with Eq. 24, and use them to compute $\Psi(\mathbf{n})$ via Eq. 23. Then evaluate $\mathrm{MSE} = \mathbf{e}_1^\top \Psi(\mathbf{n})^\dagger \mathbf{e}_1$ using Eq. 26.
5. To compute the MLBLUE estimator, for each $1 \leq i \leq \hat{L}$ with $\mathbf{n}_i \neq 0$, draw samples $\theta_k^{(i)} \sim \nu$, $1 \leq k \leq \mathbf{n}_i$, and evaluate $s_j(\theta_k^{(i)})$ for each model index $j \in \mathcal{S}_i$. Use these output evaluations in Eq. 22 to compute $\mathbf{y}(\mathbf{n})$. Then compute the minimum norm solution $\hat{\mathbf{s}}^*$ of $\Psi(\mathbf{n})\mathbf{s} = \mathbf{y}(\mathbf{n})$ using Eq. 21. Finally, compute the MLBLUE estimator $\bar{\mathbf{s}}_{\mathrm{MLBLUE}}$ using Eq. 25.
6. return $\bar{\mathbf{s}}_{\mathrm{MLBLUE}}$, MSE

---

biased with approximation errors, solving Eqs. 27 and 21 can become numerically challenging. In particular, the MLBLUE optimization tends to select dissimilar model index groups, sample sizes, and weights even for similar computational budgets. In contrast, the MFMC method has been shown to be robust towards approximation errors in Γ (see [38]). Since MLBLUE has only recently been introduced, similar advancements have not yet been made, though they can likely be expected from future work.

4 Application to Greenland ice mass loss projections

We conclude this paper with a demonstration of the introduced multifidelity UQ methods MFMC, MLMC, and MLBLUE. Specifically, we estimate the expected ice mass loss of the Greenland ice sheet high-fidelity model under uncertainties in the basal friction field and the geothermal heat flux for the 2015–2050 time period. Our surrogate modeling setup is described in Section 4.1. For our demonstration, we distinguish between two UQ use cases: First, in Section 4.2 we prescribe a target accuracy in the estimated mean of ± 1 mm SLR equivalent (or ± 361.8 Gt) ice mass loss at a 95 % confidence level; second, in Section 4.3, we compare the accuracy of the 2015–2050 estimates obtained for a fixed computational budget of five high-fidelity model evaluations. We conclude with a discussion of the results in Section 4.4.

4.1 Surrogate models

In our multifidelity modeling setting, we employ $L = 13$ models of the ice mass change in Greenland for the years 2015–2050. Our high-fidelity model s_1 is governed by the equations for ice temperature, ice thickness, and the HO stress balance model for the ice velocity as described in Section 2.1. It is discretized on our finest available mesh, denoted "fine." Surrogate models s_3 and s_5 (named "HO, medium" and "HO, coarse", respectively) are obtained by replacing the high-fidelity mesh with the two coarser meshes denoted "medium" and "coarse," with initial conditions obtained by interpolating the $t = 2015$ high-fidelity initial condition followed by a relaxation run. The mesh resolutions are provided in Table 1. Surrogate models s_7 ("SSA, fine"), s_9 ("SSA, medium"), and s_{11} ("SSA, coarse"), are obtained by replacing the HO stress balance solver with the SSA, which approximates the three-dimensional velocity fields $\mathbf{v}_x$ and $\mathbf{v}_y$ by their two-dimensional depth-averages (see Section 2.3). All models are run using ISSM with adaptive time stepping adhering to a CFL condition to guarantee numerical stability. The average computational costs are listed in Table 2. We did not encounter any model blow-ups, and all runs concluded without errors and warnings.

For each of the ISSM models s_1, s_3, s_5, s_7, s_9, and s_{11}, we define data-fit surrogate models that we denote "extrapolation" data-fit model. These data fit surrogates are denoted s_2, s_4, s_6, s_8, s_{10}, and s_{12}, and are based on the observation that the ice mass loss is dominated by two trends, the overall, almost linear decline whose slope varies by parameter samples, and yearly fluctuations. The latter is primarily caused by the surface mass balance $\dot{M}_s$, and only implicitly

depends on the parameter samples through the coupling of $\dot{M}_s$ to the ice altitude, and the coupling of the basal melt $\dot{M}_b$ to the geothermal heat flux. Both trends can clearly be observed in Figs. 4 (left) and 5. To exploit this effect, we first compute nine reference samples of the "SSA, coarse" models, fit a linear function to each, and take the mean over the misfit to approximate the parameter-independent yearly adjustments. We then define "extrapolation" surrogate models by first running our HO or SSA models for 20 years using a sampled parameter, and then fitting a linear curve and the learned yearly adjustments to the thus obtained 2015-2035 predictions; to obtain predictions for $t \in [2035, 2050]$, we then evaluate the obtained data-fit model. The procedure is illustrated in Fig. 6. As the cost is dominated by the first simulation step in ISSM, each extrapolation model's predictions for 2015–2050 are $35/20 = 1.75$ times cheaper than its corresponding ISSM model.

Finally, for our cheapest surrogate model s_{13} we interpolate linearly between the nine reference ice mass change predictions of "SSA, coarse." This interpolation is extremely cheap but is — unsurprisingly — not particularly accurate for individual predictions, c.f. Figure 6. Still, as will be seen in the following results, the interpolation model achieves a high correlation with the high-fidelity model for the 2050 ice mass loss, making it beneficial to employ in the multifidelity estimators. Thus, this interpolation model demonstrates the benefit of cheap surrogate models even at the cost of reduced accuracy, when employed within a formal multifidelity estimation framework.

The multifidelity UQ methods require an estimate of the model covariance matrix $\Sigma \in \mathbb{R}^{13 \times 13}$. We use Latin Hypercube sampling to choose 32 parameter samples for computing all entries in Σ that depend on HO models. The ISSM model predictions for the parameters are shown in Fig. 5, and for the high-fidelity model additionally in Fig. 4. We do not remove outliers from the sampled data (as seen in Fig. 5 for the "SSA, medium" model) to avoid estimating an overly confident covariance matrix. For all entries in Σ that do not depend on the HO models but on the cheaper SSA models, we include an additional 96 parameters in the sampling to reduce potential bias in Σ. Finally, for $\Sigma_{13,13}$, the variance of the interpolation model s_{13}, we use $50,000$ samples. We record the model costs $c_1, \ldots, c_{13}$ as the average CPU time for computing the predictions at these samples. Note that we do not count these computations towards our budget for estimating the high-fidelity expectation because we expect that for state-of-the-art ice sheet models, Σ can be approximated from verification and validation runs performed during model development, or using convergence rates. If not, then the pilot samples computed for its estimation could also be reused in the multifidelity estimators.

Table 2 Surrogate model overview

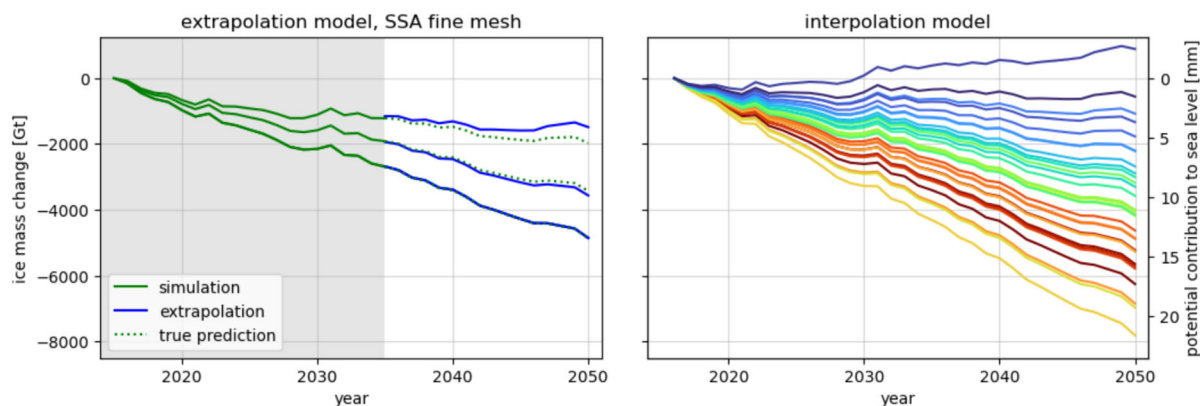| model no.	cost [CPU-h]	model setup		
		type	mesh	velocity
1	64.888	ISSM	fine	HO
2	37.079	extr.	fine	HO
3	22.889	ISSM	medium	HO
4	13.079	extr.	medium	HO
5	4.689	ISSM	coarse	HO
6	2.679	extr.	coarse	HO
7	0.455	ISSM	fine	SSA
8	0.260	extr.	fine	SSA
9	0.105	ISSM	medium	SSA
10	0.060	extr.	medium	SSA
11	0.059	ISSM	coarse	SSA
12	0.034	extr.	coarse	SSA
13	< 0.001	interp.	coarse	SSA

Fig. 6 Left: schematic explanation of extrapolation surrogate modeling setup. Right: Interpolation-based predictions for 32 samples of the geothermal heat flux and basal friction fields

4.2 Target accuracy

We start by comparing the MSE that can be achieved by MFMC, MLMC, and MLBLUE for different computational budgets when approximating the expected 2050 ice mass loss, see Fig. 7. Based on the heterogeneity of available models, the MSE for MFMC is slightly smaller than that of MLMC for the same budget, though both are larger than the MSE achieved by MLBLUE. Asymptotically, the MSE for the MFMC estimator is 13.0 times smaller than that of Monte Carlo for the same budget. Similarly, MLMC reduces the MSE by a factor of 10.2, and MLBLUE by a factor of 98.1. Interpreted in terms of accuracy, this means that the multifidelity estimators achieve a computational speedup between one and two orders of magnitude compared to MC sampling.

We compare the performance of the MFMC, MLMC, and MLBLUE methods when models and sample sizes are chosen to obtain a prescribed target accuracy of ± 361.8 Gt, corresponding to ± 1 mm SLR equivalent ice mass loss, at a 95 % confidence level. The computational budget required by the multifidelity methods to achieve this target accuracy is 30.14 CPU-days for MFMC, 37.52 CPU-days for MLMC, and 4.19 CPU-days for MLBLUE. These values are marked with dots in Fig. 7. Compared to the 381.22 CPU-days required by Monte Carlo sampling, these correspond to computational speed-ups of factors 12.6 for MFMC, 10.2 for MLMC, and 91.0 for MLBLUE. Note that the speedup for MLBLUE is smaller than its expected asymptotic value (98.1×) because the estimator is falling into the low-budget regime (see Fig. 7).

The optimal distribution of surrogate models and sampling costs for achieving the target accuracy are shown in Fig. 8. All three methods rely on information from all four surrogate model types (mesh coarsening, physical approximation, extrapolation, and interpolation). The high-fidelity model is sampled 7 times by MFMC, 12 by MLMC, and only once by MLBLUE, illustrating how all estimators are able to shift

the computational burden onto the surrogate models. MFMC and MLMC rely heavily on the computationally cheap but least accurate interpolation model with more than 180,000 samples each. In contrast, MLBLUE uses the interpolation model far less (61,573 samples), and opts for a larger variety in models. In particular, MLBLUE uses all six SSA surrogate models (both full 2015–2050 ISSM predictions and their extrapolations, each for the three available meshes) and the full 2015–2050 HO-coarse, while MFMC and MLMC only use extrapolation and interpolation type surrogates.

We next approximate the expected high-fidelity ice mass loss for the year 2050 $\mathbb{E}[s_1(2050)]$ with each multifidelity UQ method, using the optimal sample sizes from Fig. 8 to guarantee an accuracy of at most ± 1 mm SLR equivalent ice mass loss at a 95 % confidence level. The predictions are provided in Table 3, with confidence levels illustrated in Fig. 9. For comparison, we also provide Monte Carlo predictions at the same accuracy, and at ± 0.54 mm SLR equivalent ice mass loss (at 95 % confidence). All five confidence intervals overlap on the interval [15.23, 16.07].

4.3 Target budget

We next demonstrate the power of multifidelity UQ when working with a restrictive computational budget. For a budget $c = 13.5$ CPU-days, corresponding to five solves of the high-fidelity model, we compute one instance of each multifidelity estimator and its 95% confidence interval for predicting the expected 2015–2050 ice mass change. The predictions are shown in Fig. 10. The uncertainty in all estimators increases in time, with Monte Carlo sampling performing the worst — as expected for the small number of high-fidelity samples. Its final 95% confidence interval at $t = 2050$ is $[-5132.9 \pm 1913.1]$ Gt, or $[14.19 \pm 5.29]$ mm SLR equivalent ice mass loss, as illustrated in Fig. 9. In contrast, the final confidence intervals for MFMC, MLMC, and MLBLUE are $[-5581.1 \pm$

Fig. 7 Comparison of root MSE of the multifidelity UQ methods MFMC, MLMC, and MLBLUE, and Monte Carlo sampling for any given computational budget
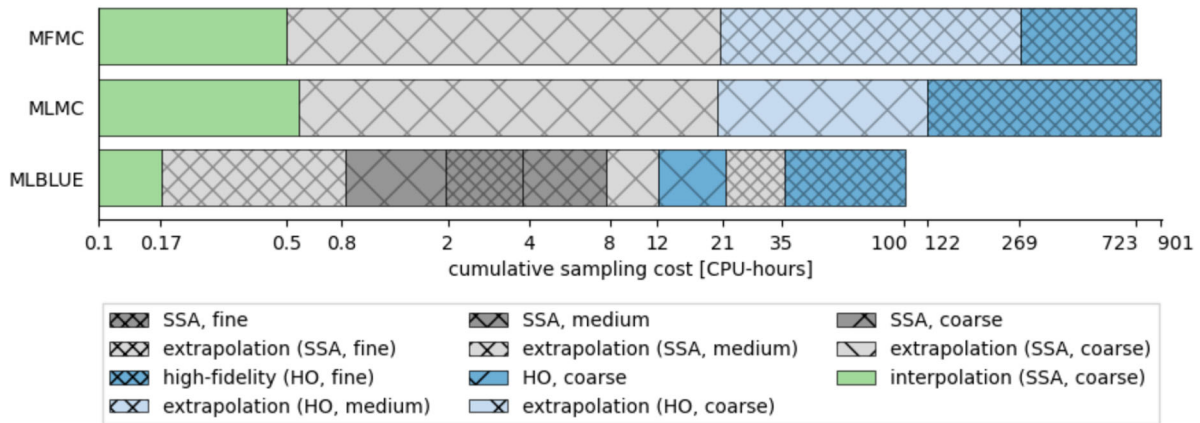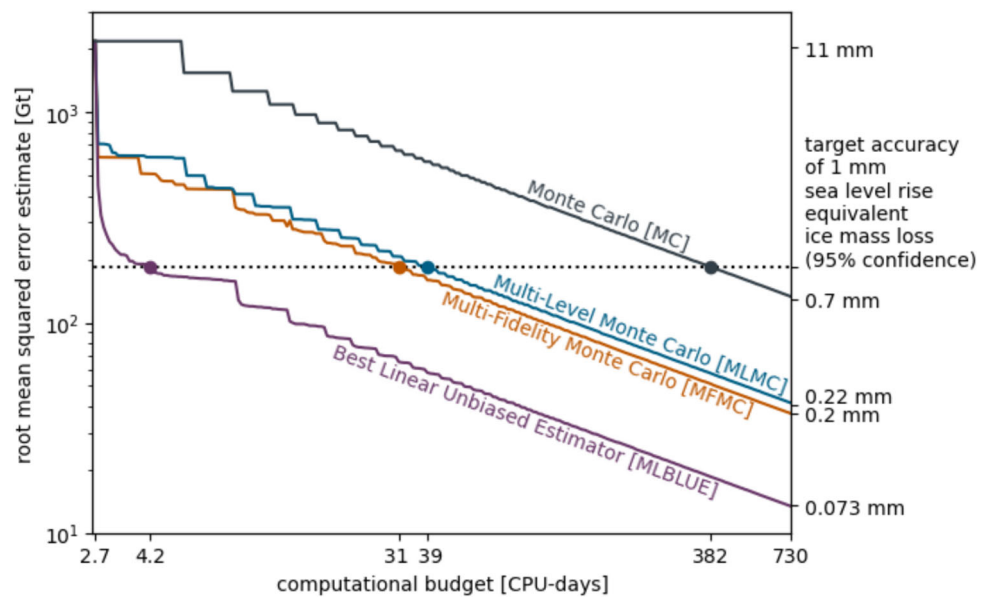

Fig. 8 Distribution of computational budget for MFMC, MLMC, and MLBLUE for reaching a target accuracy of ±361.8 Gt at 95 % confidence

Table 3 Expected Greenland ice mass loss by 2050: estimates obtained by MC and multifidelity UQ methods and associated uncertainty at a 95 % confidence level

setting	method	ice mass loss [Gt]		SLR contribution [mm]		cost [CPU-days]
		estimate	uncertainty	estimate	uncertainty	
	MC	-5511	±360	15.2	±1.00	381.2
target	MFMC	-5868	±356	16.2	±0.98	30.1
accuracy	MLMC	-5757	±352	15.9	±0.97	37.5
	MLBLUE	-5725	±346	15.8	±0.96	4.2
reference	MC	-5618	±195	15.5	±0.54	1297.8
	MLBLUE	-5572	±193	15.4	±0.53	13.5
target	MLMC	-5639	±608	15.6	±1.68	12.6
budget	MFMC	-5581	±612	15.4	±1.69	11.8
	MC	-5586	±1913	15.4	±5.29	13.5
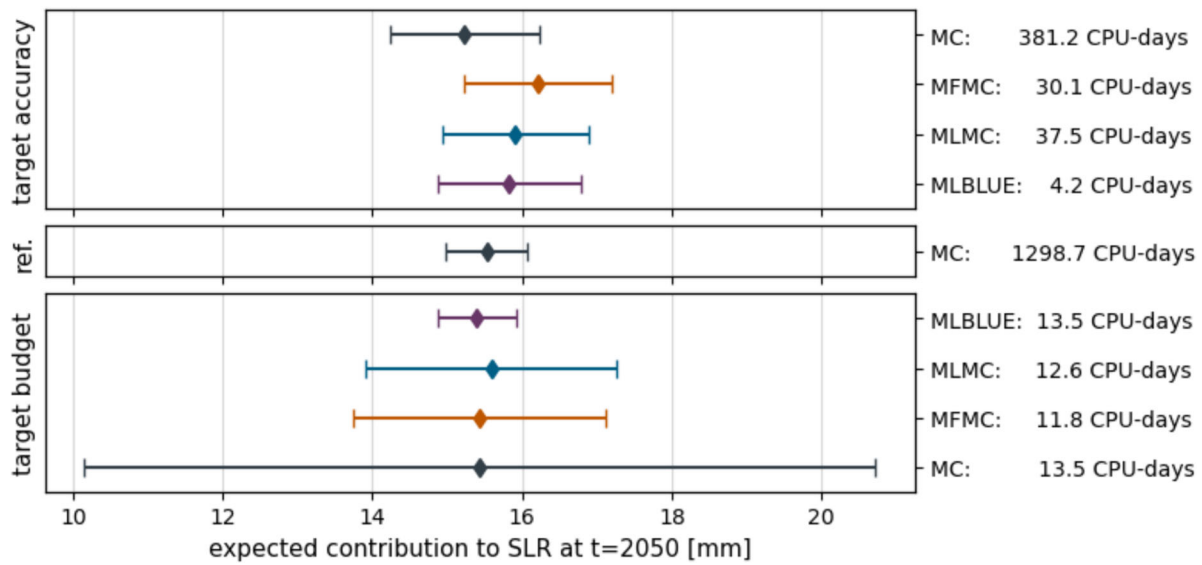
Fig. 9 Confidence intervals for expected Greenland ice mass change in 2050 relative to 2015 and control run

611.6] Gt, [−5639.4±607.7] Gt, and [−5572.2±193.1] Gt, corresponding to an accuracy of ±1.69, ±1.68, and ±0.53 mm SLR equivalent ice mass loss at 95% confidence. For Monte Carlo sampling to achieve these accuracies at this level of confidence, we would require a computational budget of 132.5, 135.2 and 1327.5 CPU-days (49, 50, and 491 high-fidelity model solves).

We note that for Fig. 10 the sample sizes and weights of the multifidelity UQ estimators were chosen to minimize the error incurring throughout the 2015–2050 prediction

regime. This stands in contrast to the values reported in Fig. 7, where the estimators were optimized for the single prediction at $t = 2050$. In particular, in Fig. 10, MLMC performed better than MFMC with up to 5% smaller confidence intervals. However, after rounding the sample sizes, neither method fully exploited the computational budget but only used 87.348 % (MFMC) and 92.882 % (MLMC), while MLBLUE used 99.998 %. Thus, the performance of both methods can still be improved by redistributing the remaining computational budget, e.g., via [81].
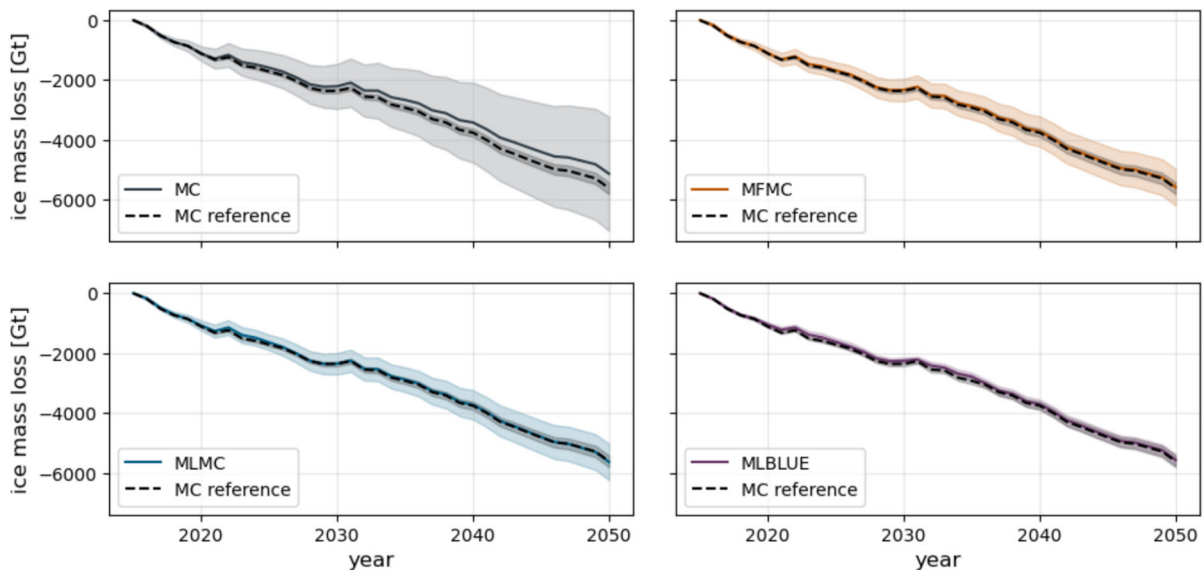
Fig. 10 Estimates of the 2015-2050 ice mass change and 95% confidence interval (shaded) for a computational budget of 324.44 CPU-h.

4.4 Conclusion

In this paper we have explained three multifidelity UQ techniques — MFMC, MLMC, and MLBLUE — and applied them to compute the expected 2050 ice mass loss of the Greenland ice sheet. Despite strong variability in individual high-fidelity predictions caused by uncertain basal friction and geothermal heat flux fields, the multifidelity UQ estimators achieve an approximation accuracy of ± 1 mm SLR equivalent ice mass loss at a 95 % confidence level with up to $91\times$ speed-up compared to Monte Carlo sampling. In the low-budget regime, the MLBLUE estimate for a computational budget of five high-fidelity solves has a variance that would require 491 Monte Carlo samples. Overall, the three multifidelity UQ estimators have proven to be well-suited for computing the output expectations of highly expensive ice sheet simulations.

In our numerical experiments, MFMC and MLMC have performed similarly well to each other but have been outperformed by MLBLUE. This result was expected as MLBLUE guarantees by construction an estimator variance that is smaller or equal to that of both the MFMC and the MLMC estimator. We still recommend both MFMC and MLMC for UQ in ice sheet simulations, primarily because both methods have a strong foundation in the literature with extensions beyond forward UQ (e.g., Markov chain Monte Carlo and sensitivity analysis). In contrast, similar extensions for MLBLUE are still in active development. Moreover, while MFMC and MLMC are generally robust against against approximation bias in the input covariance matrix, the MLBLUE sample size optimization can be challenging. Our recommended procedure after computing the input covariance matrix is to first evaluate the potential benefit of each method (e.g., in the form of Fig. 7) before committing to one. In either case, the results show clearly that the existing surrogate models in the ice sheet literature are sufficient to enable UQ even for highly expensive high-fidelity ice sheet models.

Naturally, the performance of the multifidelity UQ methods depends on the high-fidelity model and its surrogates. We have chosen our modeling setup here to be as close to the ISMIP6 protocol [43] as possible while still permitting comparisons with Monte Carlo sampling to allow an easy transfer of techniques to other models and codes. In addition, we have employed surrogate models that are readily available and do not require expert-level surrogate modeling implementations. We consequently might expect even better performance of the multifidelity UQ methods if specialized reduced-order models are employed among the surrogates.

Research data availability The BedMachine v5 dataset on the Greenland bedrock topography and ice thickness is available at nsidc.org/data/idbmg4/versions/5 . The GIMP ice and ocean mask (v2.0) is available at byrd.osu.edu/research/groups/glacier-dynamics/data/icemask. Its combination with the coastline by Jeremie Mouginot is available at https://issm.jpl.nasa.gov/documentation/tutorials/datasets/ (SeaRISE Greenland dev1.2). The MEaSUREs Multi-year Greenland Ice Sheet Velocity Mosaic (version 1) is available at nsidc.org/data/nsidc-0670/versions/1. The MEaSUREs Greenland Annual Ice Sheet Velocity Mosaics from SAR and Landsat (version 4, ID NSIDC-0725) for the years 2015-2021 are available at nsidc.org/data/nsidc-0725/versions/4. The CNRM-CM6-1 atmospheric projections are available on theghub.org/. The Greenland heat flux A20180227-001 is available at ads.nipr.ac.jp/data/meta/A20180227-001/ . The Greenland heat flux "Shapiro-Ritzwoller" is available at http://ciei.colorado.edu/~nshapiro/MODEL/.

Code availability The code and prediction data used to generate the results in this study is available at github.com/nicolearetz/Multifidelity-UQ-Greenland. The implementations of MFMC, MLMC, and MLBLUE are available at github.com/croci/bluest. The Ice-sheet and Sea-level System Model is available at https://github.com/ISSMteam/ISSM

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Goelzer, H., Nowicki, S., Payne, A., Larour, E., Seroussi, H., Lipscomb, W.H., Gregory, J., Abe-Ouchi, A., Shepherd, A., Simon, E., et al.: The future sea-level contribution of the Greenland ice sheet: a multi-model ensemble study of ISMIP6. The Cryosphere. **14**(9), 3071–3096 (2020)

2. Goelzer, H., Nowicki, S., Edwards, T., Beckley, M., Abe-Ouchi, A., Aschwanden, A., Calov, R., Gagliardini, O., Gillet-Chaulet, F., Golledge, N.R., et al.: Design and results of the ice sheet

model initialisation experiments initMIP-Greenland: an ISMIP6 intercomparison. The Cryosphere. **12**(4), 1433–1460 (2018)

3. Seroussi, H., Nowicki, S., Payne, A.J., Goelzer, H., Lipscomb, W.H., Abe Ouchi, A., Agosta, C., Albrecht, T., Asay-Davis, X., Barthel, A., et al.: ISMIP6 Antarctica: a multi-model ensemble of the Antarctic ice sheet evolution over the 21st century. The Cryosphere Discussions. **2020**, 1–54 (2020)

4. Seroussi, H., Nowicki, S., Simon, E., Abe-Ouchi, A., Albrecht, T., Brondex, J., Cornford, S., Dumas, C., Gillet-Chaulet, F., Goelzer, H., et al.: initMIP-Antarctica: an ice sheet model initialization experiment of ISMIP6. The Cryosphere. **13**(5), 1441–1471 (2019)

5. Huybrechts, P., Payne, T., et al.: The EISMINT benchmarks for testing ice-sheet models. Annal. Glaciol. **23**, 1–12 (1996)

6. Pattyn, F., Perichon, L., Aschwanden, A., Breuer, B., De Smedt, B., Gagliardini, O., Gudmundsson, G.H., Hindmarsh, R.C., Hubbard, A., Johnson, J.V., et al.: Benchmark experiments for higher-order and full-Stokes ice sheet models (ISMIP-HOM). The Cryosphere. **2**(2), 95–108 (2008)

7. Payne, A., Huybrechts, P., Abe-Ouchi, A., Calov, R., Fastook, J., Greve, R., Marshall, S., Marsiat, I., Ritz, C., Tarasov, L., et al.: Results from the EISMINT model intercomparison: the effects of thermomechanical coupling. J. Glaciol. **46**(153), 227–238 (2000)

8. Hock, R., Bliss, A., Marzeion, B., Giesen, R.H., Hirabayashi, Y., Huss, M., Radić, V., Slangen, A.B.: GlacierMIP-a model intercomparison of global-scale glacier mass-balance models and projections. J. Glaciol. **65**(251), 453–467 (2019)

9. Van Angelen, J., Lenaerts, J., Lhermitte, S., Fettweis, X., Kuipers Munneke, P., Broeke, M., Van Meijgaard, E., Smeets, C.: Sensitivity of Greenland Ice Sheet surface mass balance to surface albedo parameterization: a study with a regional climate model. The Cryosphere. **6**(5), 1175–1186 (2012)

10. Rogozhina, I., Hagedoorn, J., Martinec, Z., Fleming, K., Soucek, O., Greve, R., Thomas, M.: Effects of uncertainties in the geothermal heat flux distribution on the Greenland Ice Sheet: An assessment of existing heat flow models. J. Geophys. Res.: Earth Surface. **117**(F2) (2012)

11. Larour, E., Morlighem, M., Seroussi, H., Schiermeier, J., Rignot, E.: Ice flow sensitivity to geothermal heat flux of Pine Island Glacier, Antarctica. J. Geophys. Res.: Earth Surface. **117**(F4) (2012)

12. Barnes, J.M., Gudmundsson, G.H.: The predictive power of ice sheet models and the regional sensitivity of ice loss to basal sliding parameterisations: a case study of Pine Island and Thwaites glaciers. West Antarctica. The Cryosphere. **16**(10), 4291–4304 (2022)

13. Zhao, C., Gladstone, R.M., Warner, R.C., King, M.A., Zwinger, T., Morlighem, M.: Basal friction of Fleming Glacier, Antarctica-part 1: Sensitivity of inversion to temperature and bedrock uncertainty. The Cryosphere. **12**(8), 2637–2652 (2018)

14. Cheng, G., Kirchner, N., Lötstedt, P.: Sensitivity of ice sheet surface velocity and elevation to variations in basal friction and topography in the full Stokes and shallow-shelf approximation frameworks using adjoint equations. The Cryosphere. **15**(2), 715–742 (2021)

15. Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., Aubry, D.: Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica. Geophys. Res. Lett. **37**(14) (2010)

16. Morlighem, M., Seroussi, H., Larour, E., Rignot, E.: Inversion of basal friction in Antarctica using exact and incomplete adjoints of a higher-order model. J. Geophys. Res.: Earth Surface. **118**(3), 1746–1753 (2013)

17. Isaac, T., Petra, N., Stadler, G., Ghattas, O.: Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. J. Comput. Phys. **296**, 348–368 (2015)

18. Babaniyi, O., Nicholson, R., Villa, U., Petra, N.: Inferring the basal sliding coefficient field for the Stokes ice sheet model under rheological uncertainty. The Cryosphere. **15**(4), 1731–1750 (2021)

19. Brinkerhoff, D., Aschwanden, A., Fahnestock, M.: Constraining subglacial processes from surface velocity observations using surrogate-based Bayesian inference. J. Glaciol. **67**(263), 385–403 (2021)

20. Shapiro, N.M., Ritzwoller, M.H.: Inferring surface heat flux distributions guided by a global seismic model: particular application to Antarctica. Earth Planetary Sci. Lett. **223**(1–2), 213–224 (2004)

21. Greve, R.: Geothermal heat flux distribution for the Greenland ice sheet, derived by combining a global representation and information from deep ice cores. Polar Data J. **3**, 22–36 (2019)

22. Morlighem, M., Williams, C.N., Rignot, E., An, L., Arndt, J.E., Bamber, J.L., Catania, G., Chauché, N., Dowdeswell, J.A., Dorschel, B., et al.: Bedmachine v3: Complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation. Geophys. Res. Lett. **44**(21), 11–051 (2017)

23. Brinkerhoff, D.J., Aschwanden, A., Truffer, M.: Bayesian inference of subglacial topography using mass conservation. Front. Earth Sci. **4**, 8 (2016)

24. Guan, Y., Haran, M., Pollard, D.: Inferring ice thickness from a glacier dynamics model and multiple surface data sets. Environmetrics. **29**(5–6), 2460 (2018)

25. Bamber, J.L., Aspinall, W.: An expert judgement assessment of future sea level rise from the ice sheets. Nature Climate Change. **3**(4), 424–427 (2013)

26. Oppenheimer, M., Little, C.M., Cooke, R.M.: Expert judgement and uncertainty quantification for climate change. Nature Climate Change. **6**(5), 445–451 (2016)

27. Machguth, H., Purves, R.S., Oerlemans, J., Hoelzle, M., Paul, F.: Exploring uncertainty in glacier mass balance modelling with Monte Carlo simulation. The Cryosphere. **2**(2), 191–204 (2008)

28. Golledge, N.R., Kowalewski, D.E., Naish, T.R., Levy, R.H., Fogwill, C.J., Gasson, E.G.: The multi-millennial Antarctic commitment to future sea-level rise. Nature. **526**(7573), 421–425 (2015)

29. Ritz, C., Edwards, T.L., Durand, G., Payne, A.J., Peyaud, V., Hindmarsh, R.C.: Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. Nature. **528**(7580), 115–118 (2015)

30. DeConto, R.M., Pollard, D.: Contribution of Antarctica to past and future sea-level rise. Nature. **531**(7596), 591–597 (2016)

31. Edwards, T.L., Brandon, M.A., Durand, G., Edwards, N.R., Golledge, N.R., Holden, P.B., Nias, I.J., Payne, A.J., Ritz, C., Wernecke, A.: Revisiting Antarctic ice loss due to marine ice-cliff instability. Nature. **566**(7742), 58–64 (2019)

32. Little, C.M., Oppenheimer, M., Urban, N.M.: Upper bounds on twenty-first-century Antarctic ice loss assessed using a probabilistic framework. Nature Climate Change. **3**(7), 654–659 (2013)

33. Bulthuis, K., Arnst, M., Sun, S., Pattyn, F.: Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change. The Cryosphere. **13**(4), 1349–1380 (2019)

34. Pollard, D., Chang, W., Haran, M., Applegate, P., DeConto, R.: Large ensemble modeling of the last deglacial retreat of the West Antarctic Ice Sheet: comparison of simple and advanced statistical techniques. Geoscientific Model Develop. **9**(5), 1697–1723 (2016)

35. Aschwanden, A., Fahnestock, M.A., Truffer, M., Brinkerhoff, D.J., Hock, R., Khroulev, C., Mottram, R., Khan, S.A.: Contribution of the Greenland Ice Sheet to sea level over the next millennium. Sci. Adv. **5**(6), 9396 (2019)

36. Edwards, T.L., Nowicki, S., Marzeion, B., Hock, R., Goelzer, H., Seroussi, H., Jourdain, N.C., Slater, D.A., Turner, F.E., Smith, C.J., et al.: Projected land ice contributions to twenty-first-century sea level rise. Nature. **593**(7857), 74–82 (2021)

37. Ng, L.W., Willcox, K.E.: Multifidelity approaches for optimization under uncertainty. Int. J. Numerical Methods Eng. **100**(10), 746–772 (2014)

38. Peherstorfer, B., Willcox, K., Gunzburger, M.: Optimal model management for multifidelity Monte Carlo estimation. SIAM J. Scientific Comput. **38**(5), 3163–3194 (2016)

39. Heinrich, S.: Multilevel Monte Carlo methods. In: Large-Scale Scientific Computing: Third International Conference, LSSC 2001 Sozopol, Bulgaria, June 6–10, 2001 Revised Papers 3, pp. 58–67 (2001). Springer

40. Giles, M.B.: Multilevel Monte Carlo path simulation. Operations Res. **56**(3), 607–617 (2008)

41. Schaden, D., Ullmann, E.: On multilevel best linear unbiased estimators. SIAM/ASA J. Uncertainty Quantification. **8**(2), 601–635 (2020)

42. Larour, E., Seroussi, H., Morlighem, M., Rignot, E.: Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM). J. Geophys. Res.: Earth Surface. **117**(F1) (2012)

43. Nowicki, S., Payne, A.J., Goelzer, H., Seroussi, H., Lipscomb, W.H., Abe-Ouchi, A., Agosta, C., Alexander, P., Asay-Davis, X.S., Barthel, A., et al.: Experimental protocol for sealevel projections from ISMIP6 standalone ice sheet models. The Cryosphere Discussions. **2020**, 1–40 (2020)

44. Nowicki, S.M., Payne, A., Larour, E., Seroussi, H., Goelzer, H., Lipscomb, W., Gregory, J., Abe-Ouchi, A., Shepherd, A.: Ice sheet model intercomparison project (ismip6) contribution to cmip6. Geoscientific Model Develop. **9**(12), 4521–4545 (2016)

45. Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E.: Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. Geoscientific Model Develop. **9**(5), 1937–1958 (2016)

46. IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change vol. In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2021). https://doi.org/10.1017/9781009157896

47. Chen, D., Rojas, M., Samset, B.H., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S.H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S.K., Plattner, G.-K., Tréguier, A.-M.: In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (eds.) Framing, Context, and Methods, pp. 147–286. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2021). https://doi.org/10.1017/9781009157896.003

48. Morlighem, M.e.a.: IceBridge BedMachine Greenland, Version 5. NASA National Snow and Ice Data Center Distributed Active Archive Center (2022). https://doi.org/10.5067/GMEVBWFLWA7X . https://nsidc.org/data/IDBMG4/versions/5

49. Howat, I.M., Negrete, A., Smith, B.E.: The Greenland Ice Mapping Project (GIMP) land classification and surface elevation data sets. The Cryosphere. **8**(4), 1509–1518 (2014)

50. Joughin, I., Smith, B., Howat, I., Scambos, T.: MEaSUREs Multi-year Greenland Ice Sheet Velocity Mosaic, Version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center. (2016). https://doi.org/10.5067/QUA5Q9SVMSJG. https://nsidc.org/data/NSIDC-0670/versions/1

51. Voldoire, A.: Cnrm-cerfacs cnrm-cm6-1 model output prepared for cmip6 cmip (2018)

52. Mouginot, J., Rignot, E., Bjørk, A.A., Broeke, M., Millan, R., Morlighem, M., Noël, B., Scheuchl, B., Wood, M.: Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. Proceed. National Academy Sci. **116**(19), 9239–9244 (2019)

53. Rignot, E., Kanagaratnam, P.: Changes in the velocity structure of the Greenland Ice Sheet. Science. **311**(5763), 986–990 (2006)

54. Greve, R., Blatter, H.: Dynamics of Ice Sheets and Glaciers. Springer, ??? (2009)

55. Paterson, W.S.B.: The Physics of Glaciers, 3rd edn. Pergamon Press, Oxford, London, New York, ??? (1994)

56. Budd, W.F., Keage, P.L., Blundy, N.A.: Empirical studies of ice sliding. J. Glaciol. **23**, 157–170 (1979)

57. Chen, Q., Gunzburger, M., Perego, M.: Well-posedness results for a nonlinear Stokes problem arising in glaciology. SIAM J. Math. Anal. **45**(5), 2710–2733 (2013)

58. Leng, W., Ju, L., Xie, Y., Cui, T., Gunzburger, M.: Finite element three-dimensional Stokes ice sheet dynamics model with enhanced local mass conservation. J. Comput. Phys. **274**, 299–311 (2014)

59. Joughin, I.: MEaSUREs Greenland Annual Ice Sheet Velocity Mosaics from SAR and Landsat, Version 4. NASA National Snow and Ice Data Center Distributed Active Archive Center. (2022). https://doi.org/10.5067/RS8GFZ848ZU9. https://nsidc.org/data/NSIDC-0725/versions/4

60. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. **2**(4), 433–459 (2010)

61. Colgan, W., Wansing, A., Mankoff, K., Lösing, M., Hopper, J., Louden, K., Ebbing, J., Christiansen, F.G., Ingeman-Nielsen, T., Liljedahl, L.C., et al.: Greenland geothermal heat flow database and map (version 1). Earth Syst. Sci. Data Discussions. **2021**, 1–50 (2021)

62. Rezvanbehbahani, S., Stearns, L.A., Kadivar, A., Walker, J.D., Veen, C.J.: Predicting the geothermal heat flux in Greenland: A machine learning approach. Geophys. Res. Lett. **44**(24), 12–271 (2017)

63. Colgan, W., MacGregor, J.A., Mankoff, K.D., Haagenson, R., Rajaram, H., Martos, Y.M., Morlighem, M., Fahnestock, M.A., Kjeldsen, K.K.: Topographic correction of geothermal heat flux in Greenland and Antarctica. J. Geophys. Res.: Earth Surface. **126**(2), 2020–005598 (2021)

64. Martos, Y.M., Jordan, T.A., Catalán, M., Jordan, T.M., Bamber, J.L., Vaughan, D.G.: Geothermal heat flux reveals the Iceland hotspot track underneath Greenland. Geophys. Res. Lett. **45**(16), 8214–8222 (2018)

65. Kolster, M.E., Døssing, A., Khan, S.A.: Satellite magnetics suggest a complex geothermal heat flux pattern beneath the Greenland ice sheet. Remote Sens. **15**(5), 1379 (2023)

66. Artemieva, I.M.: Lithosphere thermal thickness and geothermal heat flux in Greenland from a new thermal isostasy method. Earth-Sci. Rev. **188**, 469–481 (2019)

67. Pollack, H.N., Hurter, S.J., Johnson, J.R.: Heat flow from the Earth's interior: analysis of the global data set. Rev. Geophys. **31**(3), 267–280 (1993)

68. Bulthuis, K., Larour, E.: Implementation of a Gaussian Markov random field sampler for forward uncertainty quantification in the Ice-sheet and Sea-level System Model v4. 19. Geosci. Model Dev. **15**(3), 1195–1217 (2022)

69. Lampkin, D.J. VanderBerg, J.: A preliminary investigation of the influence of basal and surface topography on supraglacial lake distribution near Jakobshavn Isbrae, western Greenland. Hydrol. Process. **25**(21), 3347–3355 (2011)

70. MacAyeal, D.R.: Large-scale ice flow over a viscous basal sediment: Theory and application to Ice Stream B. Antarctica. J. Geophys. Res. **94**(B4), 4071–4087 (1989)

71. Le Meur, E., Gagliardini, O., Zwinger, T., Ruokolainen, J.: Glacier flow modelling: a comparison of the Shallow Ice Approximation and the full-Stokes solution. Comptes rendus. Physique **5**(7), 709–722 (2004)

72. Brinkerhoff, D., Johnson, J.: Dynamics of thermally induced ice streams simulated with a higher-order flow model. J. Geophys. Res.: Earth Surface. **120**(9), 1743–1770 (2015)

73. Santos, T., Morlighem, M., Brinkerhoff, D.: A new vertically integrated MOno-Layer Higher-Order (MOLHO) ice flow model. The Cryosphere. **16**(1), 179–195 (2022)

74. Pattyn, F.: Antarctic subglacial conditions inferred from a hybrid ice sheet/ice stream model. Earth Planetary Sci. Lett. **295**(3–4), 451–461 (2010)

75. Pollard, D., DeConto, R.: Description of a hybrid ice sheet-shelf model, and application to Antarctica. Geoscientific Model Develop. **5**(5), 1273–1295 (2012)

76. Riel, B., Minchew, B.: Variational inference of ice shelf rheology with physics-informed machine learning. J. Glaciol. **69**(277), 1167–1186 (2023)

77. Jouvet, G., Cordonnier, G., Kim, B., Lüthi, M., Vieli, A., Aschwanden, A.: Deep learning speeds up ice flow modelling by several orders of magnitude. J. Glaciol. **68**(270), 651–664 (2022)

78. He, Q., Perego, M., Howard, A.A., Karniadakis, G.E., Stinis, P.: A hybrid deep neural operator/finite element method for ice-sheet modeling. J. Comput. Phys. **492**, 112428 (2023)

79. Gorodetsky, A.A., Geraci, G., Eldred, M.S., Jakeman, J.D.: A generalized approximate control variate framework for multifidelity uncertainty quantification. J. Comput. Phys. **408**, 109257 (2020)

80. Croci, M., Willcox, K.E., Wright, S.J.: Multi-output multilevel best linear unbiased estimators via semidefinite programming. Comput. Methods Appl. Mech. Eng. **413**, 116130 (2023)

81. Gruber, A., Gunzburger, M., Ju, L., Wang, Z.: A multifidelity Monte Carlo method for realistic computational budgets. Journal of Scientific Computing **94**(1), 2 (2023)

82. Qian, E., Peherstorfer, B., O'Malley, D., Vesselinov, V.V., Willcox, K.: Multifidelity Monte Carlo estimation of variance and sensitivity indices. SIAM/ASA J. Uncertainty Quantification. **6**(2), 683–706 (2018)

83. Peherstorfer, B.: Multifidelity Monte Carlo estimation with adaptive low-fidelity models. SIAM/ASA J. Uncertainty Quantification. **7**(2), 579–603 (2019)

84. Farcaș, I.-G., Peherstorfer, B., Neckel, T., Jenko, F., Bungartz, H.-J.: Context-aware learning of hierarchies of low-fidelity models for multi-fidelity uncertainty quantification. Comput. Methods Appl. Mech. Eng. **406**, 115908 (2023)

85. Gruber, A., Gunzburger, M., Ju, L., Lan, R., Wang, Z.: Multifidelity Monte Carlo estimation for efficient uncertainty quantification in climate-related modeling. Geoscientific Model Develop. **16**(4), 1213–1229 (2023)

86. Haji-Ali, A.-L., Nobile, F., Tempone, R.: Multi-index Monte Carlo: when sparsity meets sampling. Numerische Mathematik. **132**, 767–806 (2016)

87. McLeish, D.: A general method for debiasing a Monte Carlo estimator. Monte Carlo Methods Appl. **17**(4), 301–315 (2011)

88. Blanchet, J.H., Glynn, P.W.: Unbiased Monte Carlo for optimization and functions of expectations via multi-level randomization. In: 2015 Winter Simulation Conference (WSC), pp. 3656–3667 (2015). IEEE

89. Hoel, H., Von Schwerin, E., Szepessy, A., Tempone, R.: Adaptive multilevel Monte Carlo simulation. In: Numerical Analysis of Multiscale Computations: Proceedings of a Winter Workshop at the Banff International Research Station 2009, pp. 217–234 (2011). Springer

90. Giles, M.B., Waterhouse, B.J.: Multilevel quasi-Monte Carlo path simulation. Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics **8**, 165–181 (2009)

91. Dodwell, T.J., Ketelsen, C., Scheichl, R., Teckentrup, A.L.: Multilevel Markov chain Monte Carlo. Siam Rev. **61**(3), 509–545 (2019)

92. Giles, M.B.: Multilevel Monte Carlo methods. Acta Numerica. **24**, 259–328 (2015)

93. Schaden, D., Ullmann, E.: Asymptotic analysis of multilevel best linear unbiased estimators. SIAM/ASA J. Uncertainty Quantification. **9**(3), 953–978 (2021)

94. Groß, J.: The general Gauss-Markov model with possibly singular dispersion matrix. Stat. Papers. **45**, 311–336 (2004)