Regularization and L-curves in ice sheet inverse models, a case study in the Filchner-Ronne catchment

Michael Wolovick[1], Angelika Humbert[1,2], Thomas Kleiner[1], and Martin Rückamp[3]

[1]Glaciology Section, Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
[2]University of Bremen, Department of Geosciences, Bremen, Germany
[3]Bavarian Academy of Sciences and Humanities, Munich, Germany

Correspondence: Michael Wolovick (michael.wolovick@awi.de)

Abstract. Over the past three decades, inversions for ice sheet basal drag have become commonplace in glaciological modeling. Such inversions require regularization to prevent over-fitting and ensure that the structure they recover is a robust inference from the observations, confidence which is required if they are to be used to draw conclusions about processes and properties of the ice base. While L-curve analysis can be used to select the optimal regularization level, the treatment of L-curve analysis in glaciological inverse modeling has been highly variable. Building on the history of glaciological inverse modeling, we demonstrate general best practices for regularizing glaciological inverse problems, using a domain in the Filchner-Ronne catchment of Antarctica as our test bed. We show a step by step approach to cost function normalization and L-curve analysis. We explore the spatial and spectral characteristics of the solution as a function of regularization, and we test the sensitivity of L-curve analysis and regularization to model resolution, effective pressure, sliding nonlinearity, and the flow equation. We find that the optimal regularization level converges towards a finite non-zero limit in the continuous problem, associated with a best knowable basal drag field. Nonlinear sliding laws outperform linear sliding in our analysis, with both a lower total variance and a more sharply cornered L-curve. By contrast, geometry-based approximations for effective pressure degrade inversion performance when added to a sliding law, but an actual hydrology model may marginally improve performance in some cases. Our results with 3D inversions suggest that the additional model complexity may not be justified by the 2D nature of the surface velocity data. We conclude with recommendations for best practices in future glaciological inversions.

1 Introduction

> With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.
>
> -John von Neumann

Ever since the pioneering "tutorial" of MacAyeal (1993), inversions for ice sheet basal drag- and, less commonly, englacial rheology- have been an important part of the ice sheet modeler's toolbox. In these inversions, a numerical model of ice sheet flow is compared to observations of ice surface velocity using a cost function, and the variable to be constrained (a field of either drag or rheology) is iteratively adjusted until the cost function has dropped to an acceptable level. In the three decades

since MacAyeal's original paper, remotely sensed observations of ice sheet surface velocity have become available for nearly the entirety of the Greenland and Antarctic ice sheets, along with many smaller ice caps and mountain glaciers (Joughin et al., 2010b; Rignot et al., 2011, 2017; Fahnestock et al., 2016), while knowledge of ice sheet geometry has improved dramatically as well (Lythe and Vaughan, 2001; Fretwell et al., 2013; Morlighem et al., 2014, 2020; Bamber et al., 2001, 2009, 2013). Concurrently, the computing power available to run these models has increased enormously. Thus, it has now become routine in glaciological modeling to use inversions to study the spatial- and, with repeat observations of surface velocity, temporal-variability of the ice sheet basal drag and to initialize transient models of the ice sheet future (Joughin et al., 2004, 2006, 2009; Morlighem et al., 2010, 2013; Gillet-Chaulet et al., 2012, 2016; Habermann et al., 2013; Sergienko and Hindmarsh, 2013; Sergienko et al., 2014; Shapero et al., 2016).

However, the question remains of just how much information can truly be gleamed from ice model inversions. Fundamentally, the trouble is that a spatially resolved inversion has an infinite number of free parameters, and in a numerical implementation of the problem, the number of free parameters is determined by the discritization used, not by anything fundamental in the problem. Furthermore, the transfer function relating variations in basal shear stress to surface velocity or slope is essentially a low-pass filter (Gudmundsson, 2003), allowing dramatically different fine-scale drag fields to produce similar velocity observations (Habermann et al., 2012), and ensuring that the inverse problem (using surface velocity to infer basal stress) is ill-posed. Ill-posed inversions require regularization to create a well-posed problem and avoid overfitting noise in the data. The most common form of regularization used in glaciological inversions is Tikhonov regularization (Tikhonov and Arsenin, 1977), which involves adding an additional term to the cost function that penalizes variability in the inversion target field, usually through a measure of gradient or curvature, and combining this regularization cost (J_{reg}) with the original data cost term (J_{obs}) to create a combined cost function (J) that is minimized by the inversion. A weighting parameter, which we write as λ, controls the relative contribution of the data and regularization terms to the overall cost function that is minimized in the inversion.

Yet this introduces a new free parameter into the problem: if λ is too high, the inversion will under-fit the data, producing very little structure even if the observations could be used to infer greater detail, while if λ is too low, the inversion will over-fit the data, producing spurious structure that is not truly required by the observations. While other methods of determining the optimal regularization level, such as Morozov's discrepency priciple, have occasionally been used in glaciological inversions (e.g. Martin and Monnier, 2014), the most common way to determine the optimal value of λ is through an L-curve analysis, in which J_{obs} and J_{reg} are plotted against one another for different values of λ, ideally showing a corner (the "L" shape) at the value of λ that best balances between the extremes of over-fitting and under-fitting (Hansen, 1992; Hansen and O'Leary, 1993). This corner can be rigorously identified by measuring the curvature of the L-curve in log-log space (Hansen and O'Leary, 1993). Even when performed correctly, L-curve analysis is not perfect (e.g. Vogel, 1996); however, the actual practice of L-curve analysis in the glaciological literature is inconsistent at best, and many modelers who chose to forgo L-curve analysis do not replace it with any other method of calibrating the correct regularization level.

In this paper, we first give a brief review of the ways in which regularization and L-curve analysis have been treated in glaciological inverse models, and then we give a tutorial on how to regularize and perform L-curve analyses for glaciological

inverse problems. Using an inversion of the Filchner-Ronne catchment in Antarctica as our test bed, we demonstrate how to scale misfit components, compute L-curve curvature, and select both the optimal λ value as well as a range of acceptable λ values bracketing the corner. We explore how the spatial and spectral characteristics of the solution change as λ is varied within this acceptable range. We perform a set of experiments testing the influence of mesh resolution, effective pressure, sliding nonlinearity, and flow approximation. We generate a new independent L-curve for each change in our experimental setup. Finally, we combine our various inversion results into a single best weighted-average estimate of the basal drag within our domain. We conclude with a discussion about what this approach to glaciological inverse problems enables us to learn about basal sliding and with recommended best practices for future modelers.

2 Review

The original work of MacAyeal (1993) was a pioneering addition to the toolbox of glaciological modeling at the time. However, that work included no explicit regularization at all (MacAyeal, 1993, Eq. 17), instead avoiding overfitting by parameterizing the drag coefficient as a sum of Fourier basis functions and truncating the sum at low order. Indeed, in the synthetic example used in that paper, the "true" basal drag was given by only two Fourier terms, and the inversion consisted of an optimization to recover those same two coefficients. Subsequent papers using the MacAyeal method with real velocity data instead of synthetic examples obviously included a much greater number of Fourier coefficients, but they still relied on truncation of the series to prevent over-fitting rather than including explicit regularization within the cost function (Joughin et al., 2004, 2006, 2009). The question of where exactly to truncate this series was not investigated in detail other than to note that the governing equations were not accurate below wavelengths of a few ice thicknesses.

Later, Arthern and Gudmundsson (2010) introduced the method of solving the inverse problem by simultaneously solving Neumann and Dirichlet versions of the stress balance problem (ie, solving one forward problem where the surface is traction-free and another forward problem where the surface has a Dirichlet condition given by the observed velocity, and minimizing the difference between these two solutions). This method produced a major advance in ice sheet inversions, allowing the Full Stokes (FS) equations to be inverted without the need to derive the adjoint (an equation for the gradient of the cost function with respect to the inversion target field). In addition, and unlike the models in the direct MacAyeal lineage (Joughin et al., 2004, 2006, 2009), they did not rely on a series expansion representation of the drag coefficient in their derivation, removing the need for arbitrary truncation and instead representing drag coefficient on the same numerical mesh as the rest of their problem. However, they did not put any explicit regularization in their cost function either (Arthern and Gudmundsson, 2010, Eq. 3) and, consequently, they found that their algorithm produced erroneous spatial structure when given noisy velocity data (Arthern and Gudmundsson, 2010, Fig. 3). To solve this, they settled on a "heuristic" method of implicit regularization achieved by stopping their algorithm at a small number of iterations before the erroneous spatial structure had a chance to grow. It has long been known that certain types of iterative solvers can have a regularizing effect if they are stopped early (Hansen and O'Leary, 1993), and Habermann et al. (2012) developed a similar method of implicit regularization for glaciological inversions of synthetic data.

However, it is notable that when those authors later applied their methods to the inversion of real data, they employed explicit regularization (Habermann et al., 2013, Eq. 4), as did later users of the Neumann-Dirichlet approach (e.g. Shapero et al., 2016).

Indeed, it is rare to find inversions of real data and real glacier geometries performed without either explicit regularization or series truncation. However, that regularization is unfortunately not always discussed in detail, and often the effect of regularization is not explored and the choice of regularization is not justified. For instance, Morlighem et al. (2010) studied spatial patterns in basal drag under Pine Island Glacier determined from inverse models, yet they gave no details on regularization other than a single sentence (in their Section 2.4) confirming that it was in fact present. Thus it is difficult to evaluate whether or not conclusions drawn about the detailed structure of basal drag near the grounding line may have been influenced by the choice of regularization. In other papers, regularization is performed and explained, but L-curve analysis is not performed and the particular level of regularization is not always justified. For instance, Sergienko and Hindmarsh (2013) and Sergienko et al. (2014) explicitly showed how they regularized their inversion, and they performed sensitivity experiments with several values of λ for both real and synthetic examples. However, they explicitly chose not to present an L-curve analysis, citing the possibility of L-curve non-convergence (Vogel, 1996) as their reason.

When an L-curve analysis is performed in glaciological inversions, the presentation of the L-curve can be highly variable. Hansen and O'Leary (1993) explicitly recommended that L-curves should be presented on log-log axes, and that the curvature computation used to select the corner should also be performed in log-log space. However, while glaciologists have sometimes presented L-curves on log-log axes (e.g. Morlighem et al., 2013, Fig. 1), they have also employed linear-linear axes (Shapero et al. (2016, Fig. 1); Habermann et al. (2013, Fig. 5)) or even mixed log-linear axes (Gillet-Chaulet et al., 2012, Fig. 3). The variable presentation styles for different L-curves is potentially problematic, as L-curve analysis is explicitly predicated on the concept of "shape", and many of the above authors chose their optimal λ values based on a visual assessment of corner position rather than a quantitative measure of curvature. Finally, while L-curves in the glaciological literature have usually been monotonic (Morlighem et al., 2013; Habermann et al., 2013; Shapero et al., 2016), that has not always been the case (Gillet-Chaulet et al., 2012).

We do not claim that this review of regularization and L-curve analysis in the glaciological literature is exhaustive. Our intention here was merely to illustrate the spectrum of ways in which these issues have been treated within glaciological inversions. Inversions are a vital source of boundary conditions for projections of ice sheet dynamics in a changing climate (Seroussi et al., 2019), and inversions also provide a vital window into the conditions and processes at the ice sheet bed, an environment that is difficult to observe directly. L-curve analysis is an important method for ensuring that inversions are well-calibrated between under-fitting and over-fitting the data, yet the treatment of L-curve analysis and the visual presentation of L-curves has been highly variable within the glaciological literature. Even the basic question of whether glaciological inverse problems have a convergent L-curve remains unanswered.

3 Methods

3.1 Experimental Design

In addition to regularization, we also test the sensitivity of our results to element size, flow equation, sliding nonlinearity, and the use of effective pressure. We test the influence of regularization through L-curve analysis, and we created separate independent L-curves whenever we tested anything else in our experimental setup: for instance, when we varied element size, we created a separate L-curve for each mesh that we tested. We organized our experimental setup around a reference case using the Shelfy-Stream Approximation (SSA) in the highest-resolution mesh with a linear Weertman sliding law. We explored the influence of regularization itself by analyzing the spatial and spectral characteristics of the reference case in detail (Sections 4.2 and 4.3). In order to test the dependence of regularization on element size, we produced L-curves using linear Weertman inversions for a series of 10 meshes (Section 4.4). We tested the influence of effective pressure N by producing L-curves with Budd sliding using 3 candidate N fields in the highest-resolution mesh (Section 4.5). We tested the influence of sliding nonlinearity by producing L-curves for both the nonlinear Weertman sliding law and nonlinear Budd sliding with the highest-performing N field (Section 4.6). We performed experiments comparing SSA inversions with Higher Order (HO) inversions by producing L-curves for HO using linear Weertman sliding in Meshes # 4, 6, and 8, along with one additional HO L-curve in Mesh #6 using constant rheology instead of variable rheology (Section 4.7). All in all, the experiments presented in this paper comprise 21 independent L-curves (10x linear Weertman SSA, 2x nonlinear Weertman SSA, 3x linear Budd SSA, 2x nonlinear Budd SSA, and 4x linear Weertman HO), each composed of 25 separate inversions, for a total of 525 individual inversions. Additionally, we did a supplemental experiment where we converted the coefficient from a linear Weertman inversion into a nonlinear sliding law, and we compared the resulting converted L-curve to the L-curve computed from inversions performed directly with the nonlinear law (Section B1).

3.2 Setting

We model a domain covering the catchments of West and East Antarctica that feed into the Filchner-Ronne Ice Shelf (Fig. 1). This domain has a wide variety of glaciological settings, including both slow-flowing regions and fast-flowing ice streams, mountainous subglacial topography and deep subglacial basins, topographically confined outlet glaciers and relatively unconfined glaciers, along with a large floating ice shelf containing several ice rises and rumples. We choose to use a real setting based on real data for this study rather than constructing a synthetic example for several reasons: 1) we are interested in learning about the spatial and spectral characteristics of the true basal drag (Sections 4.2 and 4.3); 2) we are interested in evaluating how good different effective pressure fields (Section 4.5) and sliding exponents (Section 4.6) are at parameterizing reality; 3) we are interested in exploring the convergence behavior of L-curve analysis with respect to model resolution (Section 4.4), and this is known to depend on the roughness characteristics of the true solution (Vogel, 1996); and 4) we wish to conclude by offering recommendations for future glaciological modelers (Section 6), and recommendations based on simplified synthetic examples may founder when tested against the complexities that arise when applying models to real datasets. We take surface (Fig. 1a) and bed (Fig. 1b) elevations, along with ice thickness and our ocean/shelf/sheet/rock mask from Morlighem et al. (2020). Ice

surface velocity observations needed for our inversion (Fig. 1c) are taken from a combined Landsat 8 and TerraSAR-X product created by Hofstede et al. (2021), merged with MEaSUREsv2 (Rignot et al., 2017). We also can use the observed velocity and geometry to compute the driving stress projected onto the flow direction (Fig. 1d) which we use to estimate potential energy dissipation (Sections A2 and A1) and to compute our first guess of drag coefficient (Section 3.5).

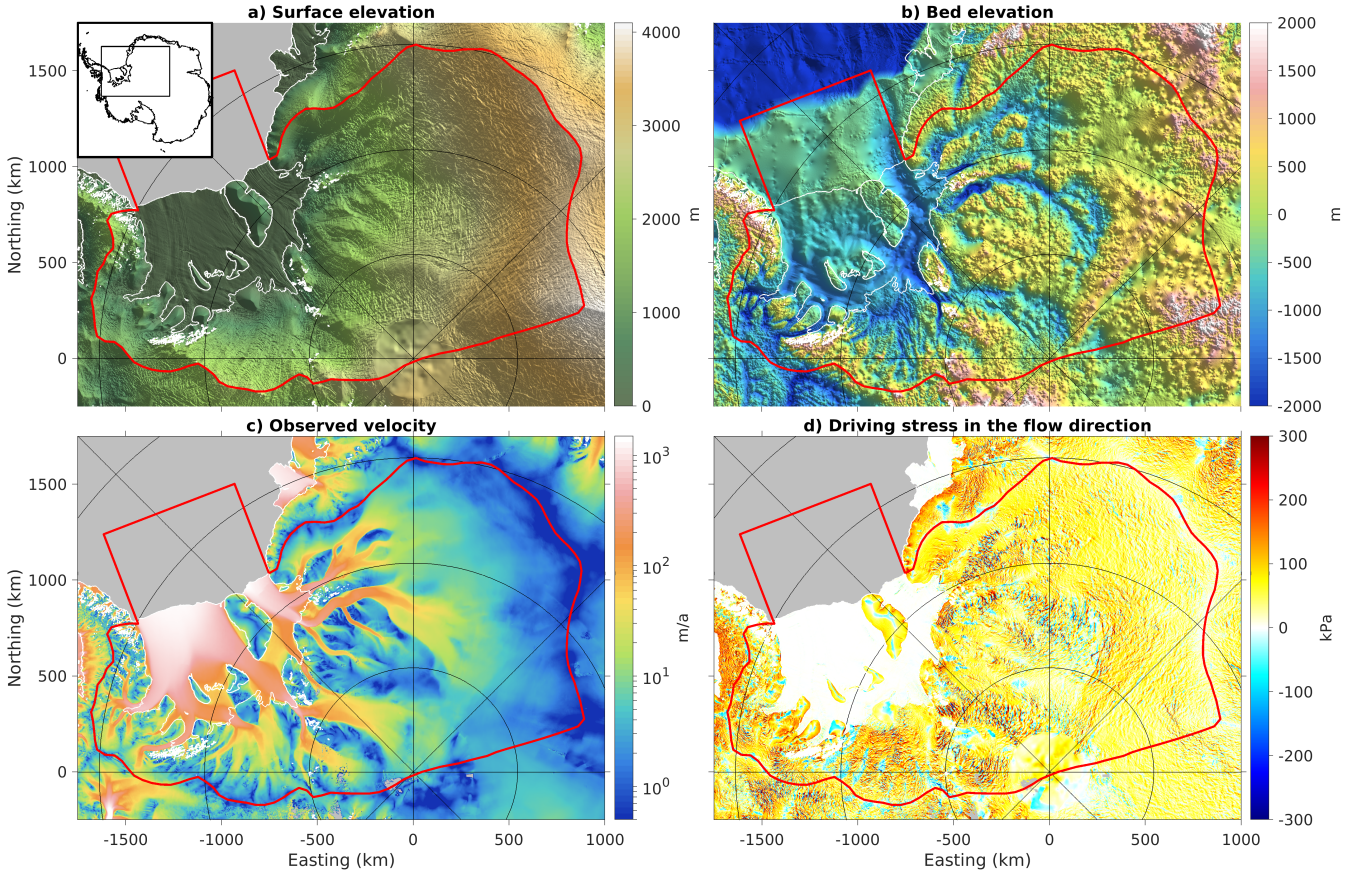

Figure 1. Geographic setting of our study area. a) Surface elevation from Morlighem et al. (2020), with inset showing map location in Antarctica. b) Bed elevation from the same source. c) Surface velocity observations from Hofstede et al. (2021) and Rignot et al. (2017). d) Driving stress in the flow direction computed from ice geometry and observed velocity vectors. Red line in all maps shows outline of model domain. The domain extends beyond the calving front to facilitate transient modeling of ice front motion, which we do not discuss in this paper. White lines show ice edge and grounding line. Graticule uses lines at 5° in latitude and 45° in longitude. Plots (a) and (b) employ hillshading to emphasize structure and texture, using two perpendicular light sources from grid north and grid east.

3.3 Ice Sheet Model

We use the Ice-Sheet and Sea-level System Model (ISSM; Larour et al. (2012)) to simulate the catchment of the Filchner-Ronne Ice Shelf. We use the Shelfy-Stream Approximation (SSA, also called Shallow Shelf Approximation; MacAyeal (1989)). SSA

reduces the 3D problem of ice flow to a vertically integrated 2D problem in the map-view plane, thus neglecting all but the xx, yy, and xy terms in the stress and strain-rate tensors. SSA is most valid in fast-flowing ice streams and outlet glaciers, where the ice is flowing primarily by basal sliding. The SSA equations are also self-adjoint for linear sliding laws (MacAyeal, 1993), and they were the first version of the ice sheet stress balance equations to be used in an inversion (MacAyeal, 1993). The SSA equations are,

$$\nabla \cdot (H\bar{\mu}(\nabla \boldsymbol{u} + \nabla^{\mathrm{T}}\boldsymbol{u} + 2(\nabla \cdot \boldsymbol{u})\mathbf{I})) + \boldsymbol{\tau}_{\mathrm{d}} - \boldsymbol{\tau}_{\mathrm{b}} = 0, \tag{1}$$

where H is ice thickness, $\bar{\mu}$ is the vertically averaged viscosity, $\boldsymbol{u}$ is the horizontal ice velocity vector, $\nabla()$ is the gradient operator in the horizontal plane and $\nabla^{\mathrm{T}}()$ is its transpose, $\mathbf{I}$ is the identity matrix, $\nabla \cdot ()$ is the divergence operator in the horizontal plane, $\boldsymbol{\tau}_{\mathrm{d}}$ is the gravitational driving stress, and $\boldsymbol{\tau}_{\mathrm{b}}$ is the basal stress, which we infer indirectly using our inversion. The driving stress is computed from the ice sheet geometry by, $\boldsymbol{\tau}_{\mathrm{d}} = -\rho_{\mathrm{i}} g H \nabla z_s$, where ρ_i is the density of ice, g is the acceleration due to gravity, and z_s is the ice sheet surface elevation. At the inland lateral boundaries of the domain we use Dirichlet conditions given by the observed velocity. The effective viscosity is computed from the nonlinear rheology of ice (Glen, 1953) using,

$$\bar{\mu} = \frac{1}{2}\bar{B}\dot{\epsilon}_0^{\frac{1-n}{n}}, \tag{2}$$

where $\bar{B}$ is the vertically averaged rheological stiffness parameter, $\dot{\epsilon}_0$ is the effective strain rate, and $n = 3$ is the rheological exponent. We use the SSA equations for the majority of our experiments, but we also ran several experiments using the 3D Higher Order (HO) equations (Blatter, 1995; Pattyn, 2002) to compare the effect of using a 3D model instead of a 2D one. We do not list those equations here for brevity. We compute the rheological stiffness needed for both sets of equations using a 1D thermal model, described in Section A1. We solve the equations for both models on spatially refined unstructured meshes described in Section A2. We used a multi-wavelength interpolation procedure to interpolate gridded data products onto our meshes in a way that prevented aliasing where the mesh is coarse while still preserving high-resolution spatial structure where the mesh is fine (Section A3).

3.4 Sliding Law

For our basic model, we use a Weertman-type sliding law (Weertman, 1957),

$$\boldsymbol{\tau}_{\mathrm{b}} = C|\boldsymbol{u}_{\mathrm{b}}|^{\frac{1-m}{m}}\boldsymbol{u}_b, \tag{3}$$

where $\boldsymbol{\tau}_{\mathrm{b}}$ is the basal shear stress vector, C is a spatially variable drag coefficient, $\boldsymbol{u}_{\mathrm{b}}$ is the basal sliding velocity vector, and m is the slip exponent. For most of our experiments we used linear Weertman ($m = 1$), however, we also ran tests with

nonlinear Weertman sliding using $m = 3$ and $m = 5$, the results of which we present in Section 4.6. When testing the effect of sliding laws with effective pressure, we use a Budd-type sliding law (Budd et al., 1979),

$$\boldsymbol{\tau}_{\mathrm{b}} = CN|\boldsymbol{u}_{\mathrm{b}}|^{\frac{1-m}{m}}\boldsymbol{u}_b \tag{4}$$

where $N = P_{\mathrm{i}} - P_{\mathrm{w}}$ is the effective pressure at the ice base, and P_{i} and P_{w} are ice and water pressures, respectively. Budd sliding is not the only sliding law to incorporate N, nor is it necessarily the most advanced law to do so (e.g. Schoof, 2007; Gagliardini et al., 2007; Tsai and Gudmundsson, 2015). However, Budd sliding continues to be used in recent glaciological publications (e.g. Brondex et al., 2019; Barnes and Gudmundsson, 2022; Kazmierczak et al., 2022), and a Budd law has the advantage of being directly comparable to a Weertman law with the same stress exponent, allowing us to test the influence of N with all else held constant. We test three different sources for N:

$$N_{\mathrm{op}} = \quad \rho_{\mathrm{i}} g H - \rho_{\mathrm{sw}} g(-z_b), \tag{5}$$

$$N_{\mathrm{opc}} = \quad \rho_{\mathrm{i}} g H - \max(0, \rho_{\mathrm{sw}} g(-z_b), \tag{6}$$

$$N_{\mathrm{CUAS}} = \quad N_{\mathrm{CUAS}}(x, y), \tag{7}$$

where N_{op} is N assuming a perfect hydraulic connection to the ocean, and allowing negative water pressures; N_{opc} is N determined from ocean pressure with a strict cutoff at $z_b = 0$; and N_{CUAS} is N taken from the hydrology model CUAS-MPI. CUAS-MPI is an equivalent thin-layer hydrology model using confined-unconfined aquifer equations and parameterizations for efficient drainage and the opening of cavity space by ice sliding (Beyer et al., 2018; Fischler et al., 2023). We force CUAS-MPI using the melt rate estimated by the same 1D thermal model used to estimate ice rheology (Fig. A1e). We run CUAS-MPI on a $1\,\mathrm{km}$ regular grid and then interpolate the results onto our ISSM mesh. N_{opc} is a commonly used approximation in the glaciological literature, and we felt it was important to test it here. However, N_{opc} contains a sharp discontinuity in gradient around the $z_b = 0$ contour, so our motivation for allowing negative water pressures in N_{op} is to eliminate that discontinuity and test for the optimal regularization when N was a simple linear function of z_b. In both approximations N is strictly positive throughout our domain.

We show a comparison of the three N fields in Figure 2. N_{op} is the smoothest, as it is a simple linear function of ice thickness and bed elevation. It also has the most range, as inland regions with a very high ice surface elevation and bed elevation above zero have a large ice overburden pressure combined with a negative water pressure, producing very high N. N_{opc} has a broadly similar large-scale structure to N_{op}, but with more visible short-wavelength structure introduced by the cutoff in water pressure at $z_b = 0$. The range of N_{opc} is also less than the range of N_{op}, as the high-elevation inland regions are no longer allowed to have negative water pressure and thus have lower N. N_{CUAS} has the lowest range of the three, with a maximum of about $10\,\mathrm{MPa}$, or a factor of 3–4 less than the other two. The spatial structure of N_{CUAS} is more complex than N_{op} but also appears more ice dynamically reasonable, at least under visual inspection, than N_{opc}. The spatial structure of N_{CUAS} is a close visual

match to the overall structure of the ice-sheet velocity field (Fig. 2c, c.f. Fig. 1c), likely as a result of the fact that we used estimated shear heating to produce the melt rate forcing for CUAS-MPI (Fig A1e).

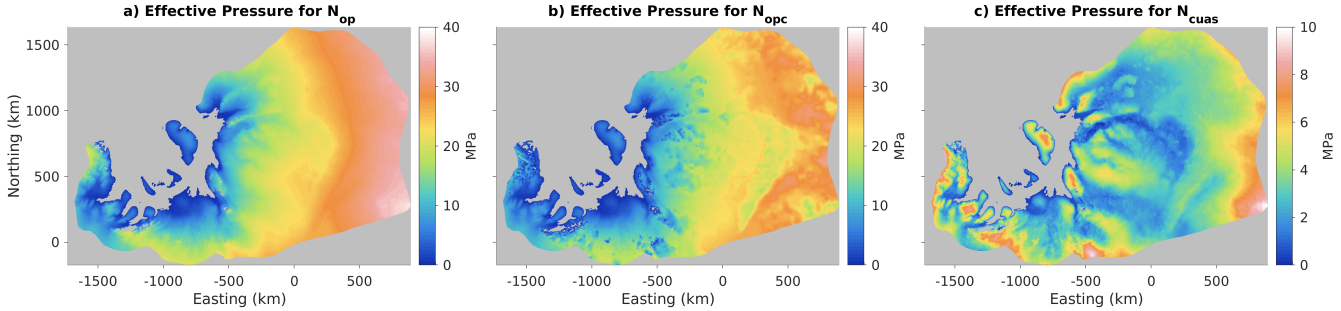

Figure 2. Comparative maps of effective pressure N. a) N_{op}, or N computed assuming a perfect connection to the ocean and allowing negative water pressures. b) N_{opc}, as N_{op} but without allowing negative water pressures. c) N computed by the hydrology model CUAS. Note change in color scale in panel (c).

3.5 Cost Function

We use a single observational misfit and a single regularization term in our cost function. Our observational misfit is the L2 norm of the absolute velocity misfit,

$$J_{\mathrm{obs,raw}} = \int_{\Omega} \frac{1}{2} \left((u_x - u_{x,obs})^2 + (u_y - u_{y,obs})^2 \right) d\Omega, \tag{8}$$

where $J_{\mathrm{obs,raw}}$ is the raw observational component of the cost function, Ω is the map-view model domain, and $\boldsymbol{u} = (u_x, u_y)$ are the two components of the horizontal velocity vector. Our regularization misfit is the L2 norm of the magnitude of the gradient of the friction coefficient,

$$J_{\mathrm{reg,raw}} = \int_{\Omega} \frac{1}{2} |\nabla k|^2 d\Omega, \tag{9}$$

where $J_{\mathrm{reg,raw}}$ is the raw regularization cost and $k = \sqrt{C}$ is the internal ISSM friction coefficient. There is no mathematical or physical reason why the leading constant of the sliding law should be squared; MacAyeal (1993) squared the constant to ensure it remained positive, but this same goal can be achieved by simply setting a lower limit within the inversion code. We therefore exclusively analyze results for C, not k, when analyzing our inversions. However, we do expect C to vary by many orders of magnitude, and the regularization term may therefore under-penalize spatial structure in low-drag areas as compared to high-drag areas. Taking the square root of the drag coefficient before computing the regularization term halves the dynamic range, and thus we expect this problem to be less of an issue for k than for C.

Before combining the data and regularization cost terms, it is useful to scale them according to estimates of their characteristic magnitude. Scaling the cost function terms is valuable for two reasons: 1) it allows us to easily identify the region of

parameter space that we need to search in our L-curve analysis, and, 2) it ensures that the regularization parameter is both unitless and readily interpretable in terms of relative weight placed on the two component terms. Without scaling, the regularization parameter would have obscure units, it would be difficult to identify the region of parameter space near the corner of the L-curve to search for the optimal regularization, and there would be no basis for judging whether a given level of regularization was "large" or "small".

The characteristic scale of the data term is readily given by the variance of the velocity observations themselves,

$$S_{\mathrm{obs}} = \int_{\Omega} |\boldsymbol{u}_{\mathrm{obs}}|^2 d\Omega = A\sigma_{\mathrm{obs}}^2, \tag{10}$$

where S_{obs} is the characteristic scale of $J_{\mathrm{obs,raw}}$, $|\boldsymbol{u}_{\mathrm{obs}}| = \sqrt{u_{\mathrm{x,obs}}^2 + u_{\mathrm{y,obs}}^2}$ is the magnitude of the observed velocity vector, A is the area of the grounded domain, and σ_{obs} is the root-mean-squared (RMS) variability of the observed velocity magnitude.

Computing the characteristic scale of the regularization term is more complex than computing the characteristic scale of the data term, because the magnitude of the regularization term depends on the gradient of the unknown drag coefficient. We computed the characteristic scale of the regularization term based on the mean-square gradient of a reference sinusoid with a wavelength given by the mean ice thickness and an amplitude given by an estimate of the standard deviation of the drag coefficient. We estimated the standard deviation of the drag coefficient by computing a first guess based on the ratio of driving stress and observed velocity,

$$k_{\mathrm{guess}}^2 = \frac{\max(0, \tau_{\mathrm{d,flow}})}{N|\boldsymbol{u}_{\mathrm{d}}|^{1/m}}, \tag{11}$$

where k_{guess} is the guess of k, and $\tau_{\mathrm{d,flow}} = -\rho_{\mathrm{i}} g H \nabla(S) \cdot \boldsymbol{u}_d / |\boldsymbol{u}_d|$ is the driving stress projected onto the flow direction (Fig. 1d). For Weertman sliding, we set $N = 1$ in the above equation. For Budd sliding, we use a minimum N value of $100\,\mathrm{Pa}$ to prevent division by zero, and for both sliding laws we use a minimum velocity value of $10\,\mathrm{cm\,a^{-1}}$ for the same purpose. We then compute σ_k, the standard deviation of k_{guess}, and we use it to define a reference sinusoid, $k = \sigma_k \sin(2\pi x/\bar{H})$. The gradient of this sinusoid is, $\nabla(k) = \frac{2\pi\sigma_k}{\bar{H}} \cos(2\pi x/\bar{H})$, and the mean-square amplitude of the gradient is $2(\pi\sigma_k/\bar{H})^2$. Plugging the mean-square amplitude of this perturbation into Eq. 9 and assuming an area of integration equal to the grounded domain yields the scale estimate,

$$S_{\mathrm{reg}} = A\left(\frac{\pi\sigma_k}{\bar{H}}\right)^2, \tag{12}$$

where S_{reg} is that characteristic scale of the regularization term $J_{\mathrm{reg,raw}}$.

We then use the characteristic scales to normalize both components of the cost function,

$$J_{\mathrm{obs}} = \frac{J_{\mathrm{obs,raw}}}{S_{\mathrm{obs}}}, \tag{13}$$

$$J_{\mathrm{reg}} = \frac{J_{\mathrm{reg,raw}}}{S_{\mathrm{reg}}}, \tag{14}$$

where $J_{\mathrm{obs}}, J_{\mathrm{reg}}$ are the scaled dimensionless cost function terms. The final cost function is a linear combination of the two scaled terms,

$$J = J_{\mathrm{obs}} + \lambda J_{\mathrm{reg}}, \tag{15}$$

where J is the total cost function that is minimized by the inversion and λ is a dimensionless Tikhonov regularization parameter (Tikhonov and Arsenin, 1977). We determine the optimal λ value through an L-curve analysis. We tested 25 logarithmically spaced samples in the range $10^{-3} \leq \lambda \leq 10^{3}$. For each sample, we run the inverse model to convergence and record J_{obs} and J_{reg}. We then used a curve-fitting procedure to fit a smooth curve to the discrete samples of J_{obs} and J_{reg}. We computed the second derivative of the smoothed curve in log-log space and used this as the basis for selecting λ_{best}, the best fit corner value of the L-curve, along with minimum and maximum acceptable values, λ_{min} and λ_{max}, that bracket the range of the corner of the L-curve. We discuss the curve-fitting and λ estimation procedure in Section A5.

4 Results

4.1 L-curves

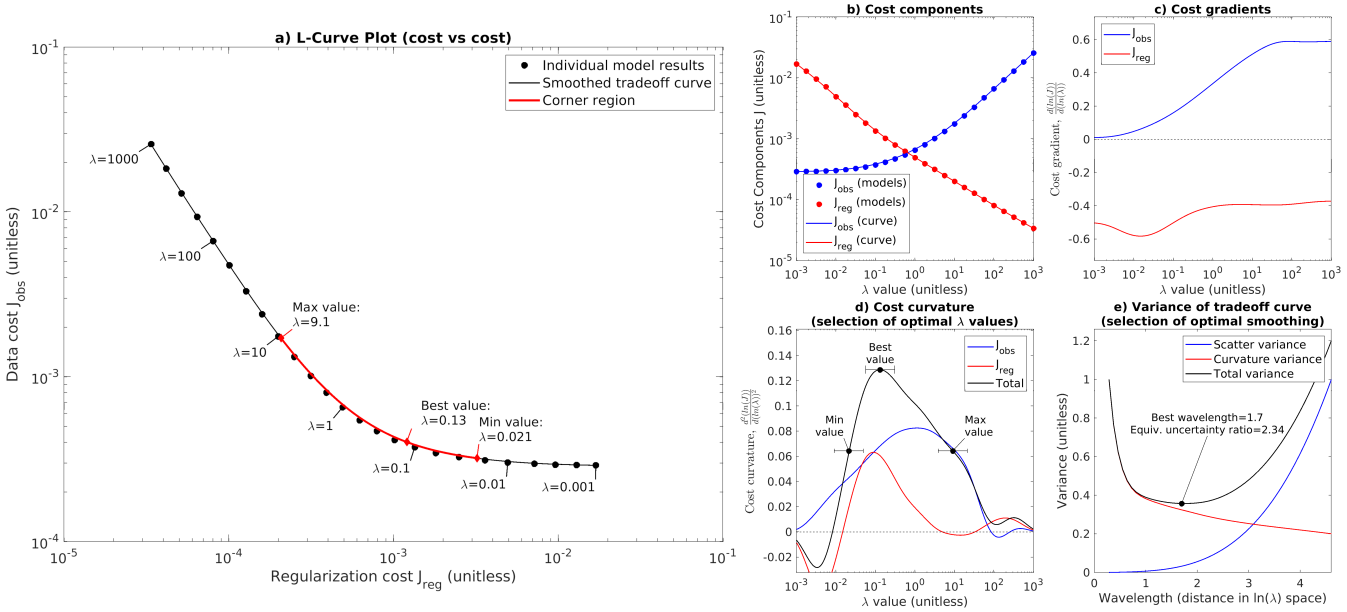

Figure 3. Example L-curve for mesh #1. Large plot (a) shows the L-curve, a log/log plot of J_{reg} vs J_{obs}. Black circles show individual inversions, black line shows the smoothed tradeoff curve fitted to the individual models, and red region highlights the corner of the L-curve. Note that we have scaled the axes in plot (a) so that one order of magnitude in J_{reg} takes up precisely as much space as one order of magnitude in J_{obs}. b) Plot of both misfit components (J_{reg} and J_{obs}) against λ. c) Cost gradient as a function of λ. d) Cost curvature as a function of λ, with selection of min, best, and max λ values annotated. e) Selection of optimal smoothing wavelength, showing scatter and curvature variance (V_{s} and V_{c}), selection of optimal smoothing wavelength, and equivalent uncertainty ratio for λ on a linear scale.

The L-curve for our reference case of mesh #1 and linear Weertman sliding is shown in Figure 3a. Our model results form a well-defined "L" shape in ($\ln(J_{\mathrm{reg}})$,$\ln(J_{\mathrm{obs}})$) space, and the visual impression of an "L" shape can be trusted in this figure

because we have taken care to scale the plot so that a decade of dynamic range takes the same amount of space on both axes. Both components of the misfit are monotonic functions of λ, and our smoothed tradeoff curve runs very close to the original model points (Fig. 3a,b), giving us confidence in our curve-fitting procedure. Previously published inversions have sometimes been vexed by problems such as deviations from L-curve monotonicity or outlier models that fall far from the smooth L-curve (e.g. Gillet-Chaulet et al., 2012, Fig. 3); while we cannot be sure of what caused these results for other authors, during development of our model setup, we found that such issues were often an indication of lower-level numerical problems such as suboptimal choice of solvers, overly lax convergence tolerances, or overly complex mask topologies. Once those issues were fixed, we found that smooth monotonic results such as those in Fig. 3 were typical for all of our inversions.

Far from the corner region, both limbs of our L-curve tend towards straight lines, but for the large-λ limb this line is sloping, indicating a power-law relationship between J_{reg} and J_{obs}, while for the small-λ limb this line is horizontal, indicating an asymptotic approach to a best achievable J_{obs} (Fig. 3a). The straight-line nature of the limbs can be confirmed by looking at the gradient of the cost terms: both terms approach constant gradient for large λ, while J_{obs} smoothly approaches zero gradient for small λ (Fig. 3c). The low-λ limb corresponds to the "flat" part of the L-curve identified by Hansen and O'Leary (1993, Section 4.1), while the high-λ limb corresponds to the "vertical" part, although for linear sliding the slope of this limb is far from vertical (for nonlinear sliding the slope of this limb increases, as we show in Section 4.6). The power-law nature of the large-λ limb may potentially explain the results found by Habermann et al. (2013). Those authors found that their L-curve lacked a corner when plotted on log-log axes, motivating them to use linear axes instead (Habermann et al., 2013, Section 3.4). While we cannot be certain what caused them to find this result, our own results suggest that they may have been testing λ values in the large-λ limb. Normalizing cost function terms by an estimate of their characteristic scale before the L-curve analysis is performed helps to prevent this possibility by ensuring that the L-curve analysis covers a range of λ that includes both limbs.

The best-λ value computed from curvature falls roughly at the visual corner of the plot (Fig. 3a), while the minimum and maximum acceptable λ values bracket the transition between the corner regime and the straight-line limbs. In this example, the smoothing applied to generate the tradeoff curve corresponds to a wavelength of 1.7 in $\ln(\lambda)$ space, or an uncertainty of $\pm$ a factor of about 2 in linear space (Fig. 3e). The uncertainty in λ was generally in the range of $\pm$ a factor of 2–3 for all of our experiments; given that the values of λ we tested differed by a factor of 1.8, this implies that we were generally able to identify λ_{best} to within 2–3 adjacent λ samples. While the position of λ_{best} is defined by peak curvature, we use a heuristic threshold of $1/2$ peak curvature to define the limits of the corner region λ_{min} and λ_{max} (Section A5). The dropoff in curvature below λ_{best} is quite steep, so we expect the arbitrary choice of this threshold to have little influence on λ_{min}, but the dropoff in curvature on the high end is more gradual (Fig 3d), so the exact position of λ_{max} should be regarded as more uncertain.

While, for simplicity's sake, we are only showing a single representative L-curve analysis in Figure 3, the features we described above are generally consistent for all of the L-curves that we produced, although the exact position of the curves (and the corresponding position of the min, max, and best λ values) systematically shifted towards smaller λ as mesh resolution increased. In addition, the minimum achievable J_{obs} in the small-λ limit dropped as mesh resolution increased. We discuss the convergence of our results as a function of mesh resolution in greater detail in Section 4.4. In Section 4.5, we discuss how our

L-curve results differ for Budd sliding laws with different N fields, and in Section 4.6 we discuss how our L-curves differ for nonlinear sliding. But first we discuss the spatial and spectral characteristics of our results for our linear Weertman reference case as a function of regularization.

4.2 Spatial Analysis

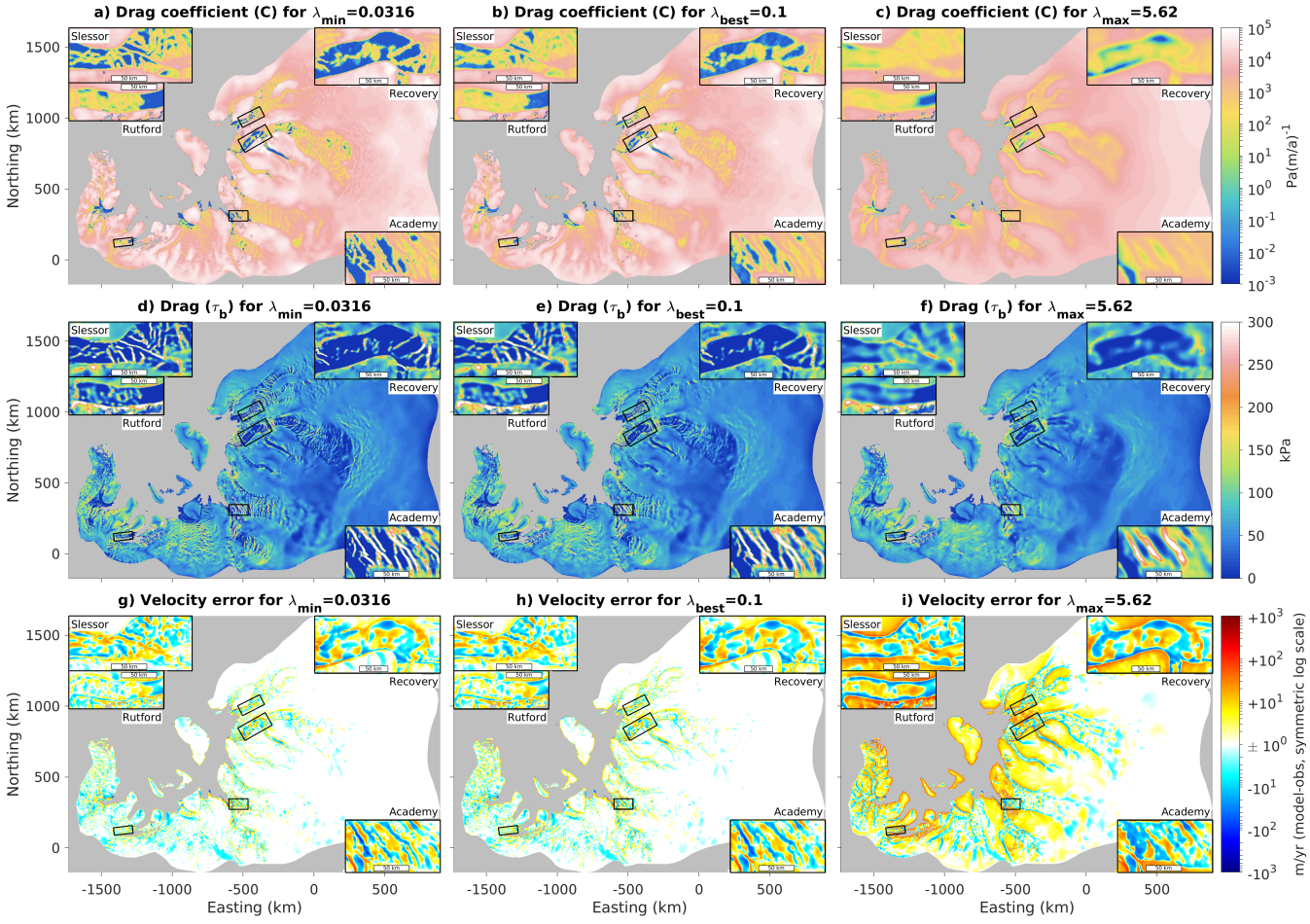

Figure 4. Comparison of inversion results for the highest-resolution mesh as a function of λ. Top row (a,b,c) shows drag coefficient C on a logarithmic scale, middle row (d,e,f) displays drag $\boldsymbol{\tau}_b$ on a linear scale, and bottom row (g,h,i) shows velocity misfit on a symmetric logarithmic scale (ie, these plots use separate logarithmic scales for positive and negative errors). Left column (a,d,g) shows minimum acceptable λ, middle column (b,e,h) shows best λ, and right column (c,f,i) shows maximum acceptable λ. All plots show zoom-in panels for Slessor, Recovery, and Academy Glaciers, along with Rutford Ice Stream.

In Figure 4, we show our inversion results for our reference case for minimum, maximum, and best λ values. In general, C values are orders of magnitude lower in fast-flowing ice streams and outlet glaciers than in slow-flowing regions of the ice

sheet (Fig. 4a,b,c). The distribution of $\boldsymbol{\tau}_\mathrm{b}$ is more complex, however, with some of the highest drag values occurring in narrow ribs or sticky spots within fast-flowing ice streams (Fig. 4d,e,f). As expected, there is a clear trend towards greater smoothness in C and $\boldsymbol{\tau}_\mathrm{b}$ for larger λ values, and greater small-scale structure in those fields for smaller λ values. Unsurprisingly, velocity misfit declines for smaller λ values (Fig. 4g,h,i). Absolute misfits tend to be higher in faster-flowing regions for all λ values, reflecting the larger velocity magnitude in those regions. However, the λ_max model also contains a large amount of positive misfit (ie, model velocity too high) in the slow-flowing regions in between fast-flowing outlet glaciers and in the islands within the floating shelf (Fig. 4i), likely because the larger regularization prevented the inversion from fully capturing the sharp rise in C across the glacier shear margins and grounding lines.

We also show details of the inverted structure within several fast-flowing ice streams and outlet glaciers. In response to the Sergienko and Hindmarsh (2013) hypothesis of regular rib-like patterns in basal drag, we choose these detail regions to cover a spectrum from very "rib-like" appearance to very "non-rib-like" appearance: Academy Glacier just upstream of its confluence with Foundation Ice Stream has some of the strongest ribs in our inversion, followed closely by downstream Slessor Glacier, while downstream Recovery Glacier has a mix of rib-like structure and non-rib sticky spots, and downstream Rutford Ice Stream has very little visible ribbing. Unsurprisingly, the ribs are strongest in the inversion with minimum acceptable λ (Fig. 4a,d) and weakest in the inversion with maximum acceptable λ (Fig. 4c,f), with the best λ inversion being intermediate. The same ribs appear in the λ_min and λ_best inversions, albeit wider in the latter; while in the λ_max inversion, adjacent ribs have begun to combine. In Academy and Slessor glaciers, almost all of the basal resistance is contained within the ribs, while in Recovery and Rutford glaciers ice flow is resisted by a combination of more rounded sticky spots in the middle of the fast-flow region along with stress transmission to higher-drag margins (Fig. 4d,e). We explore the spectral characteristics of these four detail regions next.

4.3 Spectral Analysis

For each region, we first interpolated $\boldsymbol{\tau}_\mathrm{b}$ from the model mesh onto a regular grid aligned with the rectangular detail box. In order to minimize spectral artifacts produced by the assumption of periodic boundary conditions in the FFT algorithm, we expanded the grid by 50% in each direction and filled the buffer zone by a smooth extrapolation that approached a constant value around the edges of the expanded grid. We then took the two-dimensional discrete Fourier transform to produce 2D spectra in (k_x, k_y) space, and transformed these into 1D spectra using $k_\mathrm{mag} = \sqrt{k_x^2 + k_y^2}$, followed by integration of spectral power around circles of constant k_mag. We also computed the spectral coherence between basal drag $\boldsymbol{\tau}_\mathrm{b}$ and driving stress τ_d by taking the dot product between the two spectra in (k_x, k_y) space and then integrating in a similar manner.

We show the results of this spectral analysis of basal drag for the four detail regions in Fig. 5. We have organized the regions in order of visual "ribiness", from Academy (Fig. 5a,e) to Rutford (Fig. 5d,h). All four glaciers have roughly constant spectral power in the λ_min and λ_best inversions from wavelengths of about 30 H to 5 H. Below 5 H, spectral power decays rapidly, while for the λ_max inversion, the decay begins just above 10 H. The three most "ribby" glaciers have large differences in spectral power as a function of λ at short wavelengths, especially when comparing λ_max to the other two (Fig. 5a–c), while Rutford has comparatively small differences as a function of λ (Fig. 5d). For all glaciers, the largest spread in spectral power between

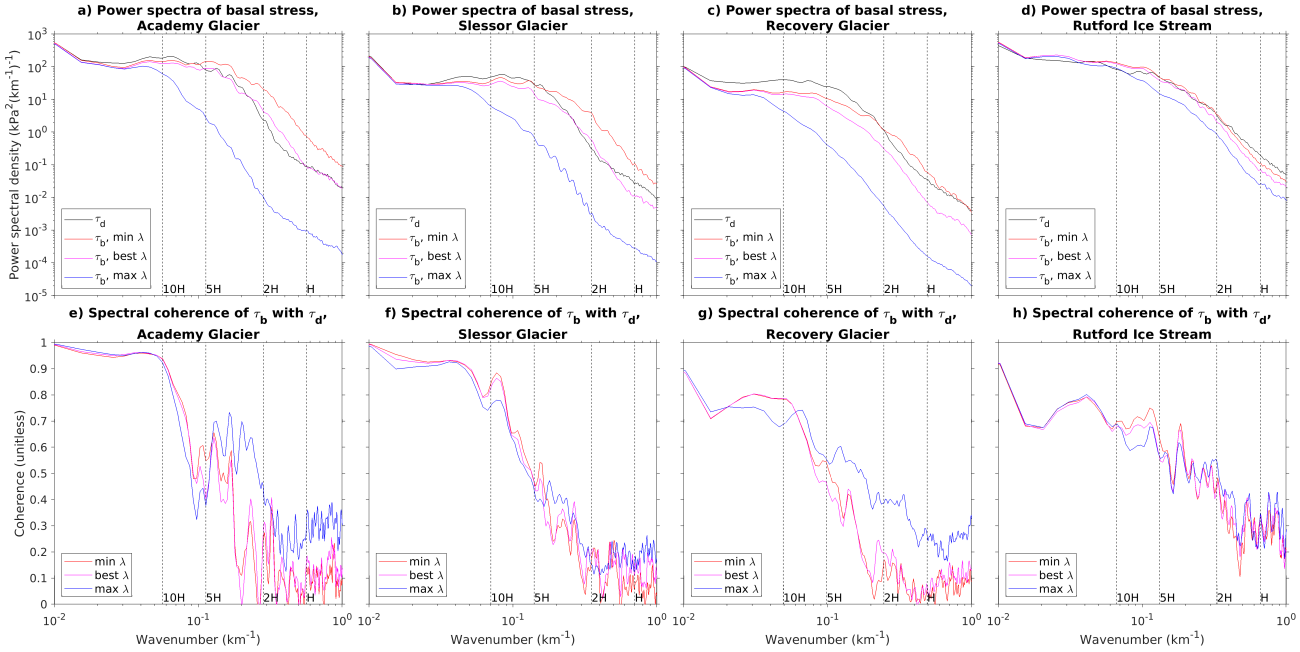

Figure 5. Spectral analysis for all four detail regions highlighted in Fig. 4. Top row (a–d) shows power spectral density of basal stress (τ). Colored lines indicate spectra for minimum acceptable, maximum acceptable, and best λ values, while black lines indicate spectra for driving stress in the flow direction. Bottom row (e–h) show coherence between basal stress $\boldsymbol{\tau}_\mathrm{b}$ and driving stress τ_d. Vertical dashed lines in all plots indicate wavelengths for select multiples of the local average ice thickness.

$\lambda_{\min}$ and $\lambda_{\max}$ occurs at a wavelength of about twice the ice thickness; for Academy Glacier the difference between $\lambda_{\min}$ and $\lambda_{\max}$ is about three orders of magnitude at this wavelength, while for Rutford Ice Stream the difference is less than one. For Academy and Slessor glaciers, the $\lambda_{\min}$ inversion actually has about an order of magnitude more spectral power than the driving stress at this wavelength, and $\lambda_{\min}$ remains above the driving stress for all wavelengths less than 4–5 H (Fig. 5a,b). However, λ_{best} has less spectral power than the driving stress at most wavelengths and analysis regions, only slightly exceeding the spectral power in driving stress for Academy and Slessor glaciers near a wavelength of 2 H (Fig. 5a–d). Academy and Slessor glaciers also have very high ($\gtrsim 90\%$) coherence values between $\boldsymbol{\tau}_\mathrm{b}$ and τ_d at long wavelengths, but a steep drop in coherence towards shorter wavelengths (Fig. 5e,f). This indicates that, in the "ribby" regions, driving stress and basal drag are almost fully balanced at scales greater than about 10 ice thicknesses, but this balance drops rapidly at scales of 5x or 2x the ice thickness. Thus, many of the high-resolution ribs revealed by the inversion in these regions are not mere copies of structure observable in the driving stress; the inversion adds value by revealing structure that cannot be easily inferred from ice sheet geometry. In Recovery Glacier and Rutford Ice Stream, by contrast, the gradient in coherence between long and short wavelengths is weaker(Fig. 5g,h), with values of $\sim 75\%$ even at wavelengths much larger than 10 H. This is probably a reflection of stress

transmission in those glaciers, in which fast-flowing trunks with nearly zero basal drag are supported by long-distance stress transmission to shear margins or isolated sticky spots (Fig. 4d–f).

Overall, these results provide qualified support to the hypothesis advanced in Sergienko and Hindmarsh (2013) and Sergienko et al. (2014) that basal drag is often concentrated within narrow elongated ribs, at least for some of the ice streams and glaciers within our domain. Our results alone cannot confirm that the specific till-water instability mechanism they proposed is responsible for creating these ribs, but clearly some process must be invoked to explain this common spatial pattern. However, in contrast to the suggestion in those papers that the finding of ribbed structure was independent of regularization, we find that the narrow ribs are highly sensitive to the choice of λ value. Spectral analysis of our inversion results reveals that the most prominently ribbed regions also display the strongest dependence of short-wavelength structure on λ, while the least ribbed glacier we analyzed (Rutford Ice Stream) also had the weakest dependence of short-wavelength spectral power on regularization.

4.4 Resolution dependence

In Figure 6, we show the resolution dependence of our results. As discussed in Section A2, we use the area-weighted median hydraulic diameter as a characteristic element size for our refined meshes, and we here restrict our analysis to element size in the fast-flowing part of the grounded ice sheet ($\gtrsim 15\,\mathrm{m\,a^{-1}}$). We find that, in the low-λ limit, the minimum achievable J_{obs} systematically rises with increasing element size, while the entire L-curve systematically shifts towards lower J_{reg} (Fig. 6a). That is, not only do inversions with coarse meshes produce less structure than inversions with fine meshes at equal values of λ, but in addition coarse meshes are also unable to fit the data as well as fine meshes, regardless of λ. The inferred values of λ_{min}, λ_{best}, and λ_{max} all systematically drop as mesh resolution improves (Fig. 6b).

The dependence of all three λ values on element size is well fit by an exponential function, indicating that our results are converging towards the following non-zero λ values in the continuous solution: $\lambda_{\mathrm{min}} = 0.02$, $\lambda_{\mathrm{best}} = 0.16$, and $\lambda_{\mathrm{max}} = 7.8$. Because we normalized our cost function components before combining them, we can interpret these λ values as "small": at least for high resolution meshes, our best estimate is that the regularization term only needs to be weighted about $1/6$ as strongly as the data term. The slope of the fit can be described by the e-folding element size, which is about 2–3 km for all three λ values. (Fig. 6b). This lengthscale is comparable to the average ice thickness in our domain of 2 km. It is also comparable to the finest common flight line spacing of the radar data that were used to generate the bed topography dataset (Morlighem et al., 2020); below this lengthscale we would not expect our mesh to resolve substantially new features in the ice geometry. It is likely that both of these factors contribute to setting the e-folding element size. Once mesh resolution is finer than this scale, diminishing marginal returns set in, λ_{best} approaches its limiting value, and further improvements in resolution will not allow the inversion to resolve finer structure.

In addition to examining variation in λ values, we would also like to explore the convergence of our inversion results themselves. We obviously do not have the continuous solution to compare our results with, but we can estimate an approximate picture by comparing our inversion results for λ_{best} in each mesh with the inversion results in Mesh #1 (Fig. 6c,d). This comparison reveals that misfits for the entire grounded domain are lower (Fig. 6c), and correlations higher (Fig. 6d), than for

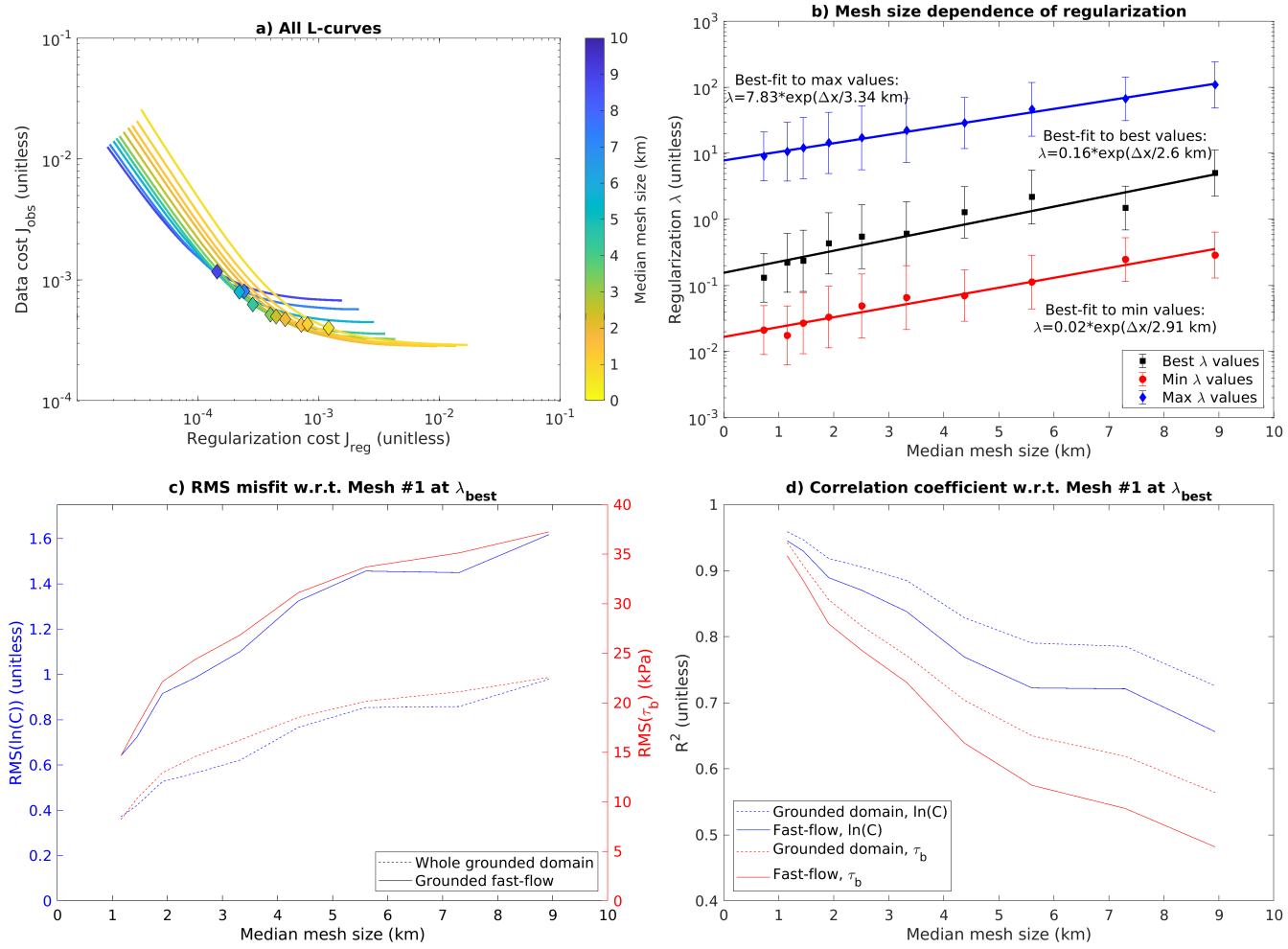

Figure 6. Convergence of L-curve analysis with mesh resolution. a) All L-curves with Weertman sliding overlain on each other. For clarity, only the fitted smooth tradeoff curves are shown, with solid diamonds marking λ_{best} for each mesh. Curves are colored by characteristic element size. b) Estimated values of λ_{min}, λ_{best}, and λ_{max}, along with their associated uncertainties, as a function of mesh resolution. Best-fit exponentials (linear plots on the semi-logarithmic axes used here) are shown as well. c) RMS misfit in drag coefficient ($\ln(C)$) and drag ($\boldsymbol{\tau}_{\mathrm{b}}$) as a function of mesh resolution. Each data point in this plot represents a comparison between λ_{best} for a particular mesh with λ_{best} in Mesh #1. Solid lines represent the fast-flow domain ($\gtrsim 15\,\mathrm{m\,a^{-1}}$), dashed lines represent the entire grounded domain. d) As (c), but showing correlation coefficients instead of RMS misfits. As discussed in the text, we use the area-weighted median of hydraulic diameter in moderately fast-flow regions ($\gtrsim 15\,\mathrm{m\,a^{-1}}$) to represent a characteristic element size for our unstructured refined meshes.

the fast-flow domain alone. This is likely because the spatial structure of the entire domain is dominated by the large-scale structure of streaming flow (Fig. 4), which is resolved even in the coarsest mesh we used, while restricting our analysis to the fast-flow domain represents a more difficult test of model convergence. The RMS misfits for $\ln(C)$ and $\boldsymbol{\tau}_{\mathrm{b}}$ track each other

closely, with values for the fast-flow domain in the range of 0.6–1.6 for $\ln(C)$ and 15–35 kPa for $\boldsymbol{\tau}_\mathrm{b}$ (Fig. 6c). However,
405 correlation R^2 values are consistently higher for $\ln(C)$ than for $\boldsymbol{\tau}_\mathrm{b}$ (Fig. 6d).

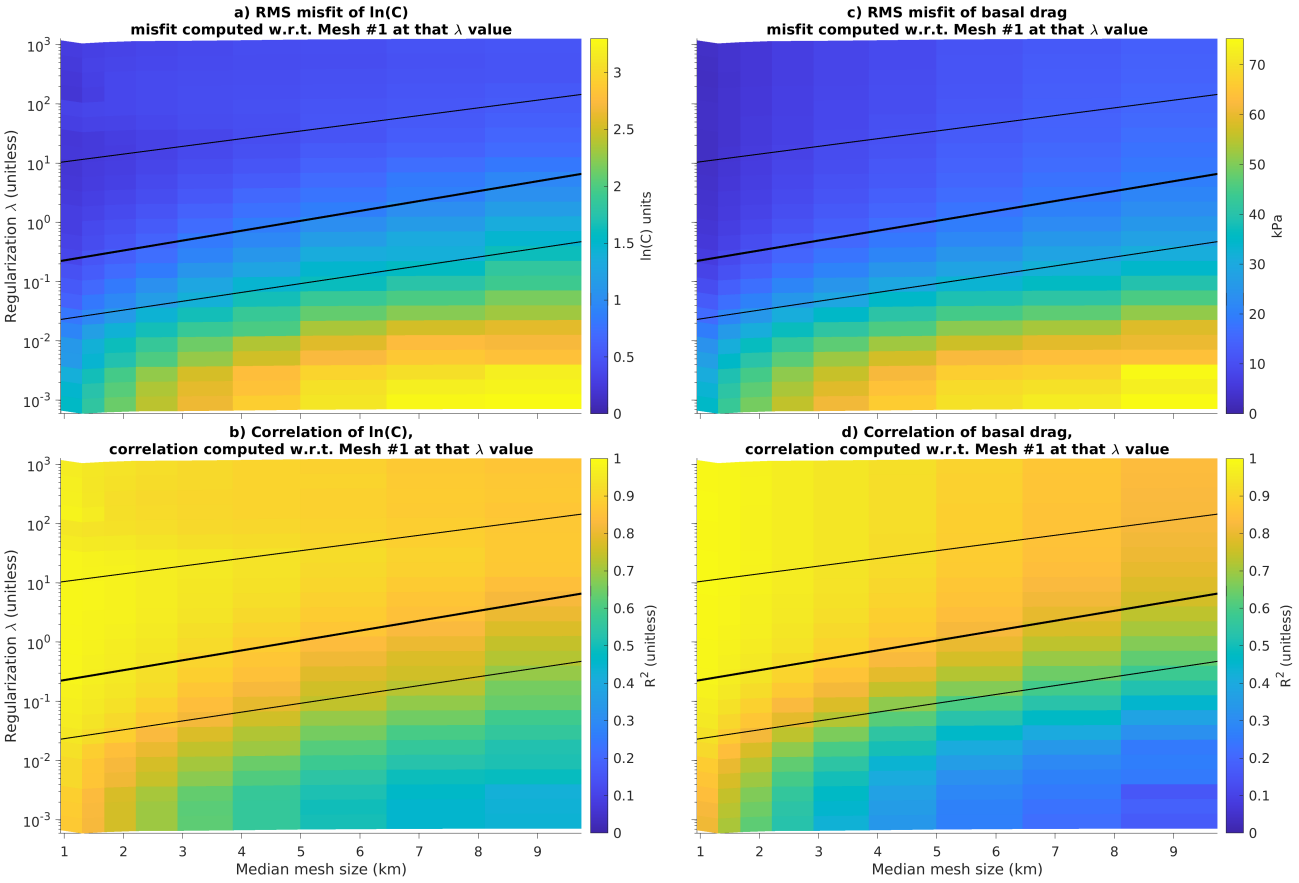

Figure 7. Comparison of inversion results as a function of mesh resolution and λ value. In these plots, each inverted C or $\boldsymbol{\tau}_\mathrm{b}$ field is compared to the corresponding field at the same λ value at the highest-resolution mesh. Left column (a,c) shows comparison of drag coefficient C while right column (b,d) shows comparison of drag $\boldsymbol{\tau}_\mathrm{b}$. Top row (a,b) shows RMS misfit while bottom row (c,d) shows squared correlation coefficient. Best-fit lines from Fig. 6 overlain for comparison. Comparisons are limited to moderately fast-flowing regions ($\gtrsim 15\,\mathrm{m\,a^{-1}}$).

In addition to comparing inversion results at the (variable) λ_best for each mesh, we can also compare our results with Mesh #1 at a constant λ value (Fig. 7). This comparison is useful because inversions with different meshes but the same λ value should, in principle, be solving the exact same set of equations, and so any differences can be treated as model error (in practice the ice sheet geometry and surface velocity data must be filtered at different wavelengths for each mesh as described in Section A3,
410 so different meshes are not solving quite the same problem). We overlay our exponential fits from Fig. 6b for reference. Above the line representing λ_best, misfit in both $\ln(C)$ (Fig. 7a) and $\boldsymbol{\tau}_\mathrm{b}$ (Fig. 7b) is low, with only a weak dependence on λ, while

correlation in both of those parameters is high (Fig. 7c,d). Between the lines representing λ_{best} and λ_{min}, misfit begins to rise (Fig. 7a,b) and correlation begins to drop (Fig. 7c,d), and below λ_{min} both fit metrics worsen rapidly.

These results confirm the utility of L-curve analysis, by demonstrating that the fine structure obtained by the inversion at λ_{best} is robust to changes in mesh resolution. There are two ways to add fine structure to the inversion results: reducing mesh size and reducing regularization. Higher-resolution meshes benefit from both improved numerical convergence and from being forced by higher-resolution geometry and velocity data, allowing them to resolve finer structure at the ice sheet bed. However, reducing λ with mesh resolution held constant will also cause the inversion to add more fine structure and reduce the observational misfit. The important question is thus: if we are limited to a coarse-resolution mesh, at what point does the fine structure obtained by reducing regularization diverge from the fine structure we would have obtained if we were able to use a finer mesh? L-curve analysis is a test that can be performed on a single mesh without reference to inversions on finer meshes. The results shown in Figure 7 demonstrate that the fine structure obtained by reducing regularization begins to diverge from the fine structure obtained by improving resolution below λ_{best}, and below λ_{min} the difference between the two increases rapidly.

4.5 Effective pressure

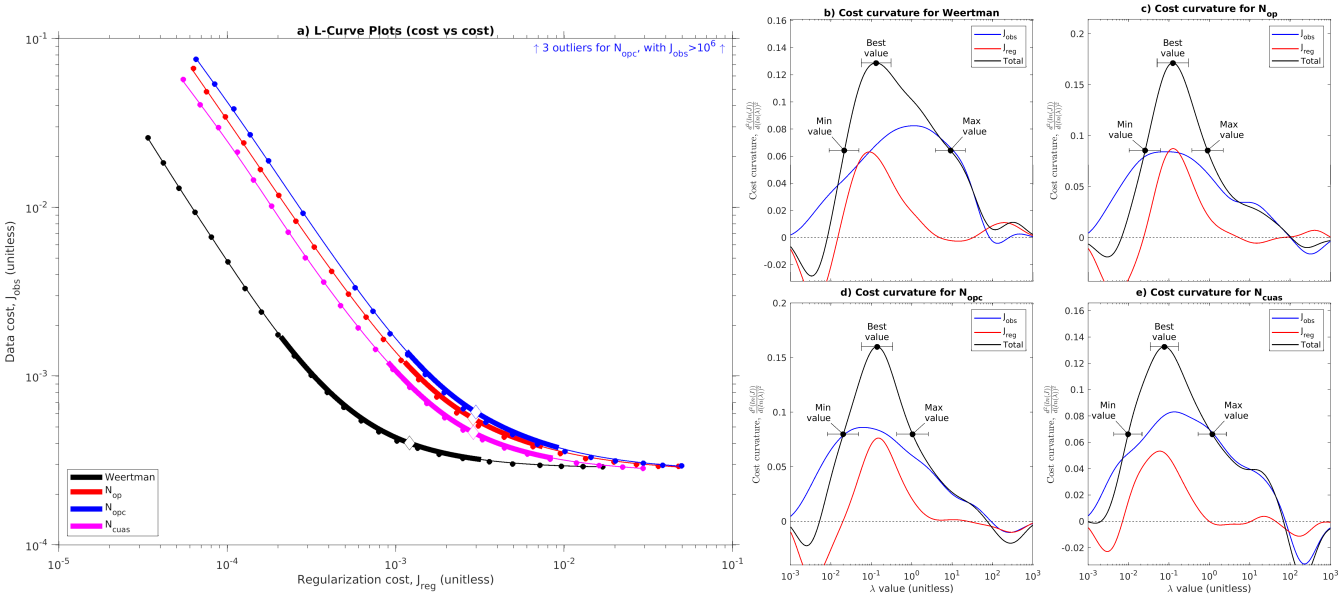

Figure 8. Comparative L-curves for different N fields. a) All L-curves overlain on each other. Small circles represent individual models, thin lines represent smoothed curves, thick lines represent inferred corner regions, and large diamonds represent inferred λ_{best} values. b–d) Cost curvature as a function of λ for each N field, with selection of min, best, and max λ values annotated. Plot (a) is equivalent to plot (a) in Fig. 3, plots (b–d) are equivalent to plot (d) in that figure. Three outlier models from N_{opc} have been excluded from this analysis.

Before analyzing our results for Budd sliding in detail, it is useful to consider Equation 4: in order to reproduce the stress and velocity fields recovered by the Weertman inversion, the Budd coefficient must satisfy the relationship, $C_{\mathrm{W}} = C_{\mathrm{B}} N$, where

C_{W} is the coefficient in the Weertman sliding law and C_{B} is the coefficient in the Budd sliding law. From this, we can form the *a priori* expectation that the utility of a given N field will be dependent on the linear correlation between N and C_{W}. If N has a strong positive correlation with C_{W}, then the inversion will not need to produce much structure in C_{B} in order to fit the observations; in the limit of a perfect N field and no material variation in the substrate, the N field would explain all of the variation in C_{W}, and C_{B} would be a constant. Conversely, if the correlation between N and C_{W} is weak or even negative, then the inversion will need to produce a lot of structure in C_{B} in order to fit the observations.

All three of our candidate N fields had a positive correlation with C_{W}, but N_{CUAS} had much better correlation than either of the two geometry-based N fields (Table 1). In addition, all three N fields had better correlation values when considering the entire grounded domain than when considering the fast-flow region alone (Table 1). For the geometry-based N fields, this is likely due to the physical setting of fast-flowing regions: many fast-flowing ice streams and outlet glaciers are located in subglacial troughs, so a simple geometry-based calculation incorporating the bed topography can approximate the basic dichotomy between slow flow and fast flow. For N_{CUAS}, this is likely because the melt rate forcing to CUAS incorporated shear heating estimated from the observed velocity field. Restricting the analysis to the fast-flowing region alone represents a much more stringent test; N_{CUAS} can explain 33% of the variance of C_W within the fast-flowing region of the domain, but the two geometry-based N fields can explain almost none of it.

Figure 8 shows comparative L-curves for all of them and Table 1 summarizes the results of a comparison between the λ_{best} inversions for each of those N fields with the λ_{best} inversion for Weertman sliding. The L-curves for all three N fields have a similar shape as the L-curve for Weertman sliding, but shifted towards higher J_{reg} (Fig. 8a). Total curvature was in the range of 0.13–0.17 for all four models (Fig. 8b–e), providing a quantitative confirmation of the visual similarity in shape. The inferred λ_{best} values are shifted towards lower λ values, mostly driven by a shift in the data curvature term (Fig. 8c–e, c.f. b). A visual inspection of the respective L-curves suggests that the asymptotic best achievable J_{obs} is similar for all sliding laws, but the λ_{best} inversions for inversions with N have higher J_{obs} than the λ_{best} inversion for Weertman sliding (Fig. 8a). Three of the inversions in the N_{opc} L-curve were unable to converge, with J_{obs} many orders of magnitude higher than the other models (Fig. 8a), but the remaining inversions were sufficient for us to recover a smooth tradeoff curve.

At face value, Figure 8a implies that N_{opc} required the most structure to fit the data, almost a full order of magnitude more structure than the inversions with Weertman sliding. However, it is possible that the increase in J_{reg} could have been caused by the different normalizations used, as the estimate of the characteristic scale S_{reg} depends on the value of N (Eqs 11, 14). To remedy this, we can obtain a scale-independent estimate of the magnitude of the structure produced by the inversion by computing the variance of $\ln(C)$. Converting C to a logarithmic scale before computing the variance ensures that the change in units and mean magnitude between Weertman and Budd sliding has no effect on the calculation. This calculation reveals that all of the candidate N fields reduced the variance in the inverted C coefficient when considered over the entire domain, but N_{CUAS} reduced the variance the most (Table 1). When the fast-flow domain is considered alone, N_{op} produced a slight *increase* in variance, N_{opc} a slight decrease, and N_{CUAS} the largest decrease. All of the inversions with N had larger J_{obs} values than the inversion with Weertman sliding, regardless of whether or not the analysis is restricted to the fast-flow domain, with N_{CUAS} having the smallest increase in observational misfit of the three (Table 1). In addition, the scaling of J_{obs} is determined only

by the statistics of the observations (Eq. 14), so we know that differences in J_{obs} between models are directly comparable with one another. Nonetheless, the absolute magnitude of error for all of the models is relatively low, with $J_{\mathrm{obs}} < 10^{-3}$ over the whole domain and $J_{\mathrm{obs}} < 2 \times 10^{-3}$ over the fast-flow region; converted back to dimensional form using $\Delta u = \sigma_d \sqrt{J_{\mathrm{obs}}}$, these correspond to characteristic velocity errors of $\sim 5\,\mathrm{m\,a^{-1}}$ over the whole domain and $\sim 10\,\mathrm{m\,a^{-1}}$ in the fast flow. We can evaluate the overall quality of each model by computing the total variance ratio, which is the product of the coefficient variance ratio and the J_{obs} ratio (Table 1). Doing so reveals that the increase in observational misfit outweighs the reduction in coefficient variance for the two geometry-based N fields (N_{op} and N_{opc}) regardless of whether the analysis is performed for the whole domain or for the fast-flow alone; for N_{CUAS}, by contrast, we obtain a reduction in total variance for the whole domain and a very slight increase in the fast flow.

Table 1. This Table summarizes inversion performance for the three sources of effective pressure N that we tested on Mesh #1, alongside results for Weertman sliding (no N). Each model is evaluated at its own λ_{best} value. We show statistics both for the entire grounded domain and for fast flow regions only ($\gtrsim 15\,\mathrm{m\,a^{-1}}$).

Model	N correlation $R^2(N, C_W)$ (unitless)	C variance $\sigma^2(\ln(C))$ (unitless)	Obs Cost J_{obs} ($\times 10^{-4}$)	Equiv Δu $\sigma_d\sqrt{J_{\mathrm{obs}}}$ ($\mathrm{ma^{-1}}$)	Var ratio (a) $\frac{\sigma^2(\ln(C))}{\sigma^2(\ln(C_W))}$ (unitless)	Var ratio (b) $\frac{J_{\mathrm{obs}}}{J_{\mathrm{obs},W}}$ (unitless)	Var ratio (tot) $a \times b$ (unitless)
Whole grounded domain:							
no N	-	3.22	3.96	4.98	1.00	1.00	1.00
N_{op}	0.18	3.07	5.33	5.78	0.95	1.35	1.28
N_{opc}	0.07	2.97	6.81	6.53	0.92	1.72	1.59
N_{CUAS}	0.51	2.18	5.07	5.64	0.68	1.28	0.87
Grounded fast flow:							
no N	-	7.04	13.8	9.31	1.00	1.00	1.00
N_{op}	0.00	7.27	17.5	10.48	1.03	1.27	1.31
N_{opc}	0.00	6.77	19.7	11.11	0.96	1.42	1.37
N_{CUAS}	0.33	5.91	16.7	10.24	0.84	1.21	1.02

4.6 Nonlinear Sliding

In Figure 9, we show comparative L-curves for linear and nonlinear Weertman sliding laws. These results reveal that, in a nonlinear sliding law, the high-λ limb of the L-curve is substantially steeper than in a linear sliding law, producing a sharper corner in the L-curve (Fig. 9a). The peak in cost curvature became narrower as sliding became more nonlinear, reducing the spread between λ_{min} and λ_{max} and allowing us to define the corner region more precisely (Fig. 9c,e,g). The increased visual sharpness of the L-curves for nonlinear sliding can be confirmed by looking at the total curvature, which doubles from 0.13 for $m = 1$ to 0.26 for $m = 5$ (Fig. 9c,e,g). The increase in L-curve sharpness for nonlinear sliding is driven by both a steeper increase in J_{obs} and a shallower reduction in J_{reg} with increasing λ (Fig. 9b,d,f). Both of these effects can be explained by an

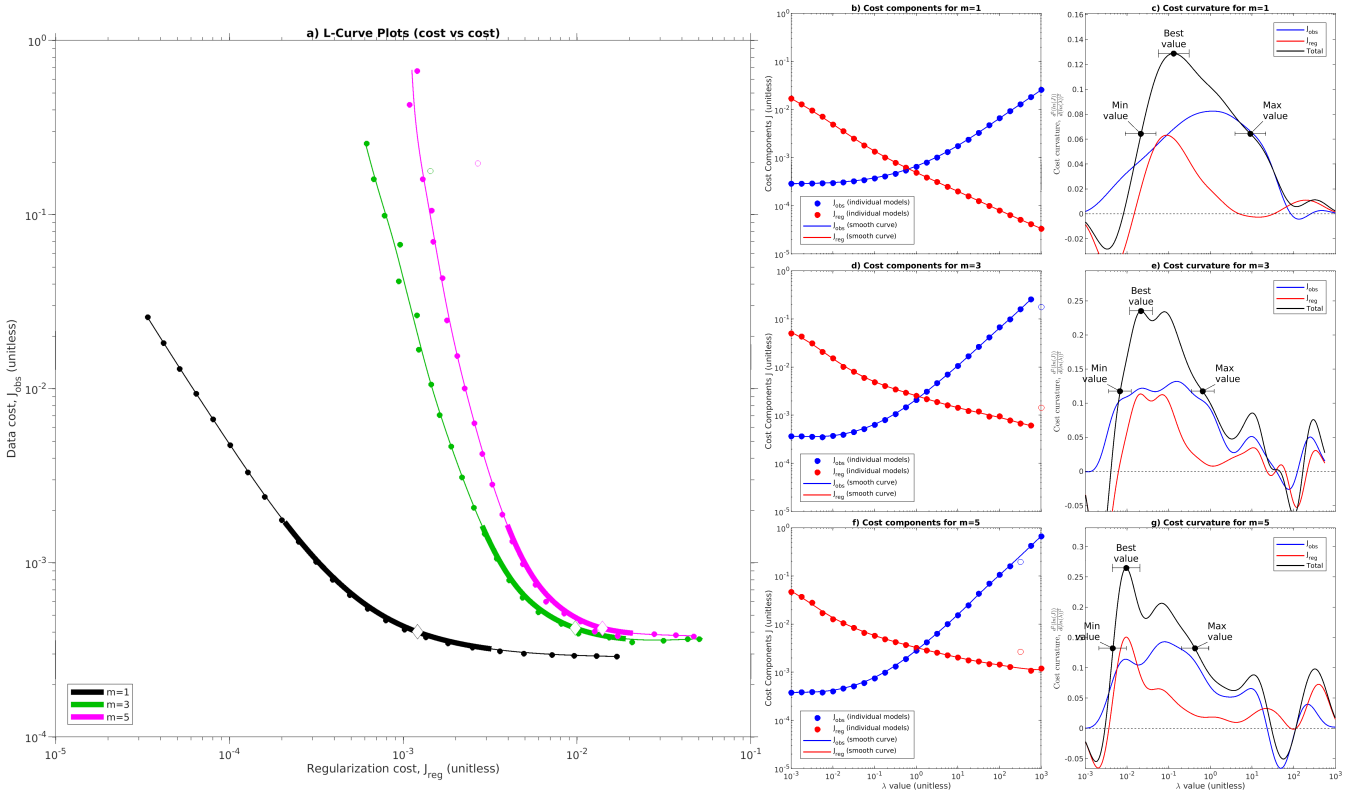

Figure 9. Comparative L-curves for different values of the sliding exponent m in Mesh #1. a) All L-curves overlain on each other. Filled circles represent individual models, thin lines represent smoothed curves, thick lines represent inferred corner regions, and large diamonds represent inferred λ_{best} values. Open circles represent two outlier models excluded from the curve-fitting (one each for $m = 3$ and $m = 5$). b) Individual cost components for $m = 1$. Filled circles represent individual models, lines represent fitted curves, and open circles represent outlier models. c) Cost curvature and selection of λ_{min}, λ_{best}, and λ_{max} for $m = 1$. d–g) As (b) and (c), but for $m = 3$ and $m = 5$.

increasing sensitivity of velocity to the coefficient in nonlinear sliding. Rearranging Eq. 3 to solve for u reveals that $u \propto C^{-m}$; thus, when C is perturbed away from a value that fits the observations closely, the resulting increase in J_{obs} will be larger for nonlinear sliding than for linear sliding. The increased sensitivity of J_{obs} to C in turn prevented the inversion from smoothing C as much as it would for linear sliding, producing higher J_{reg} values at large λ. The combination of a smaller decrease in J_{reg} with a bigger increase in J_{obs} at large λ produces a sharper L-curve.

As in our experiments with N, a first glance at our L-curves in Figure 9a seems to imply greater inversion structure with higher values of m. However, again care needs to be taken when interpreting J_{reg}, as the scaling of J_{reg} depends on the value of m (Eq. 11). Again, we can obtain a scale-independent estimate of the amount of structure present in the inverted field by computing the variance of $ln(C)$. Doing so reveals that the structure produced by the inversion drops drastically as m increases, dropping by a factor of 3–4 over the whole domain and by a factor of 2–2.5 in the fast flow (Table 2). This reduction in variance can also be seen by examining the color scales in Figure 10. At the high end, linear Weertman requires

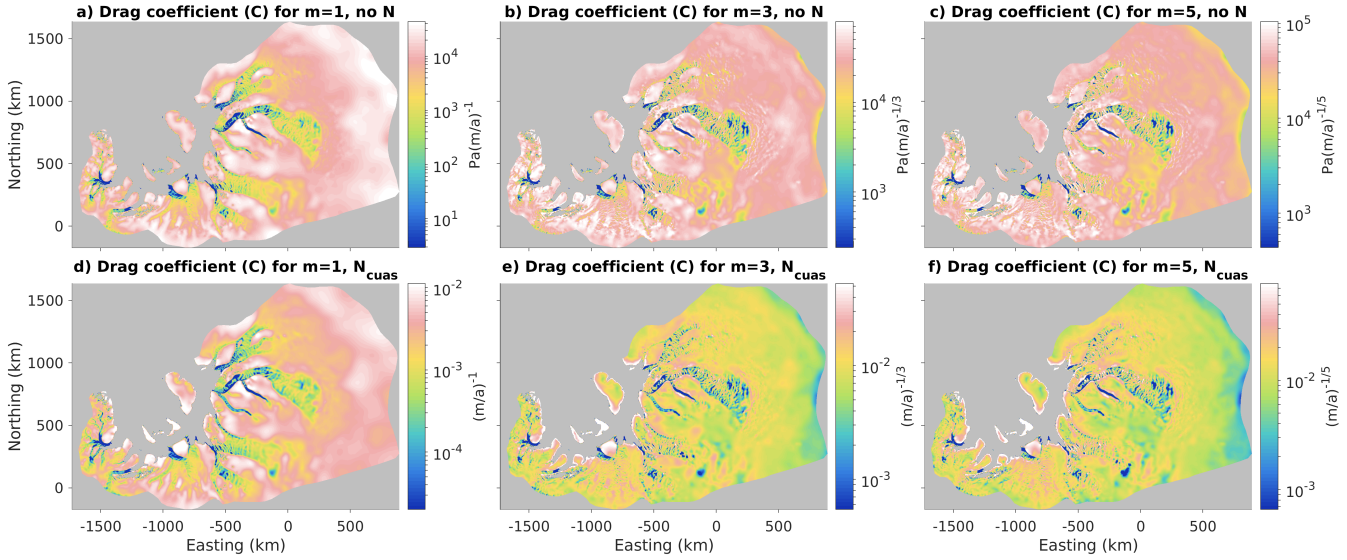

Figure 10. Comparative maps of inverted drag coefficient C for different values of the sliding exponent m, both with and without effective pressure N. Top row shows drag coefficient for Weertman sliding (without N), for $m = 1, 3, 5$. Bottom row shows the same values of m for inversions with Budd sliding using N_{cuas}. Each subplot uses a logarithmic color scale with the limits automatically set to the 1^{st} and 99^{th} percentiles of the distribution.

over five and a half orders of magnitude of dynamic range to display the 1%–99% range of its distribution, while at the low end, nonlinear Budd requires less than two and a half orders of magnitude. The spatial structure of C also changes somewhat in nonlinear sliding, with greater ribbing and less of a simple dichotomy between fast and slow regions, producing a C field that starts to have more visual features of the $\boldsymbol{\tau}_{\mathrm{b}}$ field for linear sliding (Fig. 10b,c, c.f. Fig. 4d–f). With nonlinear Budd sliding the large-scale contrast between slow and fast flow is greatly reduced, with much of that distinction being accounted for by N_{cuas} rather than C (Fig. 10e,f). The coefficient variance can also be compared with the variance of $\ln(\boldsymbol{\tau}_{\mathrm{b}})$, as we expect that $C \to \boldsymbol{\tau}_{\mathrm{b}}$ as $m \to \infty$. For linear Weertman, the variance in the coefficient is nearly three times the variance in basal drag when considered over the whole domain, but this drops to only a 20% increase in variance for $m = 5$. For N_{CUAS} with $m = 5$, the coefficient variance is only 5% higher than the drag variance over the whole domain, and it is actually 7% *less* in the fast flow. Observational misfit is slightly higher for nonlinear as compared to linear sliding, although as with our experiments with N, the absolute magnitude of J_{obs} is low for all inversions and the equivalent velocity errors are on the order of $\sim 5\,\mathrm{m\,a^{-1}}$ for the whole domain and $\sim 10\,\mathrm{m\,a^{-1}}$ in the fast flow. However, unlike in our experiments with N, for our experiments with m the large reduction in coefficient variance moving to nonlinear sliding more than makes up for the slight increase in misfit, producing total variance ratios that are an unambiguous improvement (Table 2). The improvement produced by moving from $m = 1$ to $m > 1$ in Weertman sliding is about a factor of 3 when measured across the whole domain, and about a factor of 2 within the fast flow. $m = 5$ produced a slight improvement as compared to $m = 3$, but by far the biggest jump came between

$m = 1$ and $m = 3$. With N_{CUAS} the improvement when moving from linear to nonlinear sliding was somewhat less, although still strong (Table 2).

Table 2. This table summarizes inversion performance for the three values of the sliding exponent m that we considered in Mesh #1. Each m is evaluated at its own λ_{best} value. C_0, the drag coefficient used for comparison with the other models, is the reference model ($m = 1$, no N). We show statistics both for the entire grounded domain and for fast flow regions only ($\gtrsim 15\,\mathrm{m\,a^{-1}}$).

Model	C variance $\sigma^2(\ln(C))$ (unitless)	τ_{b} variance $\sigma^2(\ln(\tau_{\mathrm{b}}))$ (unitless)	Obs Cost J_{obs} ($\times 10^{-4}$)	Equiv Δu $\sigma_d\sqrt{J_{\mathrm{obs}}}$ ($\mathrm{ma^{-1}}$)	Var ratio $\frac{\sigma^2(\ln(C))}{\sigma^2(\ln(\tau))}$ (unitless)	Var ratio (a) $\frac{\sigma^2(\ln(C))}{\sigma^2(\ln(C_0))}$ (unitless)	Var ratio (b) $\frac{J_{\mathrm{obs}}}{J_{obs,0}}$ (unitless)	Var ratio (tot) $a \times b$ (unitless)
Whole grounded domain:								
$m = 1$, no N	3.22	1.13	4.0	5.0	2.86	1.00	1.00	1.00
$m = 3$, no N	1.10	0.80	4.2	5.1	1.38	0.34	1.06	0.36
$m = 5$, no N	0.90	0.75	4.3	5.2	1.19	0.28	1.09	0.30
$m = 1$, N_{CUAS}	2.18	0.89	5.1	5.6	2.44	0.68	1.28	0.87
$m = 3$, N_{CUAS}	0.87	0.80	5.8	6.0	1.09	0.27	1.47	0.40
$m = 5$, N_{CUAS}	0.90	0.86	5.1	5.7	1.05	0.28	1.29	0.36
Grounded fast flow:								
$m = 1$, no N	7.0	4.4	14	9.3	1.59	1.00	1.00	1.00
$m = 3$, no N	3.3	2.8	15	9.7	1.18	0.47	1.08	0.51
$m = 5$, no N	2.9	2.6	15	9.7	1.11	0.41	1.08	0.45
$m = 1$, N_{CUAS}	5.9	3.4	17	10.2	1.75	0.84	1.21	1.02
$m = 3$, N_{CUAS}	2.9	2.9	20	11.3	1.01	0.41	1.48	0.61
$m = 5$, N_{CUAS}	2.9	3.1	19	10.8	0.93	0.41	1.35	0.56

4.7 Higher-Order Inversions

We produced L-curves for HO inversions with linear Weertman sliding in Meshes # 4, 6, and 8. We compare these with the corresponding SSA inversions in Figure 11. While we were able to obtain smooth monotonic L-curves with well-defined corner regions for all three experiments, the resulting L-curves clearly indicate that the HO inversions were inferior to SSA inversions using the same mesh and sliding law. Note that the scalings for both J_{reg} and J_{obs} are independent of the flow equation used, so the relative positions of the SSA and HO L-curves in Figure 11 are directly comparable to one another. This positioning indicates that the HO inversions performed worse than the SSA inversions in both components of the cost function. The difference in the asymptotic best achievable J_{obs} is particularly notable: the HO inversions were incapable of achieving values of J_{obs} less than 10^{-3}, regardless of mesh resolution. By contrast, the asymptotic best achievable J_{obs} for the SSA inversions improved with finer mesh resolution, as discussed above.

A detailed examination of the misfit maps for the HO inversions (not shown) revealed that their increased observational misfit was largely due to increased positive misfit (ie, model velocity too high) in slow-flowing regions. Thus, we hypothesized

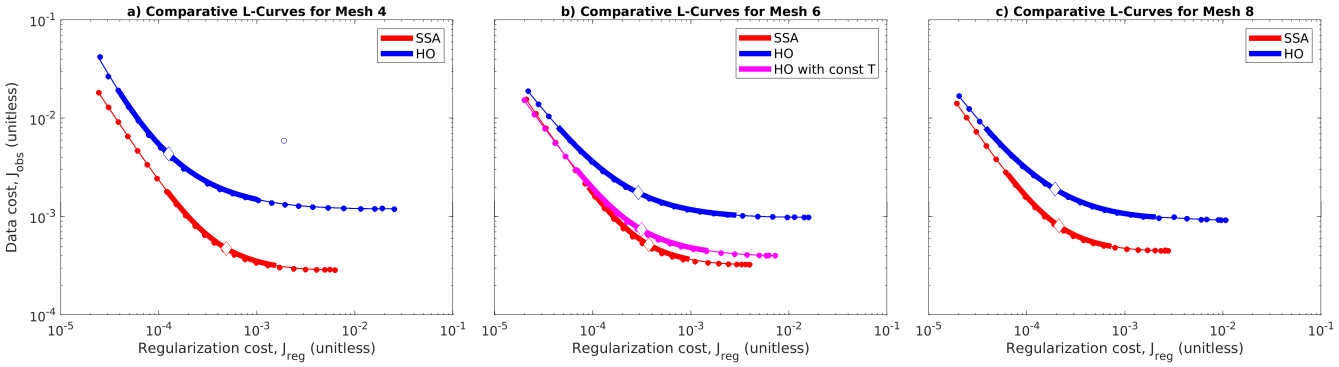

Figure 11. Comparative L-curves for HO and SSA inversions. a) Comparative L-curves for Mesh #4. Solid circles indicate model results, thin smooth lines indicate fitted curves, thick lines indicate corner region, diamonds indicate λ_{best} location, and hollow circle indicates an outlier model excluded from the curve-fitting for HO. Red represents SSA L-curve, blue represents HO. b) as (a), but for Mesh #6. Magenta represents HO with constant ice temperature of -25°C. c) as (a), but for Mesh #8.

that they may have been unable to fit the data because they contained too much deformation flow: if the temperature model used as input contained too much warm ice near the base, then a vertically resolved HO model would have too much shear deformation in the lower part of the ice column, resulting in model surface velocities that were faster than observed, with no way to bring the model velocities down by adjusting basal slip. We tested this hypothesis by running one additional L-curve experiment in Mesh #6 with HO and a constant ice rheology corresponding to a temperature of -25°C throughout the domain. The results are shown by the magenta line in Fig. 11b. With a constant, and fairly stiff, ice rheology, this HO model was able to make up almost all of the difference with the SSA L-curve.

We summarize the performance of the two HO inversions in Mesh #6 as compared to the SSA inversion for that mesh in Table 3. Basal drag was similar in both the 2D and 3D models, with RMS differences in $\boldsymbol{\tau}_{\mathrm{b}}$ on the order of $10\,\mathrm{kPa}$. The HO model with variable ice temperature required more regularization and thus had 12% less variance in its coefficient than the SSA model, but it also had a big increase in J_{obs}, resulting in a tripling in the total variance ratio (Table 3). The HO inversion with constant ice temperature performed better than this, but still had a 68% increase in total variance as compared to SSA. When the analysis is restricted to the fast-flowing domain, then the HO inversion with variable temperature improves somewhat, dropping down to a variance ratio compared to SSA of "only" 1.91, while the variance ratio for HO with constant temperature is nearly identical within the fast flow as it is in the whole domain (Table 3). These results demonstrate that, while SSA outperforms HO regardless of whether the analysis is performed over the whole domain or in the fast flow alone, the SSA model has the biggest advantage when compared against the HO model with variable temperature in a comparison region that includes both slow and fast flow. We discuss our interpretation of these results in more detail in Section 5.3.

Table 3. This table summarizes inversion performance for HO and SSA models in Meshes #4, 6, and 8. For Mesh #6, we also considered an HO model with constant ice temperature. Each model is evaluated at its own λ_{best} value. Variance ratios are computed with respect to the SSA model at the same element size. We show statistics both for the entire grounded domain and for fast flow regions only ($\gtrsim 15\,\text{m}\,\text{a}^{-1}$).

Model	C variance $\sigma^2(\ln(C))$ (unitless)	$\boldsymbol{\tau}_\text{b}$ difference $RMS(\tau_\text{b})$ (kPa)	Obs Cost J_obs ($\times 10^{-4}$)	Equiv Δu $\sigma_d\sqrt{J_\text{obs}}$ (ma^{-1})	Var ratio (a) $\frac{\sigma^2(\ln(C))}{\sigma^2(\ln(C_0))}$ (unitless)	Var ratio (b) $\frac{J_\text{obs}}{J_{obs,0}}$ (unitless)	Var ratio (tot) $a \times b$ (unitless)
Whole grounded domain:							
Mesh #6, SSA	2.56	0.00	5.8	6.03	1.0	1.00	1.00
Mesh #6, HO	2.25	8.85	20.1	11.19	0.88	3.44	3.02
Mesh #6, HO, const T	2.79	7.37	9.0	7.49	1.09	1.54	1.68
Grounded fast flow:							
Mesh #6, SSA	5.33	0.00	19.2	10.9	1.00	1.00	1.00
Mesh #6, HO	4.36	14.2	44.8	16.7	0.82	2.33	1.91
Mesh #6, HO, const T	6.38	11.4	26.5	12.9	1.20	1.38	1.66

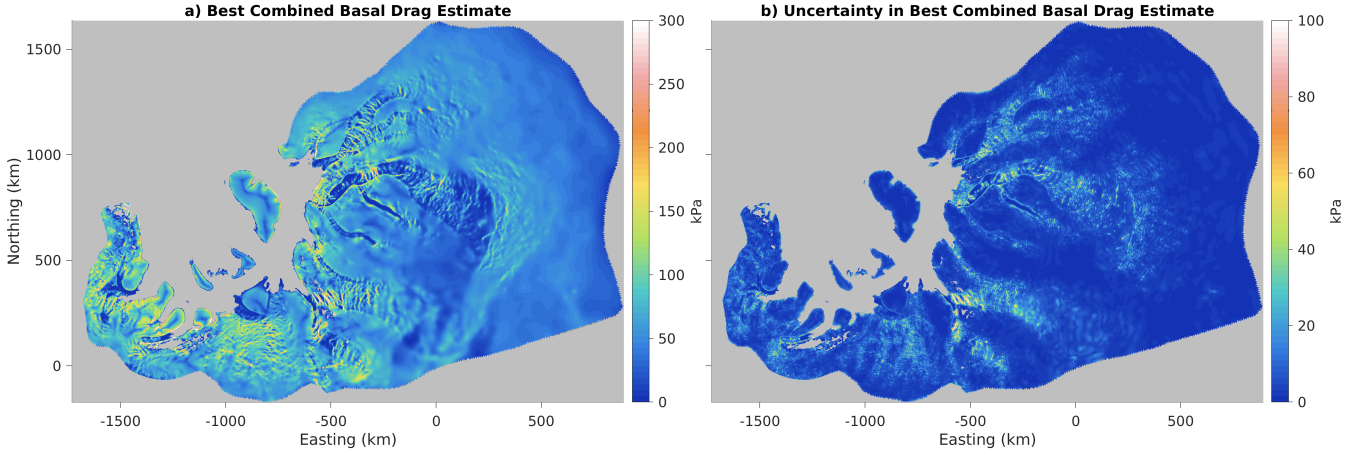

Figure 12. Our best combined estimate of the ice sheet basal drag (a) and its uncertainty (b). This map is produced by taking a weighted average of multiple models as described in the text. The uncertainty is computed from the weighted standard deviation. Each model is weighted by the reciprocal of its total variance ratio.

4.8 Best Drag Map

Throughout this paper we have produced many different inversions and competing descriptions of the ice sheet basal drag. In this subsection, we combine them to produce a single consensus picture representing our best combined estimate. We feel that it is more appropriate to produce a consensus view of drag, $\boldsymbol{\tau}_\text{b}$, rather than drag coefficient C, because drag is directly comparable between different sliding laws, while the units and mean magnitude of C vary with both m and N. In this consensus

estimate we include all eight of the L-curves that we performed on our highest resolution mesh, that is, Weertman sliding with $m = 1, 3, 5$, Budd sliding with N_{CUAS} and the same three m values, and linear Budd sliding with N_{OP} and N_{OPC}. In order to sample the range of uncertainty associated with regularization, for each L-curve we included inversions with $\lambda_{\min}$, λ_{best}, and $\lambda_{\max}$. Overall, our combined estimate is built from 24 individual inversions on our highest resolution mesh. Each inversion is weighted according to the inverse of its total variance ratio. For this purpose we use the value of total variance computed for the whole domain. We estimated the uncertainty in this combined drag map by computing the weighted standard deviation of these 24 inversions about the mean field using the same weighting.

Figure 12a shows the resulting spatial structure of basal drag. A number of faster-flowing regions are supported by rib-like structures in drag, including upstream Recovery and Slessor glaciers, Academy glacier, and the region of the Robin Subglacial Basin upstream of Institute and Möller ice streams. Generally these ribby regions are further upstream in the fast flow, while further downstream many fast-flowing trunks have nearly zero basal drag supported by stress transmission from shear margins or isolated sticky spots. Fast-flowing trunks with mostly zero basal drag are seen in Recovery and its tributaries, Rutford, Evans, Foundation, and downstream Slessor glaciers. Stress transmission from fast-flowing trunks to the shear margins produces thin strips of high drag just outside of the main trunk of many glaciers. In general, more visually ribby structure tends to be located in moderately fast-flow regions, ~ 30–$300\,\mathrm{m\,a^{-1}}$, where streaming flow is spread out over broad subglacial basins. Where fast flow is more intense ($\gtrsim 300\,\mathrm{m\,a^{-1}}$) and concentrated in narrower troughs, the tendency is towards near-zero basal drag supported by isolated sticky spots or side drag from shear margins. Recovery and Support Force glaciers have regions of high drag near their grounding lines, but otherwise most glaciers continue their low-drag troughs right into the ice shelf.

The uncertainty in our combined drag estimate is mostly low, but with large excursions co-located with areas of high drag (Fig 12b). The mean uncertainty is $5.5\,\mathrm{kPa}$ in the whole grounded domain and $12.6\,\mathrm{kPa}$ in the fast-flowing region. In general, high-drag ribs and shear margins tend to also have higher uncertainty, especially in the fast-flowing part of the domain, while the weak-bedded trunks of glaciers have low uncertainty (Fig 12b). The apparent visual correlation between mean drag and drag error can be confirmed by fitting a linear relationship between them with the constraint that the fitted relationship must pass through the origin; doing so reveals that drag uncertainty is 11% of the mean drag over the whole domain and 18% of the mean in the fast flow, with R^2 values for these relationships of 0.53 and 0.66, respectively.

5 Discussion

Fundamentally, the purpose of regularization is to determine the information content of our observations by managing the tradeoff between increased complexity in the inversion target field and reduced observational misfit. While one can usually reduce the observational misfit by adding ever more complexity to the inversion target field, this is equivalent to increasing the number of free parameters one is allowed to tune, and the meaningfulness of a good observational fit achieved with an excess of free parameters is questionable. Occam's razor suggests that we should prefer the model that explains the maximum amount of observational variance with the minimum amount of spatial structure. The ideal λ value is thus defined by the point of diminishing marginal returns: starting from a highly regularized state, reducing λ improves the observational fit substantially

without adding much spatial structure to the target field. At some point, diminishing marginal returns set in, such that further improvements require disproportionately complex spatial structure in the target field to achieve only slight reductions in misfit. Thus, regularization is not an ad hoc or arbitrary addition to the inversion process, nor should regularization be treated as an inconvenient or unfortunate necessity required to produce numerical stability. Rather, regularization is a fundamental measure of the information content achievable in geophysical inverse problems, and the selection of optimal regularization should be regarded as an essential part of glaciological inversions. A well-formed L-curve allows us to answer the question: how much structure is actually required to fit the data? Conversely, a poorly-formed L-curve casts doubt on the meaningfulness of all of the spatial structure recovered by an inversion.

This focus on information content also provides a useful framework for thinking about the question of convergence in inverse models. In forward models, "convergence" can be simply defined to mean that, in the limit that mesh resolution approaches zero, the numerical solution approaches the continuous solution. In inverse models, by contrast, there are two fields that the numerical solution could potentially approach: the continuous solution, and the true field. If the true field is sufficiently rough in comparison to the attenuating effect of the forward problem, then the limit of λ_{best} will not be 0, and the continuous solution of the regularized inverse problem will not be the same as the true field (Vogel, 1996). This split is also reflected in the fact that, for inverse problems, we must formally consider "convergence" with respect to multiple limits: the limit as model resolution approaches zero, the limit as observational error approaches zero, and the limit as observational resolution approaches zero. In our resolution tests (Section 4.4), we explored the convergence behavior with respect to both model and observational resolution (because we smoothed the observations onto coarser meshes, Section A3), but not convergence behavior with respect to observational error. We found that λ_{best} approached a non-zero limit, implying that (at least with non-zero observational error) the continuous problem would require non-zero regularization and therefore the continuous solution of the inverse problem should differ from the true basal drag.

If it is not the true basal drag, what then is the continuous solution? It is useful to think of the continuous solution of the regularized inverse problem as the best knowable drag field. When λ_{best} approaches a non-zero limit, as we found, then the best knowable drag field is not the same as the true drag field, although it is related to the true field through a low-pass filtering operation: the best knowable field is the true field filtered for the components that have a detectable influence on the observations. For any given non-zero resolution of mesh and observations, the corner of the L-curve gives the most amount of structure that can be reliably inferred at that resolution, and taking the limit as resolution approaches zero gives the most amount of structure that can be reliably inferred from observations of that precision. It remains to be seen whether the best knowable field approaches the true field in the limit that observational error also approaches zero.

5.1 Sliding Nonlinearity

It is common in the inverse modeling literature to read some version of the statement that, because it is possible to fit the data and achieve force balance with any sliding law, we therefore cannot use inverse models to distinguish between different sliding laws. For instance, Joughin et al. (2004) state "the inversion results alone do not distinguish which is the most appropriate bed model", wording which is repeated almost verbatim in Joughin et al. (2006). Under this logic, inverse models can only be used

to discriminate between different sliding laws if they have access to time-dependent information, either by simultaneously fitting data from multiple time slices with a single coefficient field (Gillet-Chaulet et al., 2016) or by evaluating the ability of transient models forced by different sliding laws to fit time-dependent geometry and velocity data (Joughin et al., 2010a). Inversions of a single time slice are thought to be neutral between different sliding laws, so long as all of the competing sliding laws obtain a similar fit to the data.

However, we argue that inverse models are not neutral between different sliding laws, even if they are limited to a single time slice. Occam's Razor dictates that, when choosing between two parameterizations that both obtain a similar fit to the data, we should prefer the parameterization that requires less complexity to do so. In the case of a sliding law, we are creating a parameterization that relates two quantities that each have about 2–3 orders of magnitude of dynamic range: velocity varies between about 1 to about $1000\,\mathrm{m\,a^{-1}}$, and basal drag varies between a few kPa and a few hundred kPa. Yet we found that linear Weertman sliding requires a coefficient with twice as much dynamic range as either of those two parameters. This result is not exactly new or controversial. Almost all published inversions of real data with linear sliding laws require many orders of magnitude of dynamic range in C (e.g. Arthern et al., 2015, Fig. 7). Indeed, this result is so well-known in the glaciological modeling community that it has been incorporated into the design of some models: for example, Elmer/Ice uses $\alpha = log_{10}(C)$ as the free parameter in their inversions rather than C itself (Gagliardini et al., 2013). However, it is less common for glaciologists to reflect on what this result means for the utility of linear Weertman sliding. Good parameterizations help us represent complex processes by simplifying them in ways that isolate the key variables and reduce the unknowns. A sliding law that doubles the amount of unexplained variance at the ice sheet bed has fundamentally failed to be a useful parameterization.

Thus, we would argue that, rather than being agnostic about the value of the sliding exponent, inverse models actually provide evidence in favor of nonlinear sliding. Our total variance measure, which combines both observational misfit and coefficient variance, was better for nonlinear sliding as compared to linear sliding by about a factor of 3 over the whole domain, or a factor of 2 when the analysis was limited to the more challenging fast-flow region. In addition, our L-curves for nonlinear sliding were more sharply curved, producing a narrower and more well-defined range of acceptable λ values. Of course, this evidence from single-snapshot inverse models is not definitive on its own; if other lines of evidence supported linear Weertman sliding, then we would be forced to accept the large variance in C as the reality. However, other lines of evidence indicate the opposite. Laboratory tests on actual samples of subglacial till consistently reveal Coulomb plastic rheology for soft basal sediments (e.g. Kamb, 1991; Tulaczyk et al., 2000). Analytical and numerical process models for hard bed sliding with cavitation show Coulomb-like behavior with basal traction bounded by a maximum yield stress (Iken, 1981; Schoof, 2007; Gagliardini et al., 2007). Transient numerical modeling of the visco-elastic response of glaciers to tidal forcing suggests that nonlinear sliding is required to fit the response recorded with GPS observations (Gudmundsson, 2011). The previously mentioned inversions of time-dependent velocity and geometry data support nonlinear sliding (Gillet-Chaulet et al., 2016), as do transient models (Joughin et al., 2010a). Thus, the evidence from single-snapshot inversions adds to a convergence of many lines of evidence indicating that the best description of basal sliding is a nonlinear one.

5.2 Effective Pressure

While our inverse model results are strongly in favor of nonlinear basal sliding, they are more ambiguous on the utility of including effective pressure in a sliding rule. Other lines of evidence, including laboratory tests (Tulaczyk et al., 2000), field observations (Kamb et al., 1985; Andrews et al., 2014), and analytic and numerical process models (Iken, 1981; Schoof, 2007; Gagliardini et al., 2007), argue that basal water pressure plays an important role in regulating basal sliding. The importance of subglacial water in regulating basal sliding is by now the unambiguous consensus of the glaciological community (e.g. Cuffey and Paterson, 2010, Ch. 7). However, the question addressed by inverse models is a more subtle one: not whether basal water pressure plays a role in the real physics of ice sliding, which it undoubtedly does, but whether our models of basal water pressure are good enough to make N a useful addition to a large-scale sliding law.

When it comes to simple geometry-based calculations for N, our answer is an unequivocal *no*: the slight reduction in inversion variance produced by these N fields is insufficient to compensate for the increase in observational misfit, regardless of whether the analysis is performed for the whole grounded domain or for the fast-flow region alone. When it comes to the physically based N_{CUAS}, however, our results are more ambiguous. When considered over the whole domain, using N_{CUAS} with linear sliding reduced total variance by 13%, but when considered in the more challenging fast-flow region alone, it increased total variance by 2%. Both of these changes are small compared with the large reductions in total variance achieved by using $m > 1$. Using $m > 1$ along with N_{CUAS} produced improvements when compared with either linear Weertman or linear Budd sliding, but not when compared with nonlinear Weertman at the same value of m (Table 2). Thus, our results do not provide support for using N in basal sliding laws- yet. If one is forced to use a linear sliding relationship, and one is mostly concerned with capturing the large-scale pattern of slow and fast flow, then our results suggest that one could improve one's model by including a physically based N field like N_{CUAS}. However, the resulting improvement is much less than the improvement that one could get by moving from linear to nonlinear sliding, and if nonlinear sliding is combined with N_{CUAS}, the improvement vanishes.

Nonetheless, there remains hope that further improvements in hydrological models may change the situation. N_{CUAS} was unambiguously better than either of the two geometry-based N fields, demonstrating that the performance of inversions using Budd sliding is sensitive to the quality of the N field; it is thus reasonable to assume that further improvements in N will improve inversion performance further. Given that we found that the performance of nonlinear Budd laws with N_{CUAS} was very close to the performance of nonlinear Weertman laws at the same value of m (within 5–10%, measured in units of the total variance in the linear Weertman inversion, Table 2), it would not take much improvement for nonlinear Budd inversions to overtake nonlinear Weertman inversions. Furthermore, we know from the basic math of our sliding laws that there is a hard limit to the performance of nonlinear Weertman inversions: the coefficient in a nonlinear Weertman sliding law must always have at least as much dynamic range as the dynamic range of basal drag, which for typical ice sheet settings is at least two orders of magnitude. By including N, Budd sliding laws open up the possibility that the variance of the coefficient can drop below this limit, and indeed we found that in our inversion with N_{CUAS} and $m = 5$, the variance of the coefficient was slightly less than the variance of basal drag in the fast-flow region. The total variance for this model was still higher than for Weertman

sliding with $m = 5$ (Table 2), but the fact that one of our Budd models was able to dip below this limit is a hopeful sign that future improvements in subglacial hydrology models may explain more of the variance in basal drag. Put simply, nonlinear Weertman may be better than nonlinear Budd right now, but it also has a much lower ceiling. With $m = 5$, nonlinear Weertman is already within 10–20% of the best possible performance it can achieve, whereas sliding laws that incorporate effective pressure can be as good (or as bad) as the hydrology model they use as input.

5.3 Full Stokes, Higher Order, and Prospects for Future Inversons

In almost all of our experiments in this paper, we used the SSA stress balance equations, which are the simplest approximation that can be used in an inversion. It is tempting to assume that inversions using more advanced stress balance equations, such as HO or Full Stokes (FS), are inherently more powerful than inversions performed with the SSA equations. However, that is forward model thinking, not inverse model thinking. From an inverse modeling perspective, the extra stress and strain rate components found in HO and FS models are a liability, not an asset. If a simpler SSA model is able to obtain a similar fit to the data as a more complex HO or FS model, then it is the simpler model that should be preferred. But "similar fit" may actually be an optimistic case for the more complex models: in our experiments comparing HO and SSA using the same mesh and sliding law, we found that the SSA inversion was able to obtain a better fit to the data than the HO model. There is thus very little justification, from our results at least, for using the more complex stress balance equations in an inversion.

The proximal cause of the poor performance of our HO inversions was that the basal ice rheology was too warm in large portions of the domain. Remedying this by using a constant cold rheology at all depths throughout the domain brought the inversion performance closer to the performance of the SSA inversion. However, a spatially invariant constant is a very poor approximation of the ice sheet thermal structure; it ignores very real variations in thermal boundary conditions such as surface temperature, accumulation rate, geothermal flux, ice thickness, and strain heating. While we probably could have produced a realistic colder temperature field that enabled our HO inversions to match the performance of our SSA inversions with a bit of work, the mere fact that this tuning is required is a mark against the HO inversions. In addition, using a stiffer ice rheology near the bed has the effect of greatly reducing shear deformation, forcing the entire domain to shift towards a regime of plug flow. In other words, in order to improve the performance of our HO inversion, we had to force it into a regime where it mimicked SSA! This explanation may also help to explain why some previous studies, such as Shapero et al. (2016), have found that the results of 3D inversions were insensitive to ice rheology. In that study, the authors performed FS inversions in restricted domains around three of the largest and fastest outlet glaciers in Greenland: Jakobshavn, Helheim, and Kangerdlugssuaq. While they did include a small amount of stagnant ice around the margins, the overwhelming majority of their three domains was composed of ice moving at least several hundred meters per year, which usually indicates that flow is dominated by basal slip, or at least, that they are unlikely to be in the regime where slip has dropped to zero and the sliding inversion is powerless to reduce velocity further. While Shapero et al. (2016) did not compare the performance of their FS inversion with an SSA inversion, our prediction would be that, if they had performed a comparative L-curve analysis, they probably would have found that their FS inversion had roughly the same performance as an SSA inversion.

Fundamentally, the problem is that, by allowing for vertical variations in rheology, strain rate, and therefore velocity, HO and FS introduce an enormous amount of variability into the problem that cannot be constrained by the available datasets. Sliding inversions typically take the rheological structure of the ice sheet to be fixed, but properly speaking, an FS or HO inversion should be simultaneously adjusting the sliding coefficient and the rheology of the lower ice column. While glaciological inversions for vertically averaged rheology have been performed (e.g. Larour et al., 2005), we are not aware of any cases of glaciologists performing mechanical inversions for vertically resolved rheology. This is unsurprising, because the data we use to constrain ice sheet inverse models are essentially two dimensional: we have maps of the two components of horizontal velocity on the upper surface of the ice sheet. If the rheology of the lower ice column is soft enough to permit a deformation velocity greater than the observed surface velocity, then an HO or FS inversion has no remedy to fix this by adjusting sliding, and it will be incapable of fitting the data. Conversely, if the rheology of the lower ice column is too stiff to permit a rapid deformation velocity, then the ice sheet will move primarily by plug flow and the HO or FS inversion will merely recover the same performance that could have been achieved by an SSA inversion with far less internal complexity. In the absence of any constraint on variation in the third dimension, a two dimensional SSA model works best because it is a match to the two dimensional nature of the data.

It is interesting that we found that our SSA inversions even outperformed our HO inversions in slow-flowing areas of the ice sheet, where the SSA equations should not be a good approximation to the ice sheet stress balance. However, this is less surprising when considering the framework of the "hybrid" approximation to the stress balance equations (Schoof and Hindmarsh, 2010; Goldberg, 2011). In this approximation, shear deformation in the lower ice column is parameterized as a component of the sliding law, with ice velocity being decomposed into the sum of sliding and deformation flow, and basal drag being determined by whichever process (slip or deformation) is weaker. Crucially, the vertically integrated stress balance equations of this approximation are identical to the SSA equations; the only difference is that the effective viscosity computation is modified to include the contribution of vertical shear to the effective strain rate, and the sliding law is modified to reflect the fact that resistance to flow could come from shear deformation in the lower ice column instead of sliding over the bed. The hybrid approximation applies throughout the ice sheet, not merely in fast-flowing streams and shelves moving by plug flow, and because its stress balance equations are identical to those of SSA, an SSA inversion can therefore be regarded as representing a smooth transition between a sliding inversion in fast-flowing areas and an inversion for near-basal rheology in slow-flowing areas. Under this framework, our SSA inversion results for drag coefficient in slow-flowing regions could be reinterpreted as a measure of the integrated stiffness of the lower ice column rather than as a measure of sliding itself. But regardless of how we choose to interpret our results, the SSA equations themselves are agnostic as to whether the resistance to horizontal flow comes from slip across the basal interface or from shear deformation a short distance above, and this very agnosticism is the key to their success at fitting surface velocity data, which cannot distinguish between sliding and deformation either.

6 Best Practices

On the basis of the results and discussion above, we suggest the following principles for guiding the use and analysis of L-curves for regularization in glaciological inversions. Note that, while we have made a number of particular choices in our model setup here (e.g., the choice to use the L2 norm of absolute velocity misfit for J_{obs}), these recommendations apply for any use of L-curve analysis in glaciological inverse modeling.

- Whenever possible, modelers should use an *a priori* estimate of characteristic scale to normalize their individual cost function terms before those terms are combined. Doing so ensures that λ values are unitless and of order unity, making it easy to identify the region of parameter space to explore in the L-curve analysis. In addition, normalization is invaluable if multiple data types or misfit metrics are included in J_{obs}, or if multiple parameters are being solved for and therefore multiple regularization terms must be included in J_{reg}.

- L-curves should always be presented on log/log axes with equal axis scaling (ie, one order of magnitude on the y-axis should be the same size as one order of magnitude on the x-axis). Given that the entire L-curve analysis is predicated on being able to identify a "corner", it is vital that the L-curve is presented in such a way that the visual impression of shape can be trusted. Linear axes are not reliable for this purpose, as inverse power laws ($J_{\mathrm{obs}} \propto J_{\mathrm{reg}}^{p}$ for $p < 0$) always appear "L"-shaped on linear axes despite having no intrinsically preferable value.

- In addition to an honest visual presentation of L-curve shape, it is vital that the optimal λ value be chosen through a quantitative measure of curvature. The specific curve-fitting procedure we presented in Section A5 is, of course, not the only way to do this, nor do we even claim that it is necessarily optimal. However, it is vital that some sort of quantitative measure of curvature be used to identify the corner of the L-curve, rather than simple visual inspection. This curvature analysis should be performed in log/log space rather than linear space.

- Inversions with noisy, uncurved, nonmonotonic, or otherwise ill-formed L-curves should be regarded as suspect. Rather than simply selecting one λ value and moving on, modelers should view a poor L-curve as a sign that something needs to be fixed "under the hood": incomplete numerical convergence, suboptimal choice of solvers, inconsistent input datasets, excessively complex mask topology, a choice of λ samples that does not include both limbs of the L-curve, or other problems in the model set-up can all contribute to a poor L-curve. If an inversion is working correctly, then varying λ with all else held constant ought to produce a smooth monotonic progression of cost function terms.

- In an inverse modeling context a modeler can not simply assume that HO or FS models are superior to an SSA model simply because the former contain more complete representations of the stress tensor. In an inverse modeling context, the increased internal complexity and degrees of freedom in the 3D models are a liability that must be justified, for instance through comparative L-curve analysis (e.g., Fig. 11).

- The utility of sliding laws that incorporate effective pressure N is dependent on the quality of the N field used as input. Commonly used geometry-based calculations for N add no value, producing a slight reduction in coefficient variance

that is not sufficient to compensate for their slight increase in observational misfit. In contrast, we found that a physically based N field added value when considered over the entire domain, although not when the analysis is restricted to the more challenging fast-flow region. Our recommendation is that glaciological modelers should only use sliding laws that incorporate N if that N field is based on the results of an actual hydrology model.

- In contrast to the small marginal utility added by including N, moving from linear Weertman sliding to nonlinear sliding improves inversion performance substantially. Nonlinear sliding laws have more sharply cornered L-curves, a reduced range between $\lambda_{\max}$ and $\lambda_{\min}$, and up to a factor of three reduction in total inversion variance. Our recommendation is that glaciological modelers should always use nonlinear sliding if they are able. Modelers who perform an inversion with a linear Weertman law and then convert the coefficient over to a nonlinear law should perform an L-curve analysis with the linear Weertman law first; this conversion is likely to give reasonable results for $\lambda \gtrsim \lambda_{\text{best}}$ but not for smaller values of λ (Section B1).

7 Conclusions

Fundamentally, inverse models are a mathematical expression of principles of knowledge and inference. We have some aspects of the ice sheet that we can observe, such as the surface velocity, and we would like to use those observations to infer aspects that we cannot observe, such as the basal drag. However, we cannot infer all variables everywhere from a finite set of observations on a single surface of the ice sheet. It is therefore vital that we structure our inference engine to be skeptical of the inferred structure that it generates. Regularization is the mathematical expression of that skepticism, ensuring that our inversions only produce structure that is actually required to fit the data.

In this paper, we have given a tutorial on how to regularize glaciological inverse models, including normalization of cost function components and L-curve analysis. We have shown that, with non-zero observational error, the glaciological L-curve converges towards finite non-zero regularization values and a best knowable basal drag field in the continuous problem. It remains to be seen whether regularization approaches zero and the best knowable field approaches the true field for real glacier settings in the limit that observational error also approaches zero. We have shown how an L-curve analysis can enable a modeler to draw more rigorous conclusions about the short-wavelength structure of basal drag, and we have shown how the optimal regularization level on coarse meshes is intimately connected with numerical convergence of spatial structure with respect to element size. We have also advocated a change in philosophy for glaciological inverse models that centers the role of regularization in the process, with the goal of inverse modeling being explicitly understood as not merely fitting the data, but fitting the data using the least amount of structure. We have shown how this shift in philosophy allows inverse models to break their agnosticism on the question of sliding nonlinearity, coming down strongly on the side of nonlinear sliding laws while providing more ambiguous conclusions on the utility of incorporating effective pressure into the sliding law. This philosophy also provides a framework for thinking about the relative performance of different types of inversions, with more complex models being required to justify their increased degrees of freedom with an improved observational fit.

But while this shift in philosophy may favor simpler 2D models for inversions of 2D datasets, a fascinating future of inverse modeling may lie in using 3D models to assimilate a greater variety of information than that which can be included in 2D models. Surface horizontal velocity is just one of many observations glaciologists have that give us information about the ice sheet state. While inverse modelers have occasionally tried to fit time-variable surface elevations (e.g. Larour et al., 2014), it is not yet common to use dH/dt observations to estimate the vertical component of surface velocity and thus construct a 3D velocity vector to use as a constraint. Beyond that, we have a wealth of information derived from radar, such as pRES observations of depth-dependent vertical velocity, and RES measurements of internal layer geometry or basal reflection characteristics, not to mention geologic and geomorphic evidence of past ice sheet dynamics. Ultimately, we have a wealth of different data streams illuminating different aspects of ice sheet dynamics. The philosophy of fitting each data type according to its characteristic scale while simultaneously minimizing complexity in the model that does so provides the best framework for integrating many sources of knowledge into a coherent picture of the ice sheet.

Code and data availability. Inversion scripts, inverse model results, L-curve summaries, and our best combined drag estimate are available at DOI:10.5281/zenodo.7798650. The best combined drag estimate is presented interpolated onto a regular grid in netcdf format for ease of use. The ice flow model ISSM is open source and freely available at: https://issm.jpl.nasa.gov/ (Larour et al., 2012). Here ISSM version 4.18 was used.

Appendix A: Supplemental Methods

A1 Thermal Structure and Rheology

Our SSA model needs an estimate of the column-average ice rheology $\bar{B}$, our HO model needs an estimate of vertically resolved B, and the Confined-Unconfined Aquifer System (CUAS-MPI) hydrology model (Beyer et al., 2018; Fischler et al., 2023) also needs an estimate of basal melt rate. We provided all of these using a 1D vertical steady-state advection-diffusion thermal model. We show the boundary conditions used to force the model in Figure A1a-c. For surface temperature, we use Comiso (2000), and for accumulation rate, we use the mean of Arthern et al. (2006) and Van de Berg et al. (2005). All surface climate inputs were accessed through Le Brocq et al. (2010). We found that early iterations of our thermal model were too warm in the floating ice shelf, resulting in insufficient buttressing and ice flow that was uniformly too fast in the shelf. This likely resulted from the neglect of horizontal advection in the 1D thermal model, thus missing the advection of cold ice into the shelf from higher elevations inland. As an ad hoc fix we reduced surface temperatures by $10°$C at low elevations (below $\sim$200 m), resulting in a mix of positive and negative velocity misfits in the shelf. We do not include velocities from the floating shelf in our cost function, nor did we attempt to optimize ice shelf rheology; the $-10°$C shift was chosen because it is a round number, and we did not attempt to tune the shift further. The basal heat flux used to force the model (Fig. A1c) is the sum of geothermal heat flow and an estimate of shear heating: we use Martos et al. (2017) for geothermal heat flow, and we

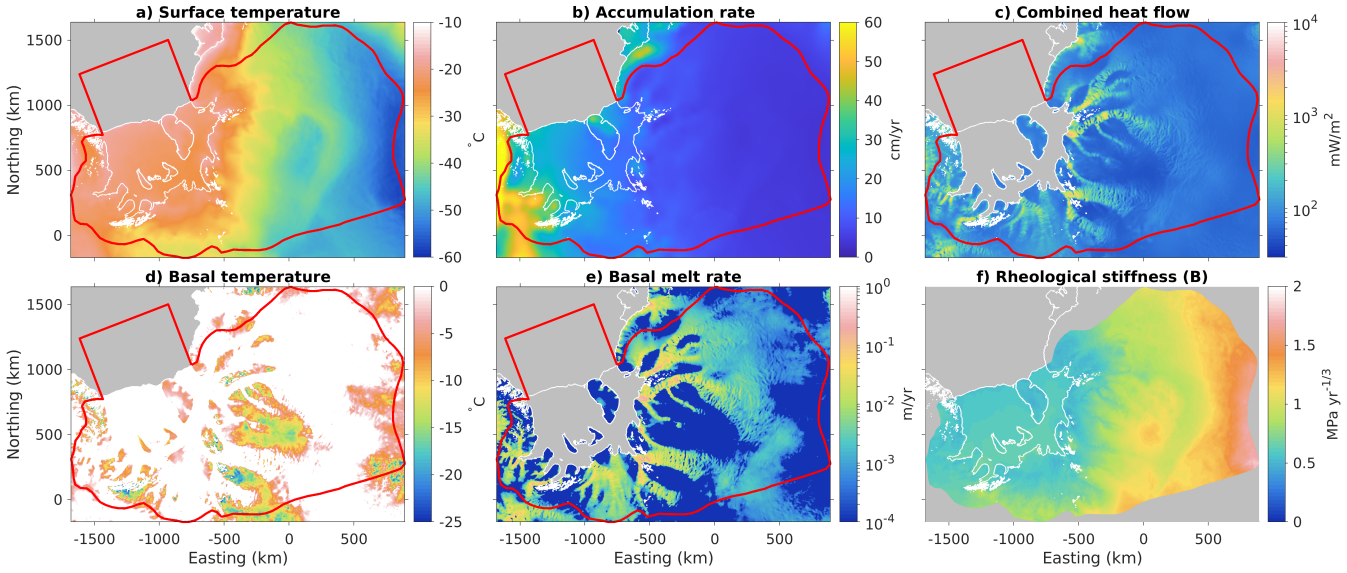

Figure A1. Thermal forcing and ice rheology. Top row (a–c) shows the boundary conditions (surface temperature, surface accumulation rate, and basal heat flow) used to force the 1D thermal model. Bottom row (d–f) shows model output: basal temperature, basal melt rate, and column-average rheological stiffness parameter. Basal temperature has been corrected for the pressure dependence of the melting point.

estimate shear heating by taking the dot product of driving stress and observed ice velocity (giving us the rate of dissipation of gravitational potential energy) and then smoothing at 10 km wavelength.

The resulting thermal structure is shown in Figure A1d–f. Basal temperature is at the pressure-dependent melting point in most of the domain, including almost all of the fast-flowing regions. Basal melt rate varies between about $0.1\,\mathrm{mm\,a^{-1}}$ and $1\,\mathrm{m\,a^{-1}}$, with values above $1\,\mathrm{cm\,a^{-1}}$ being exclusively due to high rates of shear heating. We use the thermal model to compute a vertically variable rheology stiffness parameter B based on Cuffey and Paterson (2010, p. 75), which we then average vertically to produce the $\bar{B}$ parameter needed by the SSA model (Fig. A1f). $\bar{B}$ varies spatially by about a factor of 4 between its largest and smallest values, with most of the spatial structure attributable to gradients in surface temperature (Fig. A1, c.f. plots a and f).

A2 Numerical Mesh

For our numerical mesh, we use an unstructured triangular mesh in the map-view plane, constructed using the Bidimensional Anisotropic Mesh Generator (BAMG, Hecht, 2006). As is common in finite element glaciological modeling, we want to refine our mesh in dynamically important regions, such as fast-flowing ice streams and outlet glaciers, or the grounding line and calving front. However, we also wish to investigate the influence of mesh resolution on our inversion results, which means that we need to create multiple meshes with different resolution but a similar spatial pattern of refinement. We do this by

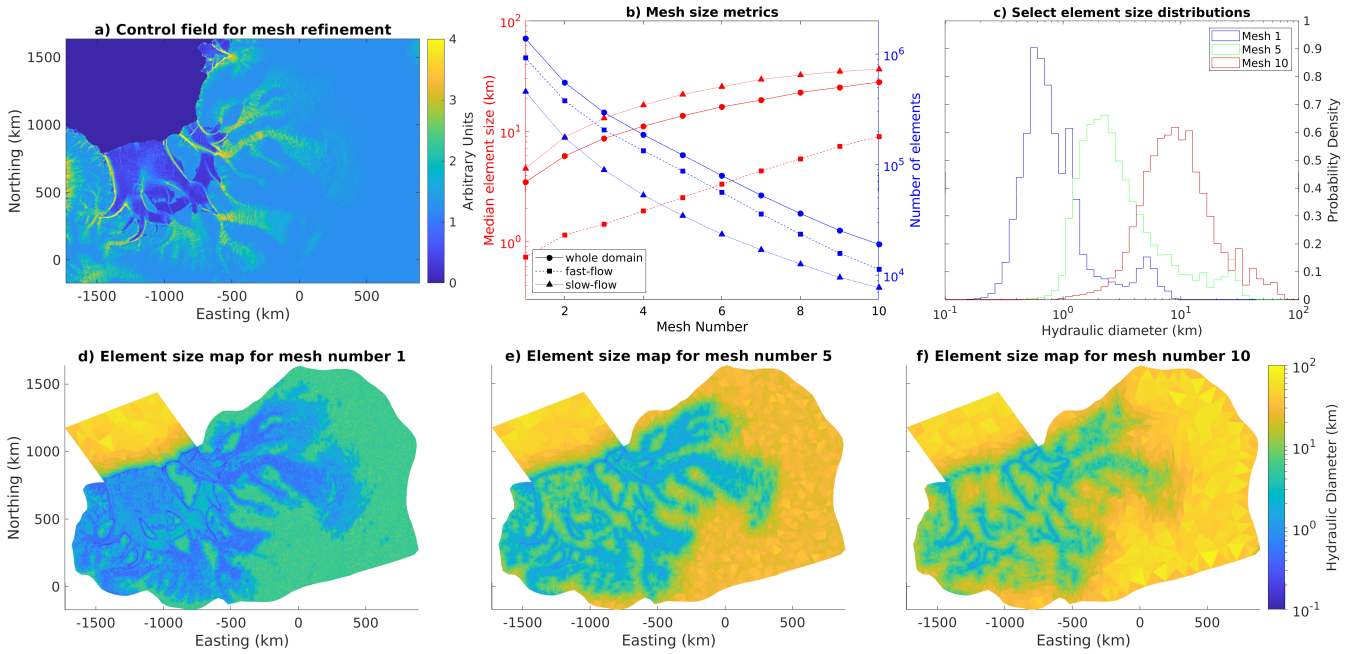

Figure A2. Element size analysis for our series of meshes. a) The common control field used to refine all of the meshes. b) Analysis of median element size and number of elements for each mesh in the series. Separate plots are shown for the entire grounded domain, as well as the grounded domain split into fast-flowing and slow-flowing regions. We define "fast-flow" as $\gtrsim 15\,\mathrm{m\,a^{-1}}$, with a smooth weighting function between $7.5\,\mathrm{m\,a^{-1}}$ and $30\,\mathrm{m\,a^{-1}}$. c) Selected histograms of element size within the grounded fast-flow domain. d–f) Maps of element size for the same selected meshes. Note that median size in (b) and probability density in (c) are defined with respect to domain area, not element number; ie, a median size of X means that 50% of the domain area is covered by elements smaller than X, not that 50% of the elements themselves are smaller than X.

constructing a common control field for all of our meshes, and then systematically varying parameters that control how the mesh is constructed from that control field.

We construct a control field on a 1 km regular grid using the logarithm of potential energy dissipation (the dot product of driving stress and ice velocity) in the grounded domain, the logarithm of strain-rate magnitude in the floating domain, and adding a constant multiple of the mask in order to produce discontinuities at the grounding line and calving front (Fig. A2a). We use BAMG to iteratively refine the mesh so as to minimize interpolation errors in the control field; thus, the mesh will be refined in areas where the control field has sharp curvature and the mesh will be unrefined in areas where the control field is smooth. We generate ten meshes using values of the BAMG parameters *err*, the acceptable level of interpolation error, ranging linearly from 0.01 to 0.1; values of *hmin*, the minimum edge length, ranging from 0.1 km to 1.0 km; and values of *hmax*, the maximum edge length, ranging from 10 km to 100 km.

The resulting meshes are summarized in Fig. A2. We use the hydraulic diameter (4 x area/perimeter) to be a representative linear size metric for our mesh elements, with the factor of 4 chosen such that the hydraulic diameter of a circle would be equal

to the regular diameter. All of our meshes share a common spatial pattern of refinement (Fig. A2d-f) that they inherited from the control field (Fig. A2a). We analyze fast-flowing and slow-flowing regions separately, using a smooth weighting function generated by applying a cosine taper to the logarithm of velocity. The 50% cutoff of the weighting function was $15\,\mathrm{m\,a^{-1}}$, with 0% – 100% limits at $7.5\,\mathrm{m\,a^{-1}}$ and $30\,\mathrm{m\,a^{-1}}$. These velocity thresholds are relatively low, but we choose them as approximate values roughly separating the beginning of streaming flow from slow flow. Element size in both the fast- and slow-flowing regions varies smoothly between different meshes, but element size in the fast-flow region is consistently ~ 7 times smaller than in the slow-flow region, and the mesher places about $2/3$ of the grounded elements within the fast-flowing region (Fig. A2b), despite the fact that the fast-flowing region is only about $1/5$ of the grounded area. An analysis of histograms of element size in the fast-flowing region reveals that our meshing procedure produces similar statistical distributions of element size for each mesh, simply shifted towards larger or smaller size (Fig. A2c). Although BAMG is an anisotropic mesh generator, in practice most of our mesh elements are fairly close to equilateral triangles, with the overwhelming majority of our grounded domain area being covered by elements having an anisotropy ratio less than 2 (Fig A3).

Once the meshes were generated, we interpolated all relevant fields (e.g., geometry, velocity) onto each mesh using a multi-wavelength technique, described below (Section A3). We reordered the mask into a series that behaves in an ice dynamically reasonable manner under smoothing and interpolation operations prior to interpolation (Section A4).

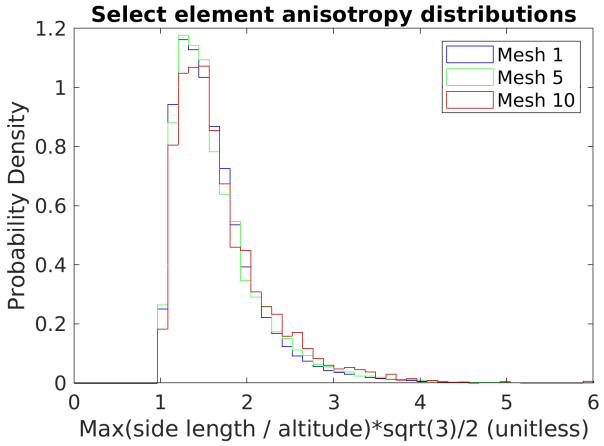

Figure A3. Element anisotropy for select meshes. This figure shows the distribution of element anisotropy for the same select 3 meshes shown in Figure A2. Element anisotropy is measured as the maximum ratio of side length to altitude, scaled such that an equilateral triangle has a value of 1. As in Figure A2c, probability density has been scaled by element area. Values shown are computed for the whole grounded domain.

A3 Multi-Wavelength Method for Data Interpolation

Our model setup requires that various data products be interpolated from their native regular square grids onto our unstructured triangular meshes. For instance, our ice geometry data is taken from BedMachine (Morlighem et al., 2020), which has a native

grid resolution of 500 m, although the underlying radar data used to generate the gridded product have a highly variable flight line spacing that is usually at least several km. Meanwhile, our model meshes have spatially variable resolution: for instance, Mesh #1 has a median element size less than $1\,\mathrm{km}$ in the fast-flow region, but an element size of $10\,\mathrm{km}$ in the slow-flow region. A simple naive interpolation, such as bilinear or spline, would therefore produce aliasing in slow-flowing areas with variable geometry, such as the Gamburtsev Subglacial Mountains. In such an area, mesh nodes that happened to fall in subglacial valleys would be assigned a very large ice thickness, while mesh nodes that happened to fall on subglacial peaks would be assigned a very small ice thickness, producing noisy and erroneous large-scale gradients in thickness. In reality, it would be better to assign each mesh node an intermediate value of ice thickness that was representative of a larger area comparable to the local element size. However, a simple naive interpolation from the original grid would still be appropriate in areas where the model mesh is fine; where element size is small, we would like to take advantage of the full information content of the gridded product without losing any information to smoothing.

Thus, we developed a multi-wavelength smoothing procedure for interpolating gridded data products onto our model mesh. The objective of this procedure is to ensure that each mesh node receives an interpolated value that is representative of the average of the original data grid over an area comparable to the local element size. Where mesh resolution is fine, mesh nodes receive values interpolated from a high-resolution grid, and where mesh resolution is course, the nodes receive values interpolated from a highly smoothed grid. The procedure is as follows:

1. First, we generate multiple copies of the original grid smoothed at a range of wavelengths. Starting from the original full-resolution grid, each subsequent grid is generated by twice smoothing the previous grid with a $3x3$ box filter and then decimating by a factor of 2 in both dimensions. Thus, the spatial resolution of each subsequent grid is a factor of 2 coarser than the previous grid (and the space used in memory is reduced by a factor of 4). This process of smoothing and downsampling is continued until we reach a resolution comparable to the coarsest elements in the mesh.

2. Next, we need to calculate a characteristic local element size for each mesh vertex. This is done by first computing the hydraulic radius for each element (half the hydraulic diameter, or 2 x area/perimeter), and then averaging hydraulic radius from the elements to the vertices.

3. Finally, we can interpolate the data from our grids to the mesh vertices. Each vertex is assigned to the first grid with a resolution coarser than its hydraulic radius. We loop through the grids, and for each grid, we perform a simple bilinear interpolation from that grid onto the vertices that were assigned to it.

This procedure produces an interpolated product that both avoids aliasing in coarse-resolution regions of the mesh and preserves high-resolution information where the mesh is fine. Additionally, while this method is not exactly conservative, it is approximately so: integrated ice volume within our domain computed on the mesh was within a factor of 10^{-4} of the value computed on the grid. We applied this method for all gridded data products that needed to be interpolated onto our meshes, including both components of the velocity observations. We applied a reordering to the BedMachine mask to ensure that it would behave logically under smoothing and interpolation operations, as described below. A Matlab function for performing this method (MultiWavelengthInterpolator.m) is available on our Zenodo code and data release page.

A4 Mask Reordering

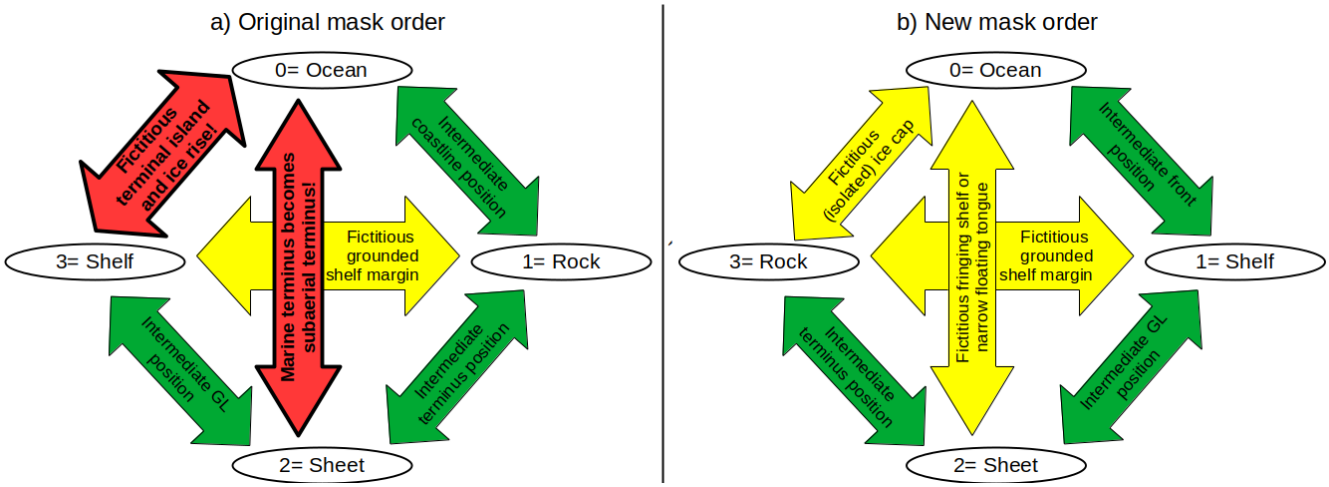

Figure A4. Flow chart showing all combinations of intermediate values for both the original (a) and the new (b) mask order. Combinations are colored green when intermediate values do not change the ice sheet configuration and can be simply interpreted as area fractions or intermediate positions, yellow when intermediate values produce new ice sheet configurations that do not have a significant influence on the force balance or overall dynamics of the ice sheet, and red when intermediate values they produce new ice sheet configurations that are extremely disruptive of the overall force balance or dynamics.

The smoothing and interpolation of mask values can be especially tricky, as there is not always a straightforward interpretation of intermediate values between the original integer classes. For instance, BedMachine has a mask variable with the following 5 values: 0- open ocean, 1- exposed rock, 2- grounded ice sheet, 3- floating ice shelf, and 4- Lake Vostok. Lake Vostok is not located within our domain, leaving us with 4 remaining integer mask values[1]. Often, modelers will interpolate mask values using nearest neighbor interpolation, thus ensuring that they only get integer results and they do not need to do the hard work of interpreting intermediate values. However, nearest neighbor interpolation will not necessarily produce values that are spatially representative. For instance, if a single mesh vertex happens to find the lone nunatak in an otherwise ice-covered region, then the apparent area of that nunatak in the model mesh would be much greater than reality. Our multi-wavelength interpolation procedure, described above, can produce values that are spatially representative, but these values will not necessarily be integer-valued and we thus need a way to interpret intermediate values of the interpolated mask.

Unfortunately, the original BedMachine mask order does not always behave in an ice dynamically reasonable manner under smoothing and interpolation operations. We consider behavior to be most "reasonable" when intermediate fractional mask

[1]If the mask reordering method described here is applied to a domain that includes Lake Vostok, then the lake should be given a mask value for either a floating ice shelf or the grounded ice sheet. If Lake Vostok is treated like a grounded ice sheet, then the bed elevation should be raised to the ice bottom and the friction coefficient set to zero, and if Lake Vostok is treated like a floating ice shelf then the local "sea level" needs to be raised to a value reflecting the actual level of hydrostatic equilibrium in the lake.

values can be easily interpreted as either area fractions of the respective regions or as an intermediate boundary position, for instance if an interpolated mask value in between grounded ice sheet and floating ice shelf can be simply interpreted as grounded area fraction. We consider behavior to be moderately unreasonable when intermediate values change the configuration but in a way that does not produce big disturbances in the force balance or overall dynamics, for instance if interpolation between a grounded ice sheet and the open ocean produces a narrow floating shelf at the terminus that provides no appreciable buttressing. And we consider behavior to be egregiously unreasonable if intermediate values change the ice sheet configuration in a way that produces a major change in force balance or dynamics, for instance if interpolation between floating ice shelf and open ocean produces a ring of grounded ice rises at the terminus.

The original mask order produces two egregiously unreasonable intermediate states (Fig A4a): interpolation between open ocean (0) and a floating ice shelf (3) produces fictitious rock islands (1) and grounded ice rises (2) at the terminus of the shelf, thus drastically increasing the amount of buttressing, and interpolation between grounded ice sheet (2) and open ocean (0) creates exposed rock (1), thus transforming a tidewater marine terminus into a grounded terminus. Additionally, interpolation between exposed rock (1) and a floating ice shelf (3) produces a thin region of grounded ice (2) around the fringe of the ice shelf, but we consider this to be only a moderately unreasonable change, since an ice shelf in contact with the rock walls of an embayment ought to be already experiencing buttressing.

However, the two egregiously unreasonable intermediate states in the original mask order can be eliminated by a simple reordering of the mask values. The only change required is that the numbers representing exposed rock and floating ice shelf are swapped, such that the mask now goes: 0- open ocean, 1- floating ice shelf, 2- grounded ice sheet, and 3- exposed rock. Under this order, no intermediate values are more than moderately unreasonable (Fig A4b) and none have a disruptive influence on the ice sheet force balance. Under this order, intermediate values at the ice shelf front (contact between 0 and 1) are easily interpreted as area fraction of ice or intermediate front position. Interpolation at a grounded terminus (contact between 0 and 2) will produce a small fictitious floating tongue (1), but this tongue should be unbuttressed and therefore will have little effect on the overall force balance. Interpolation between a floating shelf (1) and exposed rock (3) will have the same effect as before; the only new unreasonable configuration will be interpolation between exposed rock (3) and open ocean (0), which should produce a brand-new ice cap (2) and small floating shelf (1). However, this new ice cap will be isolated from the rest of the ice sheet (because it occurs at the contact between rock and ocean) and therefore will have no influence on the force balance of the true ice sheet.

This point is admittedly tangential to the main topic of our paper, but we find it striking that a simple reordering of the mask can produce behavior under smoothing and interpolation operations that is far more reasonable from an ice dynamic perspective than the original mask order. After reordering of the mask as described here, we used the same multi-wavelength interpolation procedure described above to interpolate the reordered mask onto the model mesh. Finally, we removed areas with an ice thickness less than $10\,\mathrm{m}$ from the active region of our inversion by setting those areas to either exposed rock or open ocean, depending on their bed elevation.

A5 Curve-Fitting for L-Curve Analysis

For our L-curve analysis, we sample the range $10^{-3} \leq \lambda \leq 10^{3}$ with 25 logarithmically spaced samples, resulting in 4 repeating samples per decade $(1, 1.8, 3.2, 5.6, 10, 18, 32...)$. For each sample, we run the inverse model to convergence using M1QN3 and record J_{obs} and J_{reg}. We used a stopping tolerance of 10^{-5} for the ratio of the gradient magnitude to the initial gradient magnitude, although we had to tighten this tolerance to 10^{-6} and 10^{-7} in order to produce monotonic L-curves for our inversions with $m = 3$ and $m = 5$, respectively. An example of the result is shown by the black dots in Fig. 3a.

Before selecting the optimal λ value, we next construct a smoothed tradeoff curve in $(ln(J_{\mathrm{reg}}), ln(J_{\mathrm{obs}}))$ space based on our 25 individual model results. Constructing a smoothed curve is necessary because we identify the corner of the L-curve by computing the curvature of our cost function components, and taking the 2nd derivative tends to amplify noise. Our smoothed curve is a parametric function of λ with two components, $J_{\mathrm{r,c}}$ and $J_{\mathrm{o,c}}$. To make the smoothed tradeoff curve, we first subsampled our 25 individual model results onto 1000 logarithmically spaced λ subsamples. We then tested 50 different smoothing wavelengths, ranging from a minimum of half the separation of the original model points to a maximum of $\ln(100)$. Note that all smoothing was performed in $\ln(\lambda)$ space.

The selection of the optimal smoothing wavelength in $\ln(\lambda)$ space can be regarded as a sort of meta-regularization: increased smoothing results in a worse fit between the tradeoff curve and the original model points, but reduced smoothing increases the random fluctuations in the curvature of the tradeoff curve. We would like to select a smoothing wavelength that balances these concerns. Thus, for each smoothing wavelength we computed the variance due to curvature of the smoothed curve and the variance due to scatter of the original model points about the smoothed curve. These variances are,

$$V_{\mathrm{s}} = \quad \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \left((J_{\mathrm{obs}} - J_{\mathrm{o,c}})^2 + (J_{\mathrm{reg}} - J_{\mathrm{r,c}})^2 \right), \tag{A1}$$

$$V_{\mathrm{c}} = \quad \int_{\lambda_{\min}}^{\lambda_{\max}} \left(\left(\tfrac{d^2 \ln(J_{\mathrm{o,c}})}{d \ln(\lambda)^2} \right)^2 + \left(\tfrac{d^2 \ln(J_{\mathrm{r,c}})}{d \ln(\lambda)^2} \right)^2 \right) d \ln(\lambda), \tag{A2}$$

where V_{s} and V_{c} are the scatter and curvature variances of the curves, $(J_{\mathrm{obs}}, J_{\mathrm{reg}})$ are the cost function components from the N individual inverse models, and $(J_{\mathrm{o,c}}, J_{\mathrm{r,c}})$ are the cost function components from the smoothed tradeoff curve. As in our original analysis, both V_{s} and V_{c} need to be normalized before they can be combined; we simply normalized each by their respective maximum values, then added them together and selected the wavelength that minimized the combined variance.

Once we have a smoothed tradeoff curve, we computed the total logarithmic curvature from $\frac{d^2 \ln(J_{\mathrm{o,c}})}{d \ln(\lambda)^2} + \frac{d^2 \ln(J_{\mathrm{r,c}})}{d \ln(\lambda)^2}$ (Fig. 3d). The location of maximum curvature represents the sharpest corner in the L-curve and the best-value λ. We also record where the total curvature drops to $1/2$ of its maximum value, in order to bracket the full range of the corner. Thus, we have three λ values: minimum acceptable, best, and maximum acceptable. The logarithmic uncertainty in these three λ values is given by $1/2$ the smoothing wavelength chosen above. The value of $1/2$ of maximum curvature that we used to define $\lambda_{\min}$ and $\lambda_{\max}$ is an arbitrary heuristic, and the arbitrariness of this choice introduces an additional degree of uncertainty into our estimate of those two λ values. For most of our L-curves, we found that the dropoff in total curvature below λ_{best} was quite steep (Fig. 3d), so the choice of $1/2$ as our threshold value should have little influence on our estimate of $\lambda_{\min}$. However, the dropoff in total curvature above λ_{best} was more gradual, so our estimated $\lambda_{\max}$ values should be regarded as more uncertain.

Appendix B: Supplemental Results

B1 Coefficient Conversion

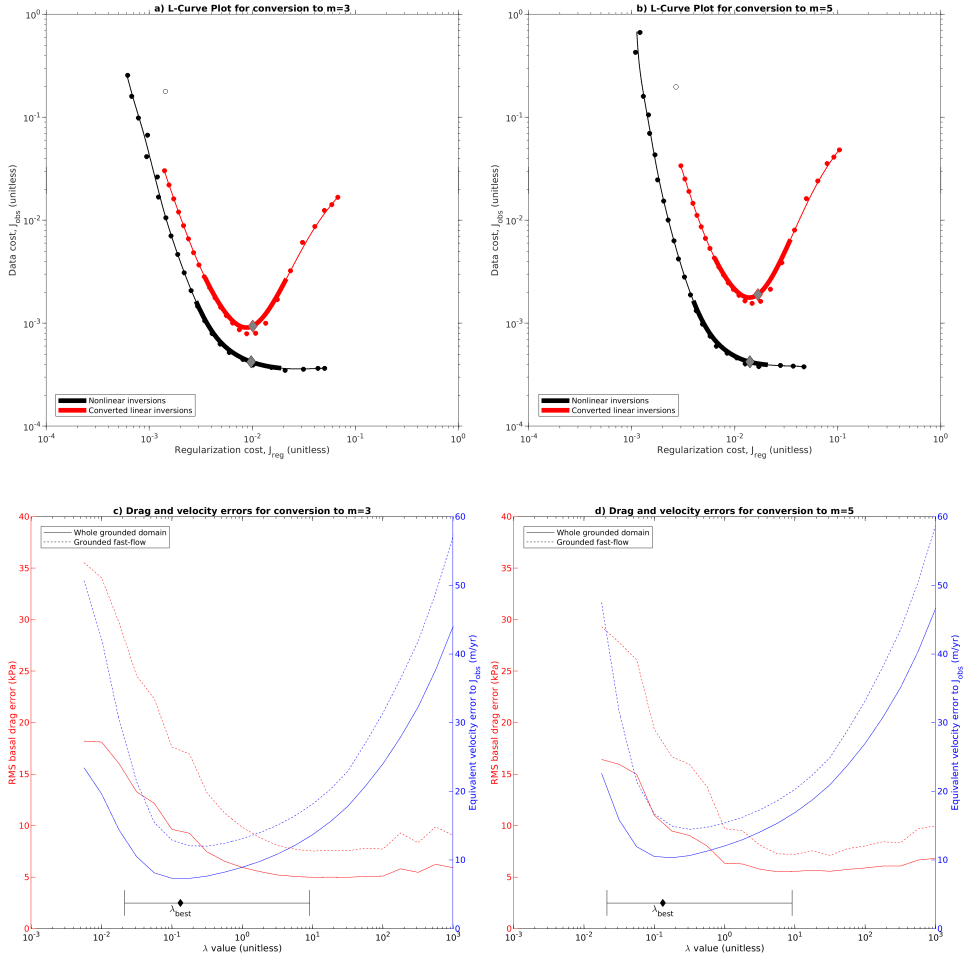

Figure B1. Effectiveness of converting the slip coefficient from a linear sliding law to a nonlinear one. a) Converted linear L-curve compared to the L-curve computed for $m = 3$. Display as in Figs 3, 8, 9, and 11. For the converted linear L-curve, the same smoothing wavelength has been used to construct the smoothed L-curve as was used for the original linear L-curve, and the indicated λ_{best} value and corner region are also those determined from the original linear L-curve. b) as (a), but for $m = 5$. c) Plot of errors in basal drag and ice velocity for the linear results converted to $m = 3$. Drag error has been computed by matching converted linear models with nonlinear models according to $\lambda/\lambda_{\mathrm{best}}$. Velocity error is computed with respect to the observations. d) as (c), but for $m = 5$.

Despite being a poor representation of reality (Section 5.1), linear Weertman sliding has the advantage of being the sliding law that is the easiest to implement numerically. Some ice sheet models, such as Elmer/Ice (Gagliardini et al., 2013) only

contain the ability to invert for a linear Weertman drag coefficient; if users of that model want to use a different sliding law constrained by observations, they must first perform an inversion for the linear Weertman coefficient and then convert that coefficient to the alternate sliding law. Thus, we think it is a useful experiment to test whether such a conversion does a good job of reproducing the inversion results and L-curve generated by the nonlinear sliding law.

We converted our linear Weertman drag coefficient to a nonlinear sliding law using,

$$C_m = C_W |\boldsymbol{u}_\mathrm{b}|^{\frac{m-1}{m}}, \tag{B1}$$

where C_m is the converted coefficient for the new exponent, and C_W is the coefficient for linear Weertman sliding. We performed this conversion separately for all 25 inversions in our linear Weertman L-curve on our highest resolution mesh. We tested the conversion from $m = 1$ to both $m = 3$ and $m = 5$. For each converted coefficient, we ran a single stress balance solve to get a new velocity field, from which we computed J_{obs}. We also computed J_{reg} for the converted coefficient using the same scale estimates we used for the corresponding nonlinear inversions. Thus, we are able to directly compare L-curve shape between the converted linear inversions and the fully nonlinear inversions.

As can be seen in Figure B1 a and b, the converted linear L-curve is nonmonotonic for both $m = 3$ and $m = 5$. For large λ, the slope of the converted linear L-curve roughly follows the large-λ limb of the nonlinear L-curve, although its position is shifted towards higher J_{reg} values (and we know this shift is meaningful because we used the same scale estimate for both L-curves). However, below λ_best, the converted linear L-curve turns around sharply. J_{obs} stops declining and begins rising again, eventually attaining values roughly as high as those in the large-λ limb. This sharp increase in velocity error for $\lambda < \lambda_\mathrm{best}$ is consistent regardless of whether the error is computed in the whole grounded domain or in the fast-flow region alone (Fig B1c,d). We also attempted to compare basal drag between the converted linear inversions and the nonlinear inversions. Such a comparison is more ambiguous because the nonlinear L-curves have a different value of λ_best, so we chose to match models between the nonlinear L-curves and the converted linear L-curves according to $\lambda/\lambda_\mathrm{best}$. Doing so reveals that the discrepancy in basal drag between the nonlinear results and the converted linear results also increases sharply for small λ (Fig B1c,d).

Thus, these results demonstrate that L-curve analysis can be a valuable tool even for modelers who intend to convert their coefficient from linear Weertman to a new sliding law. Conversion of the coefficient without performing an L-curve analysis first may produce unreliable results, especially for $\lambda < \lambda_\mathrm{best}$. However, for $\lambda \approx \lambda_\mathrm{best}$ the conversion is likely to give results that, while not quite as good as those obtained from a nonlinear inversion directly, are still broadly reasonable. Our recommendation for modelers performing coefficient conversion is that they should first construct an L-curve for their linear Weertman inversion, and then convert the coefficient obtained for λ_best.

Author contributions. MJW wrote inversion scripts, devised experimental design, made figures, and prepared the manuscript text. TK ran the CUAS hydrology model. AH and MR acquired funding and supervised the work. MR installed and maintained the ISSM installation on AWI's HPC system. All authors contributed to revising and improvements of the manuscript as well as discussions about ideas and conclusions.

Competing interests. The authors declare that we have no competing interests that would influence the conclusions drawn in this paper.

Acknowledgements. This work has been funded by AWI's INSPIRES project FRISio.

References

Andrews, L. C., Catania, G. A., Hoffman, M. J., Gulley, J. D., Luthi, M. P., Ryser, C., Hawley, R. L., and Neumann, T. A.: Direct observations of evolving subglacial drainage beneath the Greenland Ice Sheet, Nature, 514, 80–83, https://doi.org/10.1038/nature13796, 2014.

Arthern, R. J. and Gudmundsson, G. H.: Initialization of ice-sheet forecasts viewed as an inverse Robin problem, Journal of Glaciology, 56, 527–533, https://doi.org/10.3189/002214310792447699, publisher: Cambridge University Press, 2010.

Arthern, R. J., Winebrenner, D. P., and Vaughan, D. G.: Antarctic snow accumulation mapped using polarization of 4.3-cm wavelength microwave emission, Journal of Geophysical Research: Atmospheres, 111, D06 107, https://doi.org/10.1029/2004JD005667, 2006.

Arthern, R. J., Hindmarsh, R. C. A., and Williams, C. R.: Flow speed within the Antarctic ice sheet and its controls inferred from satellite observations, Journal of Geophysical Research: Earth Surface, 120, 1171–1188, https://doi.org/10.1002/2014JF003239, 2015.

Bamber, J. L., Layberry, R. L., and Gogineni, S. P.: A new ice thickness and bed data set for the Greenland ice sheet: 1. Measurement, data reduction, and errors, Journal of Geophysical Research: Atmospheres, 106, 33 773–33 780, https://doi.org/10.1029/2001JD900054, 2001.

Bamber, J. L., Gomez-Dans, J. L., and Griggs, J. A.: A new 1 km digital elevation model of the Antarctic derived from combined satellite radar and laser data - Part 1: Data and methods, The Cryosphere, 3, 101–111, https://doi.org/10.5194/tc-3-101-2009, 2009.

Bamber, J. L., Griggs, J. A., Hurkmans, R. T. W. L., Dowdeswell, J. A., Gogineni, S. P., Howat, I., Mouginot, J., Paden, J., Palmer, S., Rignot, E., and Steinhage, D.: A new bed elevation dataset for Greenland., The Cryosphere, 7, 499–510, https://doi.org/10.5194/tc-7-499-2013, 2013.

Barnes, J. M. and 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, 4291–4304, https://doi.org/10.5194/tc-16-4291-2022, 2022.

Beyer, S., Kleiner, T., Aizinger, V., Rückamp, M., and Humbert, A.: A confined–unconfined aquifer model for subglacial hydrology and its application to the Northeast Greenland Ice Stream, The Cryosphere, 12, 3931–3947, https://doi.org/10.5194/tc-12-3931-2018, 2018.

Blatter, H.: Velocity and stress fields in grounded glaciers- a simple algorithm for including deviatoric stress gradients, Journal of Glaciology, 41, 333–344, 1995.

Brondex, J., Gillet-Chaulet, F., and Gagliardini, O.: Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law, The Cryosphere, 13, 177–195, https://doi.org/10.5194/tc-13-177-2019, 2019.

Budd, W. F., Keage, P. L., and Blundy, N. A.: Empirical Studies of Ice Sliding, Journal of Glaciology, 23, 157–170, https://doi.org/10.1017/S0022143000029804, 1979.

Comiso, J. C.: Variability and trends in Antarctic surface temperatures from in situ and satellite infrared measurements, Journal of Climate, 13, 1674–1696, https://doi.org/10.1175/1520-0442, 2000.

Cuffey, K. and Paterson, W.: The Physics of Glaciers, Butterworth-Heineman/Elsevier, Burlington, MA, 4 edn., 2010.

Fahnestock, M., Scambos, T., Moon, T., Gardner, A., Haran, T., and Klinger, M.: Rapid large-area mapping of ice flow using Landsat 8, Remote Sensing of Environment, 185, 84–94, https://doi.org/https://doi.org/10.1016/j.rse.2015.11.023, 2016.

Fischler, Y., Kleiner, T., Bischof, C., Schmiedel, J., Sayag, R., Emunds, R., Oestreich, L. F., and Humbert, A.: A parallel implementation of the confined-unconfined aquifer system model for subglacial hydrology: design, verification, and performance analysis (CUAS-MPI v0.1.0), Geoscientific Model Development Discussions, 2023, 1–20, https://doi.org/10.5194/gmd-2022-312, 2023.

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Catania, G., Callens, D., Conway, H., Cook, A. J., Corr, H. F. J., Damaske, D., Damm, V., Ferraccioli, F., Fors-

berg, R., and Fujita, S.: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica., The Cryosphere, 7, 375 – 393, https://doi.org/10.5194/tc-7-375-2013, 2013.

Gagliardini, O., Cohen, D., Råback, P., and Zwinger, T.: Finite-element modeling of subglacial cavities and related friction law, Journal of Geophysical Research: Earth Surface, 112, https://doi.org/10.1029/2006JF000576, 2007.

Gagliardini, O., Zwinger, T., Gillet-Chaulet, F., Durand, G., Favier, L., de Fleurian, B., Greve, R., Malinen, M., Martín, C., Råback, P., Ruokolainen, J., Sacchettini, M., Schäfer, M., Seddik, H., and Thies, J.: Capabilities and performance of Elmer/Ice, a new-generation ice sheet model, Geoscientific Model Development, 6, 1299–1318, https://doi.org/10.5194/gmd-6-1299-2013, 2013.

Gillet-Chaulet, F., Gagliardini, O., Seddik, H., Nodet, M., Durand, G., Ritz, C., Zwinger, T., Greve, R., and Vaughan, D. G.: Greenland ice sheet contribution to sea-level rise from a new-generation ice-sheet model, The Cryosphere, 6, 1561–1576, https://doi.org/10.5194/tc-6-1561-2012, 2012.

Gillet-Chaulet, F., Durand, G., Gagliardini, O., Mosbeux, C., Mouginot, J., Rémy, F., and Ritz, C.: Assimilation of surface velocities acquired between 1996 and 2010 to constrain the form of the basal friction law under Pine Island Glacier, Geophysical Research Letters, 43, 10,311–10,321, https://doi.org/10.1002/2016GL069937, 2016.

Glen, J. W.: Rate of flow of polycrystalline ice, Nature, 172, 721–722, https://doi.org/10.1038/172721a0, 1953.

Goldberg, D. N.: A variationally derived, depth-integrated approximation to a higher-order glaciological flow model, Journal of Glaciology, 57, 157–170, 2011.

Gudmundsson, G. H.: Transmission of basal variability to a glacier surface, Journal of Geophysical Research: Solid Earth, 108, https://doi.org/10.1029/2002JB002107, 2003.

Gudmundsson, G. H.: Ice-stream response to ocean tides and the form of the basal sliding law, Cryosphere, 5, 259–270, https://doi.org/10.5194/tc-5-259-2011, 2011.

Habermann, M., Maxwell, D., and Truffer, M.: Reconstruction of basal properties in ice sheets using iterative inverse methods, Journal of Glaciology, 58, 795–808, https://doi.org/10.3189/2012jog11j168, publisher: Cambridge University Press (CUP), 2012.

Habermann, M., Truffer, M., and Maxwell, D.: Changing basal conditions during the speed-up of Jakobshavn Isbrae, Greenland, The Cryosphere, 7, 1679–1692, https://doi.org/10.5194/tc-7-1679-2013, 2013.

Hansen, P. C.: Analysis of Discrete Ill-Posed Problems by Means of the L-Curve, SIAM Review, 34, 561–580, https://doi.org/10.1137/1034115, publisher: Society for Industrial & Applied Mathematics (SIAM), 1992.

Hansen, P. C. and O'Leary, D. P.: The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, SIAM Journal on Scientific Computing, 14, 1487–1503, https://doi.org/10.1137/0914086, publisher: Society for Industrial & Applied Mathematics (SIAM), 1993.

Hecht, F.: BAMG: Bidimensional Anisotropic Mesh Generator, Tech. rep., FreeFem++, https://www.ljll.math.upmc.fr/hecht/ftp/bamg, 2006.

Hofstede, C., Beyer, S., Corr, H., Eisen, O., Hattermann, T., Helm, V., Neckel, N., Smith, E. C., Steinhage, D., Zeising, O., and Humbert, A.: Evidence for a grounding line fan at the onset of a basal channel under the ice shelf of Support Force Glacier, Antarctica, revealed by reflection seismics, The Cryosphere, 15, 1517–1535, https://doi.org/10.5194/tc-15-1517-2021, 2021.

Iken, A.: The effect of the subglacial water pressure on the sliding velocity of a glacier in an idealized numerical model, Journal of Glaciology, 27, 407–421, 1981.

Joughin, I., MacAyeal, D., and Tulaczyk, S.: Basal shear stress of the Ross ice streams from control method inversions, Journal of Geophysical Research: Solid Earth, 109, https://doi.org/10.1029/2003JB002960, 2004.

Joughin, I., Bamber, J. L., Scambos, T., Tulaczyk, S., Fahnestock, M., and MacAyeal, D. R.: Integrating satellite observations with modelling: basal shear stress of the Filcher-Ronne ice streams, Antarctica, Philosophical Transactions of the Royal Society A- Mathematical, Physical, and Engineering Sciences, 364, 1795–1814, https://doi.org/10.1098/rsta.2006.1799, 2006.

Joughin, I., Tulaczyk, S., Bamber, J. L., Blankenship, D., Holt, J. W., Scambos, T., and Vaughan, D. G.: Basal conditions for Pine Island and Thwaites Glaciers, West Antarctica, determined using satellite and airborne data, Journal of Glaciology, 55, 245–257, https://doi.org/10.3189/002214309788608705, 2009.

Joughin, I., Smith, B. E., and Holland, D. M.: Sensitivity of 21st century sea level to ocean-induced thinning of Pine Island Glacier, Antarctica, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044819, 2010a.

Joughin, I., Smith, B. E., Howat, I. M., Scambos, T., and Moon, T.: Greenland flow variability from ice-sheet-wide velocity mapping, Journal of Glaciology, 56, 415–430, https://doi.org/10.3189/002214310792447734, 2010b.

Kamb, B.: Rheological nonlinearity and flow instability in the deforming bed mechanism of ice stream motion, Journal of Geophysical Research: Solid Earth, 96, 16 585–16 595, https://doi.org/10.1029/91JB00946, 1991.

Kamb, B., Raymond, C. F., Harrison, W. D., Engelhardt, H., Echelmeyer, K. A., Humphrey, N., Brugman, M. M., and Pfeffer, T.: Glacier Surge Mechanism: 1982-1983 Surge of Variegated Glacier, Alaska, Science, 227, 469–479, https://doi.org/10.1126/science.227.4686.469, 1985.

Kazmierczak, E., Sun, S., Coulon, V., and Pattyn, F.: Subglacial hydrology modulates basal sliding response of the Antarctic ice sheet to climate forcing, The Cryosphere, 16, 4537–4552, https://doi.org/10.5194/tc-16-4537-2022, 2022.

Larour, E., Rignot, E., Joughin, I., and Aubry, D.: Rheology of the Ronne Ice Shelf, Antarctica, inferred from satellite radar interferometry data using an inverse control method, Geophysical Research Letters, 32, https://doi.org/10.1029/2004GL021693, 2005.

Larour, E., Seroussi, H., Morlighem, M., and Rignot, E.: Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM), Journal of Geophysical Research: Earth Surface, 117, n/a–n/a, https://doi.org/10.1029/2011JF002140, 2012.

Larour, E., Utke, J., Csatho, B., Schenk, A., Seroussi, H., Morlighem, M., Rignot, E., Schlegel, N., and Khazendar, A.: Inferred basal friction and surface mass balance of the Northeast Greenland Ice Stream using data assimilation of ICESat (Ice Cloud and land Elevation Satellite) surface altimetry and ISSM (Ice Sheet System Model), The Cryosphere, 8, 2335–2351, https://doi.org/10.5194/tc-8-2335-2014, 2014.

Le Brocq, A. M., Payne, A. J., and Vieli, A.: An improved Antarctic dataset for high resolution numerical ice sheet models (ALBMAP v1), Earth System Science Data, 2, 247–260, https://doi.org/10.5194/essd-2-247-2010, 2010.

Lythe, M. B. and Vaughan, D. G.: BEDMAP: A new ice thickness and subglacial topographic model of Antarctica, Journal of Geophysical Research: Solid Earth, 106, 11 335–11 351, https://doi.org/10.1029/2000JB900449, 2001.

MacAyeal, D.: Large-scale ice flow over a viscous basal sediment: Theory and application to ice stream B, Antarctica, Journal of Geophysical Research: Solid Earth, 94, 4071–4087, https://doi.org/10.1029/JB094iB04p04071, 1989.

MacAyeal, D.: A tutorial on the use of control methods in ice-sheet modeling, Journal of Glaciology, 39, 91–98, 1993.

Martin, N. and Monnier, J.: Adjoint accuracy for the full Stokes ice flow model: limits to the transmission of basal friction variability to the surface, The Cryosphere, 8, 721–741, https://doi.org/10.5194/tc-8-721-2014, 2014.

Martos, Y. M., Catalán, M., Jordan, T. A., Golynsky, A., Golynsky, D., Eagles, G., and Vaughan, D. G.: Heat Flux Distribution of Antarctica Unveiled, Geophysical Research Letters, 44, 11,417–11,426, https://doi.org/10.1002/2017GL075609, 2017.

Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., and Aubry, D.: Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica, Geophysical Research Letters, 37, n/a–n/a, https://doi.org/10.1029/2010GL043853, 2010.

Morlighem, M., Seroussi, H., Larour, E., and Rignot, E.: Inversion of basal friction in Antarctica using exact and incomplete adjoints of a higher-order model, Journal of Geophysical Research: Earth Surface, 118, 1746–1753, https://doi.org/10.1002/jgrf.20125, 2013.

Morlighem, M., Rignot, E., Mouginot, J., Seroussi, H., and Larour, E.: Deeply incised submarine glacial valleys beneath the Greenland ice sheet, Nature Geoscience, 7, 418–422, https://doi.org/10.1038/ngeo2167, 2014.

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., Goel, V., Greenbaum, J. S., Gudmundsson, H., Guo, J., Helm, V., Hofstede, C., Howat, I., Humbert, A., Jokat, W., Karlsson, N. B., Lee, W. S., Matsuoka, K., Millan, R., Mouginot, J., Paden, J., Pattyn, F., Roberts, J., Rosier, S., Ruppel, A., Seroussi, H., Smith, E. C., Steinhage, D., Sun, B., Broeke, M. R. v. d., Ommen, T. D. v., Wessem, M. v., and Young, D. A.: Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet, Nature Geoscience, 13, 132–137, https://doi.org/10.1038/s41561-019-0510-8, 2020.

Pattyn, F.: Transient glacier response with a higher-order numerical ice-flow model, Journal of Glaciology, 48, 467–477, https://doi.org/10.3189/172756502781831278, 2002.

Rignot, E., Mouginot, J., and Scheuchl, B.: Ice flow of the Antarctic Ice Sheet, Science, 333, 1427–1430, https://doi.org/10.1126/science.1208336, 2011.

Rignot, E., Mouginot, J., and Scheuchl, B.: MEaSUREs InSAR-Based Antarctica Ice Velocity Map, Version 2, DOI:10.5067/D7GK8F5J8M8R, 2017.

Schoof, C.: Cavitation on Deformable Glacier Beds, SIAM journal on applied mathematics, 67, 1633–1653, https://doi.org/10.1137/050646470, 2007.

Schoof, C. and Hindmarsh, R. C. A.: Thin-Film Flows with Wall Slip: An Asymptotic Analysis of Higher Order Glacier Flow Models, The Quarterly Journal of Mechanics and Applied Mathematics, 63, 73–114, https://doi.org/10.1093/qjmam/hbp025, 2010.

Sergienko, O. V. and Hindmarsh, R. C. A.: Regular patterns in frictional resistance of ice-stream beds seen by surface data inversion, Science, 342, 1086–1089, https://doi.org/10.1126/science.1243903, 2013.

Sergienko, O. V., Creyts, T. T., and Hindmarsh, R. C. A.: Similarity of organized patterns in driving and basal stresses of Antarctic and Greenland ice sheets over extensive areas of basal sliding, Geophysical Research Letters, 41, 3925–3932, https://doi.org/10.1002/2014GL059976, 2014.

Seroussi, H., Nowicki, S., Simon, E., Abe-Ouchi, A., Albrecht, T., Brondex, J., Cornford, S., Dumas, C., Gillet-Chaulet, F., Goelzer, H., Golledge, N. R., Gregory, J. M., Greve, R., Hoffman, M. J., Humbert, A., Huybrechts, P., Kleiner, T., Larour, E., Leguy, G., Lipscomb, W. H., Lowry, D., Mengel, M., Morlighem, M., Pattyn, F., Payne, A. J., Pollard, D., Price, S. F., Quiquet, A., Reerink, T. J., Reese, R., Rodehacke, C. B., Schlegel, N.-J., Shepherd, A., Sun, S., Sutter, J., Van Breedam, J., van de Wal, R. S. W., Winkelmann, R., and Zhang, T.: initMIP-Antarctica: an ice sheet model initialization experiment of ISMIP6, The Cryosphere, 13, 1441–1471, https://doi.org/10.5194/tc-13-1441-2019, 2019.

Shapero, D. R., Joughin, I. R., Poinar, K., Morlighem, M., and Gillet-Chaulet, F.: Basal resistance for three of the largest Greenland outlet glaciers, Journal of Geophysical Research: Earth Surface, 121, 168–180, https://doi.org/10.1002/2015JF003643, 2016.

Tikhonov, A. and Arsenin, V.: Solutions of Ill-Posed Problems, Winston, Washington, 1977.

Tsai, V. C. and Gudmundsson, G. H.: An improved model for tidally modulated grounding-line migration, Journal of Glaciology, 61, 216–222, https://doi.org/10.3189/2015JoG14J152, 2015.

Tulaczyk, S., Kamb, W. B., and Engelhardt, H. F.: Basal mechanics of Ice Stream B, West Antarctica: 1. Till mechanics, Journal of Geophysical Research: Solid Earth, 105, 463–481, https://doi.org/10.1029/1999JB900329, 2000.

Van de Berg, W., Van den Broeke, M., Reijmer, C., and Van Meijgaard, E.: Characteristics of the Antarctic surface mass balance, 1958-2002, using a regional atmospheric climate model, Annals of Glaciology, 41, 97–104, https://doi.org/10.3189/172756405781813302, 2005.

Vogel, C. R.: Non-convergence of the L-curve regularization parameter selection method, Inverse Problems, 12, 535, https://doi.org/10.1088/0266-5611/12/4/013, 1996.

Weertman, J.: On the sliding of glaciers, Journal of Glaciology, 3, 33–38, 1957.