



Parameter and state estimation with a time-dependent adjoint marine ice sheet model

D. N. Goldberg and P. Heimbach

Earth, Atmospheric and Planetary Sciences, M.I.T., Cambridge, USA.

Correspondence to: D. N. Goldberg (dgoldber@mit.edu)

Received: 16 May 2013 – Published in The Cryosphere Discuss.: 17 June 2013

Revised: 30 September 2013 – Accepted: 10 October 2013 – Published: 6 November 2013

Abstract. To date, assimilation of observations into large-scale ice models has consisted predominantly of time-independent inversions of surface velocities for basal traction, bed elevation, or ice stiffness, and has relied primarily on analytically derived adjoints of glaciological stress balance models. To overcome limitations of such “snapshot” inversions – i.e., their inability to assimilate time-dependent data for the purpose of constraining transient flow states, or to produce initial states with minimum artificial drift and suitable for time-dependent simulations – we have developed an adjoint of a time-dependent parallel glaciological flow model. The model implements a hybrid shallow shelf–shallow ice stress balance, solves the continuity equation for ice thickness evolution, and can represent the floating, fast-sliding, and frozen bed regimes of a marine ice sheet. The adjoint is generated by a combination of analytic methods and the use of algorithmic differentiation (AD) software. Several experiments are carried out with idealized geometries and synthetic observations, including inversion of time-dependent surface elevations for past thicknesses, and simultaneous retrieval of basal traction and topography from surface data. Flexible generation of the adjoint for a range of independent uncertain variables is exemplified through sensitivity calculations of grounded ice volume to changes in basal melting of floating and basal sliding of grounded ice. The results are encouraging and suggest the feasibility, using real observations, of improved ice sheet state estimation and comprehensive transient sensitivity assessments.

1 Introduction

Simulation of land ice evolution is hampered by a great number of sources of uncertainties regarding poorly or unknown quantities which exert control over ice dynamics. These uncertainties must be dealt with to address questions regarding the estimation of past ice flow and future behavior of ice sheets and glaciers. Unknown quantities often take the form of spatial fields rather than scalars, requiring computational techniques that can handle sets of unknowns which scale with model dimension, but with computational costs largely independent of that dimension. The adjoint or Lagrange multiplier method is an ideal candidate.

The adjoint of a (generally nonlinear) model is essentially the transpose of its Jacobian, the Jacobian being the linear map of perturbations to model input (e.g., its initial and boundary conditions) to perturbations in output, for a given model realization. Assuming the model is differentiable, the Jacobian can, in principle, be estimated with finite differences. However, when the size of the parameter set numbers in the thousands or millions, as is the case in land ice models, such an approach quickly becomes intractable. With the adjoint method, on the other hand, gradient or sensitivity information of the model output with respect to its input is obtained efficiently. Its computational cost is independent of the size of the input set, making gradient-based optimization/estimation for large-scale problems with high-dimensional input spaces tractable.

Model adjoints have been valuable tools in climate research for a number of years (see Errico (1997) and Wunsch and Heimbach (2013) for partial reviews in the context of meteorology and oceanography). In the field of land ice modeling, adjoint methods have, until recently, been used

almost exclusively to estimate basal or interior properties of ice sheets and ice shelves (e.g., basal traction parameters, stiffness parameters) based on observations of surface velocities. The first such application was by MacAyeal (1992), who used a depth-integrated stress balance (Morland, 1987; MacAyeal, 1989). In recent years, similar inversions have been carried out with so-called “higher-order” models, i.e., models that incorporate vertical inhomogeneity and even nonhydrostatic effects (e.g., Maxwell et al., 2008; Raymond and Gudmundsson, 2009; Goldberg and Sergienko, 2011; Morlighem et al., 2010; Arthern and Gudmundsson, 2010; Gillet-Chaulet et al., 2012; Petra et al., 2012). However, all these inversions consider only the nonlinear stress balance; i.e., they do not take the time-dependent mass balance into account. Furthermore, they only consider uncertainties in surface velocities (not, for instance, in surface elevation or thickness). Brinkerhoff et al. (2011) applied an adjoint method to a flowline Stokes model coupled to a steady-state thermal balance, but there was no time dependence in the model. An example of the use of a time-dependent ice model is that of Heimbach and Bugnion (2009), in which the adjoint of SICOPOLIS (Greve, 1997), a thermo-mechanical ice model which makes the shallow ice approximation (SIA; Hutter, 1983), was generated.

A natural extension to the work of Heimbach and Bugnion, then, is the application of adjoint methods to time-dependent ice models that include horizontal stress coupling in the nonlinear momentum balance. Such an approach has the advantage that the prognostic components of the ice model, such as thickness and temperature evolution, are accounted for in the model adjoint, thus enabling assimilation of time-dependent data to produce a dynamically consistent state estimate with associated optimized parameters. In the field of ocean modeling, state estimation efforts based on the adjoint method were first introduced by Thacker and Long (1988) and Tziperman and Thacker (1989). Since then, estimation systems, in which the adjoint was derived by means of automated differentiation of full-fledged ocean general circulation models, have provided solutions that are consistent with observational data, suitable for model initializations and in-depth data analysis, as well as a framework for estimating the information content of new observations (Stammer et al., 2002; Wunsch et al., 2009; Wunsch and Heimbach, 2013). While ice sheet state estimation is still in its infancy, we view ocean state estimation as a model paradigm, and time-dependent ice model adjoints as a step toward this goal.

In this paper, we present the adjoint generation of a time-dependent ice sheet–stream–shelf model. The ice model implements a “hybrid” stress balance (e.g., Bueler and Brown, 2009; Pollard and DeConto, 2009; Schoof and Hindmarsh, 2010; Goldberg, 2011), which is the simplest form of higher-order stress balance, yet still it accounts for horizontal stress coupling, which makes our approach novel. Prior to this study, the adjoint method has not been applied to a time-dependent ice model with a nonlocal stress balance. Ray-

mond and Gudmundsson (2009) developed a method for a maximum a posteriori (MAP) estimate of basal properties in terms of observed surface properties, taking the steady state of the continuity equation into account. Their forward model contains the full Stokes stress balance without approximation. However, it is in a single horizontal dimension, and is based on analytical transfer functions that assume Newtonian rheology as well as small perturbations about a mean state (Gudmundsson, 2003). In addition to the conceptual novelty of using a time-evolving adjoint model for inversion, a technical novelty of our study is the use of algorithmic (or automatic) differentiation (AD) (Griewank and Walther, 2008) for an ice model that involves longitudinal (nonlocal) stress balance terms. AD tools are a powerful array of software that are capable of generating adjoint model source code via line-by-line differentiation of the numerical model. The pliability of the AD tools means that sensitivities to diverse sets of inputs can be examined, and changes to the input set, the cost function, and the forward model can be reflected in the adjoint much more easily than through separate error-prone by-hand adjoint code extensions. Finally, the approach provides the exact adjoint of the discrete model. In some of the glaciological studies mentioned above, the nonlinear dependence of viscosity on strain rates has been ignored in the adjoint calculation, which is generally a safe approximation, but in some cases has been shown to cause difficulties with model inversion (Goldberg and Sergienko, 2011).

The ice sheet stress balance equations (depth-integrated or otherwise) have a differentiable structure which considerably simplifies the derivation of their adjoint. This, and the fact that they do not involve time stepping, enables straightforward discretization of the adjoint differential equations. This option is, in practice, much simpler than applying AD, which is not well-suited to iterative methods, such as those used to solve the nonlinear elliptic partial differential equations of the ice models. On the other hand, other conservation equations involved in solving the ice flow, such as the thickness (or continuity) equation or conservation of heat, only solve for local interdependencies and are well suited for AD. The adjoint of Heimbach and Bugnion, as well, was generated with AD tools – the only instance of AD application to a land ice model. Here, we shall adopt a mix-and-match strategy, using both AD-generated code and “hand-written” solvers where this appears more efficient to produce the exact model adjoint.

In the following, we briefly describe the forward model used in our study, and then discuss the application of the AD tool to the model, notably the steps taken to deal with the nonlocal stress balance. We then proceed to demonstrate the utility of the adjoint through several idealized sensitivity calculations as well as state and parameter estimation experiments.

Table 1. Pseudocode version of forward model time-stepping procedure.

```

DO    $t_n = 0, \dots, t_f$ 
    • CALL  UPDATE_FLOATATION
        Evaluate floatation condition
        
$$s^{(n)} = \begin{cases} (1 - \frac{\rho_w}{\rho})h^{(n)} & \text{if } h^{(n)} \leq -\frac{\rho_w}{\rho}R \\ R + h^{(n)} & \text{otherwise} \end{cases}$$

    • CALL  VEL_SOLVE
        Find  $\mathbf{u}^{(n)}$ 
        • CALL  CALC_DRIVING_STRESS
            Evaluate  $\boldsymbol{\tau}_d = \rho g h \nabla s$  based on  $h^{(n)}, s^{(n)}$ 
        • DO    $m = 0, \dots, \text{max\_iter\_nl}$ 
            • CALL  BUILD_STRESS_MATRIX
                 $\mathbf{A}_m$  constructed
                from  $\nu^{(m)}, \beta_{\text{eff}}^{(m)}$ 
            • CALL  STRESS_CG_SOLVE
                 $\mathbf{u}^{(m+1)} = \mathbf{A}_m^{-1} \boldsymbol{\tau}_d$ 
            • CALL  UPDATE_VISC_BETA
                 $\nu^{(m+1)}, \beta_{\text{eff}}^{(m+1)}$  found from  $\mathbf{u}^{(m+1)}$ 
            • [check for convergence]
                if converged,  $\mathbf{u}^{(n)} = \mathbf{u}^{(m)}$ 
        • END   DO
    • CALL  ADVECT_THICKNESS
         $h^{(n+1)}$  found from continuity
        equation with  $\mathbf{u}^{(n)}$ 
END   DO
    
```

2 Forward model

The ice model used in this study is an extension of the stress balance solution presented in Goldberg (2011). It is a hybrid model, also referred to as L1L2 under the Hindmarsh (2004) classification, meaning it accounts for vertical shear in its stress balance, although not as completely as the Blatter–Pattyn balance (Blatter, 1995; Pattyn, 2003; Greve and Blatter, 2009). On the other hand, the balance requires the solution of a two- rather than three-dimensional nonlinear elliptic differential equation, greatly reducing computational expense. The balance is derived by making an approximation to the variational principle corresponding to the Blatter–Pattyn equations rather than to the equations themselves. It has been demonstrated to be a good approximation to Blatter–Pattyn and to Stokes flow (Sergienko, 2012), especially when some level of basal sliding is present. In addition, the model solves the depth-integrated continuity equation for ice thickness and accounts for grounded and floating ice through a hydrostatic floatation condition.

Table 1 is a pseudocode version of the ice model. We present this diagram for clarity, but also in order to aid the description of adjoint generation in the following section. At the beginning of a given time step, $h^{(n)}$, the thickness at time t_n is known. The cells that are floating are determined from the hydrostatic floatation condition:

$$h^{(n)} \leq -\frac{\rho_w}{\rho} R, \tag{1}$$

where ρ_w and ρ are ocean and ice densities, respectively, and R is bed elevation (negative when below zero). This also determines surface elevation $s^{(n)}$, because when the ice is floating a fraction ($\frac{\rho_w}{\rho}$) of total column thickness is below sea level. These operations are represented in the pseudocode by UPDATE_FLOATATION, which also sets the contribution of basal sliding coefficients to zero in floating cells.

Following this call the nonlinear hybrid stress balance is solved for velocity $\mathbf{u}^{(n)}$, using the scheme from Goldberg (2011). This involves first evaluating the discretized form of the glaciological driving stress $\boldsymbol{\tau}_d = \rho g h \nabla s$ (CALC_DRIVING_STRESS), which depends on $s^{(n)}$ and $h^{(n)}$. This is then followed by Picard iteration on viscosity and basal coefficients. In each iteration m of the loop, the matrix \mathbf{A}_m (corresponding to the two-dimensional elliptic PDE mentioned above) is constructed, using current iterates of nonlinear ice viscosity $\nu^{(m)}$ and basal coefficient $\beta_{\text{eff}}^{(m)}$ (BUILD_STRESS_MATRIX).

β_{eff} is a variable unique to the stress balance derived in Goldberg (2011), and thus its form is given here for the benefit of the reader:

$$\beta_{\text{eff}} = \frac{f(u_b)}{u_b \left(1 + \frac{\omega f(u_b)}{h u_b} \right)}. \tag{2}$$

Here u_b is the magnitude of basal sliding velocity ($\mathbf{u}|_{z=b}$), and $f(u_b)$ determines the functional relation between basal

Table 2. Pseudocode version of adjoint model reverse-time-stepping algorithm corresponding to Table 1. (AD) indicates that this step is processed by the algorithmic differentiation tools.

```

DO    $t_n = t_f, \dots, 0$ 
    • CALL  ADADVECT_THICKNESS
                                      $\delta^* \mathbf{u}^{(n)}, \delta^* h^{(n)}$  from  $\delta^* h^{(n+1)}$ 
                                     via adjoint to continuity equation (AD)

    • CALL  ADVEL_SOLVE
        • DO    $m = m_{\text{term}}, \dots, 0$ 
            • CALL  ADUPDATE_VISC_BETA
                                      $\delta^* \mathbf{u}^{(m+1)}, \delta^* \beta$  updated from
                                      $\delta^* \mathbf{v}^{(m+1)}, \delta^* \beta_{\text{eff}}^{(m+1)}$  (AD)

            • CALL  ADSTRESS_CG_SOLVE
                                     Equations (7) and (8) (non-AD)

            • CALL  ADBUILD_STRESS_MATRIX
                                      $\delta^* \mathbf{v}^{(m)}, \delta^* \beta_{\text{eff}}^{(m)}, \delta^* B_{\text{glen}}, \delta^* h^{(n)}$ 
                                     updated from  $\delta^* A_m$  (AD)

        • END   DO
        • CALL  ADCALC_DRIVING_STRESS
                                      $\delta^* h^{(n)} = \rho g \nabla_s^{(n)} (\delta^* \boldsymbol{\tau}_d),$ 
                                      $\delta^* s^{(n)} = \rho g h^{(n)} (\nabla \cdot \delta^* \boldsymbol{\tau}_d)$ 
                                     (AD)

    • CALL  ADUPDATE_FLOATATION
           
$$\begin{cases} \delta^* h^{(n)} = \delta^* h^{(n)} + (1 - \frac{\rho_w}{\rho_w}) \delta^* s^{(n)} & \text{if } h^{(n)} \leq -\frac{\rho_w}{\rho} R \\ (\delta^* h^{(n)}, \delta^* R) = (\delta^* h^{(n)}, \delta^* R) + \delta^* s^{(n)} & \text{otherwise} \end{cases}$$

           (AD)

END   DO

```

stress ($\boldsymbol{\tau}_b$) and sliding velocity:

$$\boldsymbol{\tau}_b = \frac{f(u_b)}{u_b} \mathbf{u}|_{z=b}. \quad (3)$$

u_b is determined from basal stress and depth-averaged velocity under the approximation of depth-independent longitudinal stresses, and ω is a factor that depends on the current iterate of viscosity, as defined in Goldberg (2011) (Eq. 35).

Next the resulting linear system $A_m \mathbf{u}^{(m+1)} = \boldsymbol{\tau}_d$ is solved for the new iterate of \mathbf{u} (STRESS_CG_SOLVE). The nonlinear ice viscosity and sliding coefficients are then updated with this new guess for velocity (UPDATE_VISC_BETA).

Following the velocity solve, thickness is updated via the depth-integrated continuity equation, using a simple second-order flux-limited finite volume scheme. If calving front advance is allowed, this is carried out using an algorithm based on that of Albrecht et al. (2010), though no calving parameterization has been implemented. This all takes place in ADVECT_THICKNESS, and concludes the time step.

In the hybrid balance, the viscosity ν and sliding coefficient β_{eff} (which depends on velocity even for a linear sliding law) have slightly more complicated expressions than in the SSA balance. However, under the Picard iterative scheme employed, the updates of these fields are straightforward. Additionally, UPDATE_VISC_BETA can seamlessly be re-

placed by an SSA viscosity update, effectively making the model a shallow-shelf model.

The matrix A_m referred to above is constructed based on a finite element method with bilinear $Q1$ elements on a rectangular mesh. $\nu^{(m)}$ and $\beta_{\text{eff}}^{(m)}$ are considered constant within an element. Finite elements were chosen, not to use a mesh that conforms to irregular boundaries, but rather because of the ease with which complicated nonlinear boundary conditions can be implemented. A conjugate gradient method is used to solve the linear system, with a simple Jacobi preconditioner.

The model was implemented as an extension (a “package”) within the Massachusetts Institute of Technology general circulation model (MITgcm; Marshall et al., 1997). As such the code takes advantage of various grid definitions and metrics, file I/O subroutines, and the MITgcm parallel domain decomposition and message-passing facilities. The ice flow model run time has been observed to scale favorably with domains of $\sim 40,000$ degrees of freedom and tens of processors. Developing the ice model within the MITgcm framework has other potential advantages: while the model is currently isothermal, there is the potential to rapidly implement temperature transport (and transport of other scalars) using the generic tracer code developed for MITgcm. An ice model that shares grid and modeling components with an ocean model also lends itself to ice–ocean coupling, which is

a future development direction. However, the most important benefit for present purposes is that we are able to take advantage of the adjoint generation framework within the MITgcm, as discussed in the following section.

3 Model adjoint

The notion of “adjoint” is best understood in terms of its construction. Assume that, in a time-dependent ice model, one were concerned with how the initial thickness at a *single location* affected the thickness field at the end of the run. Assume the initial thickness field is represented by a vector $\mathbf{h}^{(0)}$, and the thickness at the final time step t_n by $\mathbf{h}^{(n)}$. If the model is differentiable, one can then consider a perturbation to a single element of $\mathbf{h}^{(0)}$ (i.e., a directional derivative) and propagate the perturbation forward to the final time to the resulting perturbation in the model output. This forward-propagation of perturbations is known as a *tangent linear model* (TLM). The TLM provides information about sensitivities of model output to a single variable (i.e., $h^{(0)}(i)$, the value of $\mathbf{h}^{(0)}$ in cell i). If one wanted information about sensitivities to all such $h_k^{(0)}$, $k = 1, \dots, M$, where M is the size of the grid, the TLM would need to be run M times. (Note that in this context, the “single location” referred to above is not a point in space, but rather a localized area that depends on the model’s discrete representation of a continuous field. If the representation were spectral in nature, a localized value could not be well represented by a single degree of freedom.)

On the other hand, the adjoint can provide this information in a single run, provided the output of interest is scalar-valued. This scalar is often referred to as a *cost*, *objective*, or *target function*, for instance, the total volume of ice in the domain at the end of the run. The adjoint model is referred to as such because it is the mathematical adjoint of the TLM. This seemingly trivial distinction has important implications for how the models are constructed. For the TLM, the forward perturbation is found by successive compositions of the TLMs of sequential time steps. The adjoint model operates backwards in time, by composing the adjoints of individual time steps (or operations within time steps) in reverse order. The eventual result of the TLM is the sensitivity of the output to a single input variable, whereas the result of the adjoint model is the sensitivity of a *single scalar output* to a set of input variables. In the remainder of this section we give an outline of the generation of our ice model adjoint, highlighting the varied approaches employed. For a more comprehensive discussion of the mathematical meaning and implications of adjoint models, see Heimbach and Bugnion (2009).

First we introduce notation that will aid our discussion. Each state variable of the model has a corresponding variable in the adjoint model state (or more formally, a co-vector in the dual to the tangent space of the model at the given point in state space). For a state variable x , we denote its tangent space counterpart by δx , and its adjoint by $\delta^* x$. In gener-

ating the adjoint we relate the adjoint state variables to one another. Assume the state variable y derives from x through the atomic operation $y = g(x)$. To generate the tangent linear model, we find δy , the perturbation to y , by applying the linear operator $\left(\frac{\partial g}{\partial x}\right)$ to δx . Conversely, we track sensitivities of a cost function J from the final time backward. This means, if the sensitivities of J to y are stored in $\delta^* y$, we find $\delta^* x$ by applying the adjoint of $\left(\frac{\partial g}{\partial x}\right)$ to $\delta^* y$.

In Table 3 we give a pseudocode version of the adjoint ice model, corresponding to the version of the forward model presented in Table 1. The time-stepping loop is now from the final time t_f to 0. Note that the intermediate steps of a single time step occur in reverse order, and the adjoint of the Picard iteration loop begins from `m_termin`, the termination step of the forward Picard loop. From this it can be seen how initial condition sensitivities, denoted $\delta^* h^{(0)} \equiv \nabla_{[h^{(0)}]} J$, might be found, as well as parameter sensitivities. For instance, β_{eff} derives in part from the sliding parameter field β in the pseudo-subroutine `UPDATE_VISC_BETA`, and so contributions to $\delta^* \beta$ are calculated in `ADUPDATE_VISC_BETA`. Note that the divergence operator in `ADCALC_DRIVING_STRESS` is actually the discrete divergence, corresponding exactly to the discretized gradient operator `CALC_DRIVING_STRESS`. Note also that the updates of $\delta^* h^{(n)}$ and $\delta^* R$ from $\delta^* s^{(n)}$ in `ADUPDATE_FLOATATION` involve the same conditional statement as in `UPDATE_FLOATATION`. Also in this pseudo-subroutine, the backward propagation of β -sensitivities are terminated where the floatation condition is satisfied.

The majority of the adjoint generation is carried out with AD software. Source-to-source transformation AD tools generate adjoint code by treating each line of source code as an atomic step and finding its adjoint – similar to the transformation implied by Tables 1 and 3 but at a higher level of granularity. Several products are available, including open-source software such as OpenAD (Utke et al., 2008). We use the tool TAF (Transformation of Algorithms in Fortran; Giering et al. (2005)), which has long been used for adjoint generation with the MITgcm (Heimbach et al., 2005), and which was used in Heimbach and Bugnion (2009). This highlights an additional advantage of developing the ice model with MITgcm: the readiness with which the AD tools can be applied. The code needs to satisfy certain requirements in order to be easily parsed by TAF (e.g., there should not be do-loops of indeterminate length, and the use of pointers should be minimal), and conforming to the standards of MITgcm helped to ensure this. Additionally, MITgcm defines rules for steps involved with parallel message-passing, such that they can be treated as adjointable operations, and thus the adjoint model is parallel as well. The same is true for our ice model adjoint.

In Table 3, the steps to which AD tools are applied are indicated by (AD). These are straightforward operations and we had minimal difficulties applying the tools. Only one step, the

matrix solve, is not handled by AD. The conjugate gradient algorithm employed involves a large number of intermediate variables relevant only to the solver, and can require many iterations. A direct application of AD tools to the solver would involve large memory requirements, as well as a great deal of either code modification or manual “directing” of the tools, to prevent costly recomputation loops. On the other hand, TAF (as well as other AD tools) offers a facility to replace the adjoint code that is automatically generated by manually written code at the subroutine level, if the adjoint of a subroutine is known, as is the case with the linear solver. As in Table 1, the linear solve can be written

$$\mathbf{u}^{(m+1)} = \mathbf{A}_m^{-1} \boldsymbol{\tau}_d, \quad (4)$$

where $\mathbf{u}^{(m+1)}$ is the iterate of velocity found at step m of the Picard loop, \mathbf{A}_m is the matrix constructed with the previous iterates of viscosity and basal coefficient, and $\boldsymbol{\tau}_d$ is the driving stress $\rho gh \nabla s$ at the current time step. The solve can be viewed as an operator g with arguments \mathbf{A}_m and $\boldsymbol{\tau}_d$, i.e.,

$$g : (\mathbb{R}^{n \times n} \times \mathbb{R}^n) \rightarrow \mathbb{R}^n; \quad (5)$$

thus the adjoint operator must have the form

$$g^* : \mathbb{R}^n \rightarrow (\mathbb{R}^{n \times n} \times \mathbb{R}^n). \quad (6)$$

The adjoint of Eq. (4) is given by

$$\delta^* \boldsymbol{\tau}_d = \delta^* \boldsymbol{\tau}_d + \mathbf{A}_m^{-T} \delta^* \mathbf{u}^{(m+1)}, \quad (7)$$

$$\delta^* \mathbf{A}_m = -\delta^* \boldsymbol{\tau}_d (\mathbf{u}^{(m+1)})^T. \quad (8)$$

Equation (7) is written as an accumulation of adjoint sensitivities of $\boldsymbol{\tau}_d$; the adjoint of each Picard iteration has an effect on $\delta^* \boldsymbol{\tau}_d$. In our adjoint model, a subroutine carries out these operations; since \mathbf{A}_m is self-adjoint, the same conjugate gradient solver (with the same matrix coefficients) is used in Eq. (7). Note that in the MITgcm ocean model, which solves a linear system for either rigid-lid surface pressure or free surface elevation, an equation similar to Eq. (7) is solved by the adjoint model, and the symmetry of the matrix is similarly exploited. However, the coefficients of the matrix are fixed, and so Eq. (8) has no counterpart. Following the solution of Eq. (7) and Eq. (8) by hand-written code, evaluation of the adjoint via AD-generated code is resumed: $\delta^* \boldsymbol{\tau}_d$ and $\delta^* \mathbf{A}_i$ are passed to the adjoint of the matrix construction, ADBUILD_STRESS_MATRIX.

An important by-product of this approach (“hiding” the matrix inversion from the AD software) is that it allows us to potentially replace our linear solver with faster, optimized “black-box” solvers (such as those available in external packages such as PETSc) without affecting the accuracy of the adjoint. We point out that the advantages mentioned here are not limited to our model, i.e., to one that implements a shallow-shelf or hybrid stress balance. Matrix inversion is

the most time-intensive component of any ice model with a nonlocal stress balance, and the part of the code that is most likely to lead to difficulties in application of AD software. As long as it can be handled in a similar way to our model (i.e., as long as the matrix is self-adjoint), efforts can be focused on ensuring that the remainder of the code is suitable for algorithmic differentiation.

It was found that, in order that the adjoint model produce accurate results, the CG tolerance for the linear solve in the adjoint needed to be several orders of magnitude smaller than that used in the linear solve of the forward model. (The accuracy is assessed by comparing the derivatives calculated by the adjoint to finite-difference approximations. Relative agreement to within 10^{-6} was considered accurate.) This suggests that without special treatment of the convergence criteria, a fully AD-generated adjoint might have low accuracy, and further supports our decision to let the AD software “bypass” our linear solver.

4 Nonlinear optimization

For optimization problems, our model uses the M1QN3 library, publicly available Fortran code which is based on the algorithm described in Gilbert and Lemaréchal (1989). The M1QN3 algorithm solves large-scale unconstrained minimization problems. It provides a search direction and step size based on a limited-memory quasi-Newton approximation to the Hessian of the objective function J , using gradients of J that are provided by the user throughout the minimization process. The gradient of J is calculated by the adjoint. The M1QN3 software has been adapted for use with the control package of MITgcm, and so we are able to make use of it as well.

5 Experiments

We present a suite of experiments that showcases the adjoint and optimization capabilities of our model. The optimizations consist mainly of “identical-twin” experiments, in which the field being inverted for is known a priori, and the “observed” data is a perfect solution of the model. Such inverse problems have been termed “inverse crimes” (Colton, 1998), because they are considered too optimistic to provide a reliable test of performance. However, an inverse model must at a minimum be able to perform well on these idealized experiments, which demonstrate the strengths and limitations of the model.

Table 3. Pseudocode version of adjoint model reverse-time-stepping algorithm corresponding to Table 1. (AD) indicates that this step is processed by the algorithmic differentiation tools.

```

DO   tn = tf, ..., 0
    • CALL  ADADVECT_THICKNESS
                                     δ*u(n), δ*h(n) from δ*h(n+1)
                                     via adjoint to continuity equation (AD)

    • CALL  ADVEL_SOLVE
        • DO   m = mterm, ..., 0
            • CALL  ADUPDATE_VISC_BETA
                                     δ*u(m+1), δ*β updated from
                                     δ*v(m+1), δ*βeff(m+1) (AD)

            • CALL  ADSTRESS_CG_SOLVE
                                     Equations (7) and (8) (non-AD)

            • CALL  ADBUILD_STRESS_MATRIX
                                     δ*v(m), δ*βeff(m), δ*Bglen, δ*h(n)
                                     updated from δ*Am (AD)

        • END   DO
        • CALL  ADCALC_DRIVING_STRESS
                                     δ*h(n) = ρg∇s(n)(δ*τd),
                                     δ*s(n) = ρgh(n)(∇ · δ*τd)
                                     (AD)

    • CALL  ADUPDATE_FLOATATION
        { δ*h(n) = δ*h(n) + (1 -  $\frac{\rho_w}{\rho_w}$ )δ*s(n)   if h(n) ≤ - $\frac{\rho_w}{\rho}R$ 
        { (δ*h(n), δ*R) = (δ*h(n), δ*R) + δ*s(n)   otherwise
                                     (AD)

END   DO
    
```

5.1 Experiment 1: sensitivity to ice shelf stiffness and melt rates

The first experiment does not involve optimization, but simply demonstrates the interpretive powers of the adjoint model. We consider the sensitivity of the volume of grounded ice in a marine ice stream to thermodynamic effects on its adjacent ice shelf. Such effects are of considerable importance, given observations of the Antarctic coastline made over the last several decades. Confined ice shelves are known to act as logjams to ice stream flow (a phenomenon referred to as ice shelf buttressing; Dupont and Alley (2005)) and therefore exert a large control on grounded ice mass balance. The heat contained in Southern Ocean subsurface waters is able to cause melting at the underside of Antarctic ice shelves, most notably those in the Amundsen Sea embayment (Jacobs et al., 1996; Jacobs, 2006; Jenkins et al., 2010; Jacobs et al., 2011). Meanwhile, widespread speedup, thinning, and mass loss have been observed in these ice shelves and the ice streams that feed them (Rignot, 1998; Rignot et al., 2002; Shepherd et al., 2002, 2004).

A number of modeling studies have been carried out exploring the magnitude and distribution of sub-ice shelf melting that results from intrusions of warm water into an ice shelf cavity, as well as how these quantities depend on the geometry of the cavity and the strength of the forc-

ing (e.g., MacAyeal, 1984; Jenkins, 1991; Hellmer et al., 1998; Holland et al., 2008; Little et al., 2009; Heimbach and Losch, 2012). Less frequently asked is how the response of grounded ice depends on the magnitude and spatial distribution of melting, though Walker et al. (2008) and Gagliardini et al. (2013) have investigated this question. To understand how large-scale changes in ocean heat content and circulation can affect ice sheets, both questions are important.

Here we address the second question with an idealized ice sheet–stream–shelf system. Ice is allowed to flow in a rectangular domain of 150 km × 150 km, where ice flux input is constant along the $x = 0$ boundary (simulating flow from the ice sheet interior). Along the $x = L = 150$ km boundary a calving front boundary condition is imposed, whether ice is grounded or floating (Weertman (1957); Schoof (2007), Appendix B. A channel runs the length of the domain, deepening away from the front. The bedrock elevation is expressed

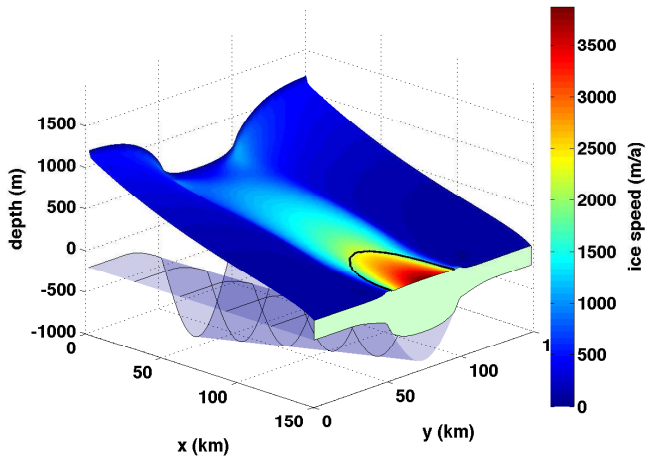


Fig. 1. Background steady state of the ice stream–shelf system in Experiment 1. Coloring on the upper surface is velocity magnitude. The thick black contour denotes the grounding line.

(in meters) by

$$R(x, y) = R_x(x)R_y(y) - 200, \quad (9)$$

$$R_x(x) = 1 + \frac{5}{6} \left(\frac{150 - x}{150} \right), \quad (10)$$

$$R_y(y) = \begin{cases} -100 - 600 \sin\left(\frac{\pi(y-50)}{50}\right) & 50 \leq y \leq 100 \text{ km}, \\ -100 - 100 \sin\left(\frac{\pi(y-50)}{50}\right) & (25 \leq y \leq 50) \text{ or} \\ (100 \leq y \leq 125 \text{ km}), & \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

where x and y are in km. Sliding is governed by a linear sliding law, i.e.,

$$\boldsymbol{\tau}_b = \beta^2 \mathbf{u}_b. \quad (12)$$

Within the channel, i.e., where $50 \text{ km} \leq y \leq 100 \text{ km}$, β^2 is set to $30 \text{ Pa}(\text{m/a})^{-1}$. Outside of the channel it is 9 times larger. The lateral boundaries at $y = 0 \text{ km}$, 150 km are no-slip boundaries, but the resistance to ice stream flow comes from basal stress in the outer “sheet” region, not the sidewalls. The Glen’s law parameter A is set uniformly to $9.5 \times 10^{-18} \text{ Pa}^{-3} \text{ a}^{-1}$, which corresponds roughly to a temperature of -15°C . The model is run to equilibrium, shown in Fig. 1. An ice shelf about 50 km long forms over the channel.

To examine changes in grounded ice, we consider volume above floatation (VAF), defined as volume of ice that would contribute to sea level rise if all of the ice in the domain were to melt, and the loss thereof (Dupont and Alley, 2005). Note that the floating columns of ice do not contribute to VAF. We calculate the adjoint sensitivities of VAF loss during the ten-year run to two different input fields: basal melting under the ice shelf (\dot{m}), and the Glen’s law flow parameter A . A realistically depends on the temperature and fabric of the ice,

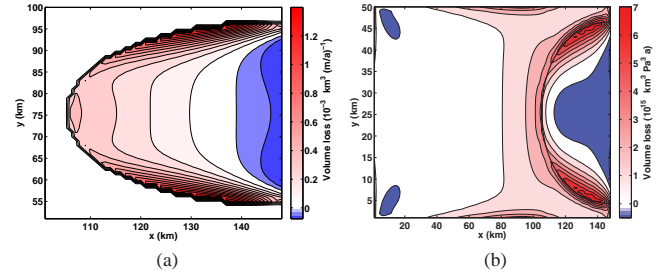


Fig. 2. (a) Sensitivity of grounded ice volume after 10 yr to sub-shelf melt rates, in 10^{-3} km^3 per (m a^{-1}) of melting. The non-colored section of the figure is within the grounded part of the domain. (b) Similarly for Glen’s law parameter A , but with units of $10^{-15} \text{ km}^3 \text{ Pa}^{-3} \text{ a}^{-1}$. A thick black contour denotes the grounding line. The largest values in both figures are in the ice shelf margins. Note the differing x and y bounds on the figures.

but here we consider dependence on A directly. We define a scalar function

$$J = \sum_i \text{HAF}(i) \Delta x \Delta y, \quad (13)$$

$$\text{HAF}(i) = \left(h(i) + \frac{\rho_w}{\rho} R(i) \right)^+, \quad (14)$$

where “HAF” is height above floatation, $R(i)$ and $h(i)$ are bed elevation and thickness at final time in cell i , and Δx and Δy are spacings on the (here uniform) grid. J is the cost function, or objective function, for this experiment (equal to final-time VAF), and it is the scalar function of which the gradient is found, with respect to \dot{m} and A , by the adjoint model.

Melt rate sensitivities are shown in Fig. 2a. The value at a location (i.e., in cell i) can be interpreted as the loss in VAF after 10 yr with a constant melt rate of 1 m a^{-1} in cell i . Sensitivities are only nonzero in locations where ice is floating. This is due to a rule in the model that melt rates cannot be applied under grounded ice: so even though adjoint sensitivities are propagated backward in time, they terminate at the point in the code where melt rates are applied. The pattern of sensitivities is interesting: they are largest not in the deepest part of the shelf near the grounding line, where melt rates are generally highest, but rather in the margins of the shelf. This implication (which should be taken with many qualifications, as explained below) that shifts in melt rates near relatively shallow ice shelf margins could have stronger impacts on grounded ice than shifts of equal magnitude near the grounding line. Also curious is the fact that sensitivities are actually negative (though very small) near the ice shelf front. This effect is in fact realized in forward runs: thinning of the ice shelf front leads to flux across the ice shelf margin (into the shelf) and drawdown of the grounded ice cliffs, lessening their contribution to VAF loss.

Note that these results differs from those of Walker et al. (2008) and Gagliardini et al. (2013), who found that melt rate

patterns which concentrate toward the grounding line have the greatest impact on grounded ice. However, these studies used flowline models and thus could not resolve the effects of thinning near shear margins. Therefore, this experiment highlights the need to resolve both horizontal directions when assessing the impact of melting on ice shelf buttressing.

Figure 2b shows sensitivities with respect to A . (Values are large because the nominal value of A is on the order of $10^{-17} \text{ Pa}^{-3} \text{ a}^{-1}$.) Here values are nonzero in grounded and floating ice, and are largest in the margins of the ice shelf (and negative, but small toward the ice shelf front). A is sometimes referred to as the “fluidity” of ice; i.e., the larger it is the more easily the ice flows. A positive change corresponds to weakening of the ice, and weakening in the margins leads to the most grounded ice loss.

That the thinning (through melting) and weakening of an ice shelf can lead to grounded ice loss is well established on theoretical (Thomas, 1979) and modeling (Dupont and Alley, 2005; Goldberg et al., 2009; Little et al., 2012) bases. But less-oft discussed is which *parts* of the shelf are most sensitive to this mechanism, that is, the structural integrity of the ice shelf. The result of our idealized case, with a small, confined ice shelf, suggests that the margins are the weak points of the shelf. It is not clear to what extent this applies to ice shelves in general, although intuition suggests that margins play a similar role in all confined, relatively small shelves. This demonstration shows that a similar analysis can be done on any ice shelf–ice stream system (although a baseline solution free of artificial model drift would be required for meaningful results).

The computational advantage of the adjoint in producing data sets such as those shown in Fig. 2 is considerable. Using a domain decomposition over 9 processors (with a 50×50 grid belonging to each process) took approximately 12 minutes, or 5 seconds per time step. The adjoint run that produced both data sets in Fig. 2 (and could have produced additional adjoint sensitivity fields as well) took approximately 4 times longer, giving a total run time of about an hour. (The additional run time is because parts of the forward model must be run multiple times to provide state information for the reverse-time adjoint run.) If, on the other hand, one were to approximate sensitivities by perturbing single parameter values and using one-sided finite differencing, the melt rate sensitivities would take approximately $50 \times 50 \times 0.2 \text{ h} \approx 25$ days (since only a portion of the domain is ice shelf) and the A -sensitivities about 9 times as long.

While the efficiency of the adjoint in finding sensitivities is obvious, it should be kept in mind that the analysis is inherently linear: a forward trajectory is needed around which to perturb. In this case the trajectory was a quiescent one, as the run began in steady state with no melting. As discussed above, melting increases flux across the grounding line through loss of ice shelf buttressing; sufficiently high melt rates would lead to grounding line retreat. This, in turn, would result in increased grounding line thickness (due to the

shape of the bedrock), leading to further mass flux increase (Schoof, 2007). The latter effect is nonlinear, however, and is not detected by our adjoint results. When calculating adjoint sensitivities, the results should not be taken at face value, but rather serve as a starting point for further investigation. It is worth noting that Goldberg et al. (2012b), using a coupled ice–ocean model that allowed grounding line migration, found that thinning of an ice shelf at the margin due to melting was a key factor in unstable retreat of grounded ice – giving credence to the high sensitivities of VAF loss to values in the ice shelf margin.

5.2 Experiment 2: inversion of basal sliding coefficients from surface velocities

In our first inversion experiment we infer basal sliding coefficients from surface velocities, considering only the momentum balance of the model (i.e., no time dependence). This is an identical-twin experiment, in that the surface velocities come from model output with prescribed parameter values (or “true” values), and the inverted parameters are then compared with the truth. The forward model is one-dimensional (only one horizontal direction is considered, and the SSA balance is implemented) with periodic boundary conditions, and surface slope and thickness are constant.

Many authors have carried out similar inversions over the past two decades, both with synthetic observations and real ones (e.g., MacAyeal, 1992, 1993; Vieli and Payne, 2003; Khazendar et al., 2007; Sergienko et al., 2008; Morlighem et al., 2010; Joughin et al., 2009). The purpose of this experiment is not to introduce a new type of glaciological inversion, or to improve upon an existing one, but rather to enable direct comparison with existing methods. As one of the simplest glaciological inverse problems, our adjoint optimization framework must, at a minimum, perform as well as other inversion methods.

The experiment is based on Experiment B from the ISMIP-HOM intercomparison (Pattyn et al., 2008) with $L = 40$ km, and 1 km resolution. The intercomparison specifies a constant thickness of 1000 m, a constant surface slope of 0.1° , and a linear sliding law with a mode-one sinusoidally varying sliding coefficient, or lubrication, β^2 (Fig. 3a). This profile of β^2 represents our true parameter values. The Glen’s law parameter A is uniformly set to $10^{-16} \text{ Pa}^{-3} \text{ m a}^{-1}$. We define as \mathbf{u}_1^* the x velocities from the model with these parameters.

For the inverse problem to determine β^2 , we define a cost function on the model output:

$$J = \frac{1}{2} \sum_i^N \frac{(u_1^*(i) - u(i))^2}{\sigma_i^2}, \quad (15)$$

where the summation is over all cells i , and $u_1^*(i)$ and $u(i)$ are the nodal values of the observed velocities \mathbf{u}_1^* and model output velocities \mathbf{u} , respectively, where the latter depend on

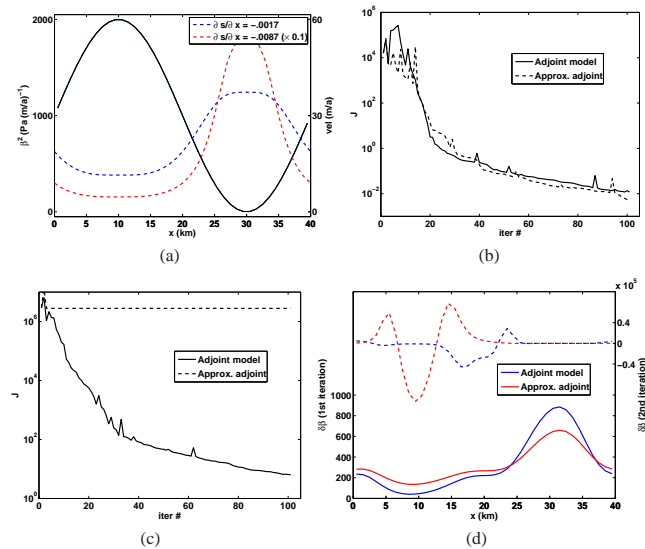


Fig. 3. (a) Data for inverse problem. “True” basal sliding coefficient β^2_{true} (left axis) and corresponding velocity profiles (right axis) for low and high surface slopes. (b) Cost function J versus number of model evaluations for full and approximate adjoint, low surface slope. (c) Same for high surface slope. (d) Gradient found by full and approximate adjoint in first iteration (solid) and second iteration (dashed), high surface slope.

the current guess for β through the stress balance. (In this inversion, we attempt to recover β , not β^2 , which is the easiest way to impose the constraint that sliding coefficients are nonnegative.) The numbers σ_i are scaling factors for the cost function. Generally these scaling factors represent a priori knowledge or first guesses regarding observations or parameters; for instance, σ_i might be the uncertainty in the velocity observation. In practice, this prevents poorly constrained observations from leading to overspecification. In this example, the scaling factors are set uniformly to 1: this presents no loss of generality, as long as values of J are compared to the initial value.

For a given β , the adjoint finds sensitivities of J with respect to β . The cost function is then minimized using the optimization algorithm described in Sect. 4. Notice that no regularization terms have been added to the cost function to ensure a priori properties of the lubrication field, e.g., smoothness and boundedness, although we include such terms in later experiments.

Figure 3b shows how J evolves, eventually reaching cost reduction, defined as $\left(\frac{J}{J_0}\right)$, on the order of 10^{-6} . J_0 is the value of J using the initial guess for β^2 (described below). For comparison, an inversion is also carried out where the “approximate” adjoint of MacAyeal (1993) is used, rather than the adjoint sensitivities from the AD-generated adjoint (the “full” adjoint). The approach is termed approximate because the dependence of viscosity on strain rates is ignored.

In both cases, the same optimization algorithm is used, so calculated adjoint sensitivities are the only difference between the two inversions.

The inverted β^2 is not shown, but it is very close to the true profile. The same is true of the experiment with higher surface slope, described below. This is because of the lack of high-frequency variability in the initial guess. If the initial guess had a broad spectrum with power at high frequencies, such as a Gaussian shape (not shown), the inversion would still decrease J as in Fig. 3b, and the inverted β^2 would have the same broad pattern as the true profile, but there would be high-frequency variability as well. This is because the velocities in the model are insensitive to high-frequency variability in β^2 ; they are muted by longitudinal stresses.

Both optimizations begin with the same initial guess for β^2 , a half-mode sinusoid of the same amplitude. A uniform β would be the simplest guess, but with a constant thickness and surface slope there is zero strain rate, and this leads to very high values for viscosity. The performance of the inverse model then depends on the viscosity regularization parameter. With the hybrid balance, this is not an issue; but in this experiment the SSA balance is used so that different inversion approaches can be compared.

The full and approximate adjoints perform equally well in the experiment with a surface slope of 0.1° , with the full adjoint actually leading to greater values of J early on. However, when surface slope is increased to 0.5° (as in the ISMIP-HOM experiments with flow over a wavy bed), their performances differ. Now the target surface velocities $u_5^*(i)$ are an order of magnitude larger (Fig. 3a), and it is important to maintain the nonlinear dependence of viscosity on strain rates in the adjoint, as evidenced by the poor performance of the approximate adjoint (Fig. 3c). This can also be seen by examining the adjoint sensitivities from the two models. After the first iteration of the inversion (Fig. 3d) the sensitivity profiles are similar, yielding similar search directions in the optimization algorithm. On the next iteration, however, the profiles look very different, and upon being provided with the approximate adjoint sensitivities, the optimization algorithm is unable to lower the cost function.

This is not to say that the AD-generated adjoint model is in all cases an improvement over the approximate adjoint in this type of inversion. The optimization algorithm has a number of associated parameters; it is possible that, for the experiment considered, different parameters might yield better performance with the approximate adjoint, or worse with the full adjoint. It is clear, nevertheless, that instances exist where the full adjoint is advantageous.

5.3 Experiment 3: estimation of past conditions

We test the ability of our inversion framework to recover two different parameter fields simultaneously based on time-dependent data. Previous studies have considered time-dependent observations, for example Jay-Allemand et al.

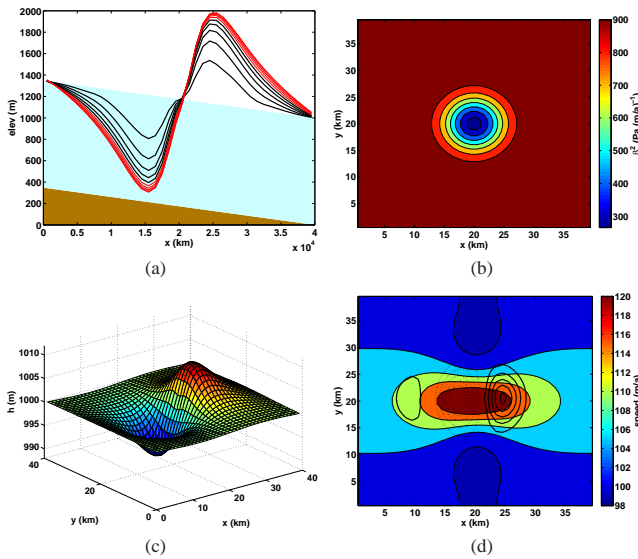


Fig. 4. (a) “True” profile of glacier at time $t = 0$ and transient surface profiles spaced one year, exaggerated by a factor of 100. The last 5 profiles (in red) correspond to data used to constrain β^2 and $h(t = 0)$. (b) “True” β^2 . (c) Thickness at final time ($t = 10$ a). (d) Surface speed at final time. Region where flow is strongly divergent (convergent) is indicated by solid (dotted) black contours.

(2011). However, the assimilation in this study consisted of a series of “snapshot” inversions of surface velocities; no dynamic consistency between the states at different times was enforced. We re-emphasize that the propagation of sensitivity/misfit information back in time by the adjoint model implies that the inversion takes advantage of available observations not only forward in time (as do filter or sequential assimilation methods) but also backward in time (smoother or variational methods), mediated through the model dynamics.

We consider a mountain glacier undergoing adjustment in response to a perturbation in basal conditions. We assume we have knowledge of surface velocities at the end of this adjustment (the “present”) and of thickness and surface elevations at certain discrete times during the adjustment, but not of the initial thickness, nor of the basal conditions during this adjustment. Perfect knowledge of bed elevation is also assumed (this is not, in general, the case, and in the next experiment we deal with bed topography uncertainty). We ask whether we can recover this initial thickness and basal traction with adjoint-based optimization. Reconstructions of past glacier configuration, coupled with inversions of basal properties, could be very useful in certain glaciological settings. For instance, if an ice stream is known to be out of balance due to relatively recent changes in its basal environment, such an inversion could give us thickness of the stream prior to its observational history or, conversely, could pinpoint the time of onset of the changes.

The forward model is again a periodic domain with a constant bed slope of -0.5° in the x direction. Both horizon-

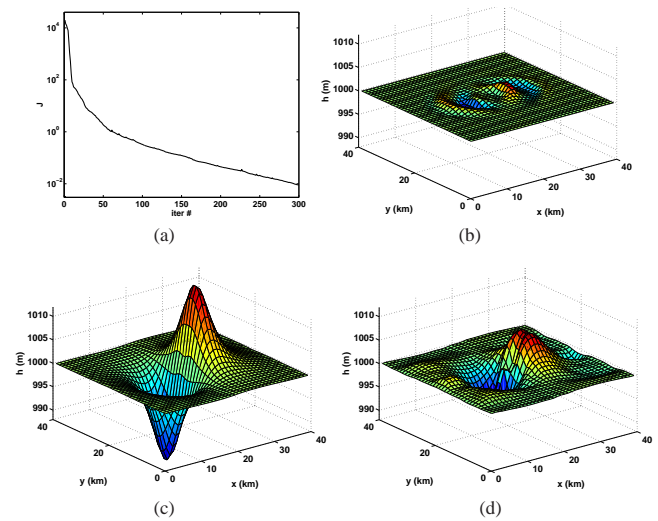


Fig. 5. (a) Value of cost function versus iteration. (b) Final inverted initial thickness ($h(t = 0)$). “True” field is uniformly 1000 m. Compare with Fig. 4c, which is used as the initial guess. Maximum misfit is reduced from 9 to 1.8 m; rms misfit is reduced from 2.6 to 0.25 m. (c) Inverted initial thickness when velocity is “frozen” in its observed state (Fig. 4d). (d) Inverted initial thickness when only the observed thickness at $t = 10$ yr is used to constrain the problem.

tal dimensions are resolved, and the domain is y periodic as well, with zero bed slope in the y direction. The domain is 40×40 km with 1 km resolution. The initial thickness is uniform with a value of $H_0 = 1000$ m. The Glen’s law parameter A is as in the previous experiment. A hybrid stress balance is used. The sliding law is again linear, with sliding coefficient β^2 . The true β^2 is defined as

$$\beta^2 = 1000 - 750 \exp\left(-\left(\frac{r}{5}\right)^2\right), \quad (16)$$

with units of Pa m a^{-1} , where r is distance from the center of the domain in km. The model is integrated forward with 30-day time steps for 10 yr, at the end of which the model is close to a new equilibrium. Figure 4a shows the initial surface and thickness, as well as annual surface elevations during the adjustment (magnified a hundred-fold), along the center line $y = 20$ km.

For data in our identical-twin inversion to recover β^2 and the initial thickness, we take annual surface elevation over the last 5 yr (the red curves in Fig. 4a) and surface velocity averaged over the last year (Fig. 4d). These are referred to below as the “target data”. We define a cost function

$$J = \sum_i^N \left(\frac{1}{N\sigma_u^2} |\mathbf{u}_s^*(i) - \mathbf{u}_s(i)|^2 + \frac{1}{N\sigma_s^2} \sum_{n=6}^{10} (s^{(n),*}(i) - s^{(n)}(i))^2 \right), \quad (17)$$

where the outer summation is over all cells i . $\mathbf{u}_s^*(i)$ is the target surface velocity in cell i , and $s^{(n),*}(i)$ is the target surface elevation in cell i in year n . σ_u and σ_s are meant to signify uncertainties in velocity and surface measurements, but

since the observations are synthetic there is no rationale to use spatially varying uncertainties. The relative values of σ_u and σ_s affect the results of the inversion, however. We use 1 m/a for σ_u and 2 cm for σ_s . Thus when $J \sim O(1)$ then the model misfit is on the order of observational uncertainties on average. This value is motivated by the spread of surface elevations over the 5 yr sampling period; field measurements of surface elevation generally have $O(\sim 1 \text{ m})$ errors, at least in relatively flat regions (Griggs and Bamber, 2008). GPS measurements are capable of much higher accuracy than remote sensing, provided the presence of a nearby exposed promontory, or ice for which the surface elevation is changing on much slower timescales (N. Gourmelen, personal communication, 2013), but we concede that GPS observations are unlikely to be collected with such high spatial resolution.

For initial guesses, a uniform value of $\beta_{\text{guess}}^2 \equiv 400 \text{ Pa m a}^{-1}$ is used for β^2 ; as for $h^{(0)}$, the observed thickness at $t = 10 \text{ yr}$ is used. That is,

$$h_{\text{guess}}^{(0)} = R + s^{(10),*} \quad (18)$$

where R is the bedrock, the rationale being that, in this artificial experiment, the glacier is in this configuration (or close to it) during the period of observation.

Figure 5a shows the cost function trajectory. About 60 to 80 gradient evaluations are required for J to be $O(1)$, but the inversion is carried farther than that. Figure 5b shows the final estimate of $h^{(0)}$. Remnants of the initial guess can be seen, but the associated misfit (when compared with the true $h^{(0)}$) has a maximum amplitude of about 1.8 m, whereas that of the initial guess has a maximum of 9 m. In terms of root mean square error,

$$\text{err}_{\text{rms}} = \left(\overline{(h^{(0)} - 1000)^2} \right)^{1/2}, \quad (19)$$

this value is reduced from 2.6 m for the initial guess to 0.25 m. The estimated β^2 is not shown; it differs from the true β^2 by at most 2%.

To appreciate the complexity of this inverse problem, one should contrast with that of reconstructing the history of an advected field in a flow that is fixed, or independent of the field. Such problems have been dealt with frequently in generic flow fields. Our problem is more complex, in that the velocities depend nonlinearly and nonlocally on the advected quantity (thickness). To illustrate this, another experiment is carried out, in which depth-averaged velocity is fixed to the target surface velocity, independent of time. The cost function consists only of the second term in Eq. (17), and the only control field is initial thickness, with initial guess the same as before. The forward model is essentially just the mass balance equation. The estimated $h^{(0)}$ is shown in Fig. 5c. In this experiment the cost function only decreases by a factor of 20 (as opposed to the $O(10^6)$ decrease shown in Fig. 5a). More interesting is that the final estimate of $h^{(0)}$ is actually worse than the initial guess. This is because the divergence pattern

in the target velocity field (Fig. 4d) differs from that throughout much of the “truth” simulation.

It is also interesting to consider how the time dependence of the constraints affects the inversion. The forward model is close to steady state by the end of the 10 yr integration, and the observed surface fields $s^{(n),*}$ differ on the order of centimeters. Since they are so close, one might guess that constraining the surface at years 6 through 9 of the simulation adds little information beyond that contained in the surface elevation at year 10. However, this is not the case. We carry out another inversion in which the cost function given by Eq. (17) is modified so that the summation over n only contains one term, $n = 10$. That is, only the final surface elevation is constrained. In this experiment, the stress and mass balances are again used, rather than solely the mass balance. Figure 5d shows the result of this inversion. In this case the estimated initial thickness is much closer to the initial guess than the true $h^{(0)}$. Valuable information about the thickness trajectory is contained in the intermediate surface observations, even though the temporal variability is small. This particular result hinges on a level of measurement accuracy that is not generally attained; still, it demonstrates the importance of time-dependent information in estimating past ice sheet behavior and other unknown parameters.

To our knowledge, a similar inversion experiment has not been carried out with a higher-order model; although numerical inversions for historic data, including prior ice thicknesses, have been carried out using a model that makes the shallow ice approximation (e.g. Waddington et al., 2007). For this reason, Experiment 3 is intended as a preliminary investigation in inferring unknown parameters based on time-dependent data, in which different types of data (surface elevation and velocities) are not co-temporal. There are many other questions that can be asked of such inversions: for instance, how is the inversion affected if data is not spatially complete, and if the “holes” in data at different times are not coincident? Does the quality of the estimation decay with the hindcast horizon? The choice of duration of our experiment was based on the adjustment time of the forward model, but it is unclear at which point the signal of observations gets lost in the noise of the model. Also, we did not allow for time-dependent lubrication, as in Jay-Allemand et al. (2011); how does this affect the ease of solution?

Further, the assumed accuracy of surface elevation measurement (centimeter scale) is far below that expected by satellite or airborne sensing (meter scale). Nevertheless, the exploratory nature of our experiment is warranted and provides considerable substance, with more realistic accuracy and spatial resolution of measurements being an avenue of further investigation.

5.4 Experiment 4: simultaneous inversion of basal topography and sliding coefficients

Two prevalent challenges in modeling dynamic behavior of real glaciers and ice sheets are those of basal topography and of model initialization. Basal topography is often collected by sparse flight lines of airborne ice-penetrating radar, leading to very low-resolution representations of ice thickness much lower than that required by models used to study highly dynamic features such as ice streams. Recent published errors of gridded Antarctic ice thickness uncertainty are tens of meters or more (e.g., Fretwell et al., 2013). Given the sensitivity of models to representations of bed topography (Durand et al., 2011; Seroussi et al., 2011), this introduces large uncertainties into model response.

Model initialization is important for studies that aim to assess the time-dependent response of ice sheets and glaciers in their current configurations to external forcing. In these studies it is important to start from a quasi-balanced state, in which there is no unrealistic model adjustment or drift that will contaminate the results. Long-timescale spinups can be computationally expensive, and due to parameter uncertainty there is no guarantee that the steady state produced is close to the observed state.

A number of studies have made use of the adjoint method introduced by MacAyeal (1993) to estimate these parameters. However, these inversions are problematic in that the various data sets used are gathered using different methods and resolutions at different times. Very often the solutions can lead to large anomalous mass balances when used to initialize a time-dependent model. Contributing greatly to the error inherent in the inverted solutions are the uncertainties in basal topography (Morlighem et al., 2011).

Two related issues emerge: uncertainties in ice thickness and basal topography, and their impact on the usefulness of assimilated products in model initialization. To deal with the latter, some studies run spinups (albeit shorter ones due to improved parameter guesses) or add “flux corrections” in the term of surface or basal mass balances (e.g., Joughin et al., 2009; Larour et al., 2012a). Such approaches are not ideal, because the model response may be influenced by the adjusted configuration or artificial mass balance. Morlighem et al. (2011) carries out an adjoint-based method that is constrained by the continuity equation with observed velocities, thus guaranteeing a stable mass balance. Basal lubrication and ice stiffness can then be estimated using other methods. However, the method does not consider the stress and mass balance equations together, and thus it does not truly provide a balanced state. In particular, the resulting model velocities may agree well with observed velocities in the L_2 -norm, but this does not guarantee that the divergence patterns will be identical.

With a time-dependent ice model adjoint it is possible to constrain both the continuity equation and the momentum equations in simultaneous inversions for basal lubrication

and topography. In this section we explore the potential of using this approach both to minimize drift in model initializations and to provide improved estimates of basal topography. We note that methods for such inversions have been previously developed (Thorsteinsson et al., 2003; Gudmundsson and Raymond, 2008). However, these methods assume Newtonian rheology and rely on linear transfer functions of small perturbations, so it is not clear that their results carry over to large deviations and nonlinear rheologies.

We also point out that, in the general case, the retrieval of basal lubrication and topography is ill-posed. Consider a sliding glacier of infinite extent, with constant thickness, surface slope, and basal lubrication. With negligible horizontal velocity divergence, the glacier would be in steady state. There are an infinity of lubrication/bed elevation parameter pairs that would give the same surface velocity with the given surface elevation. This is an extreme case, but we will keep this potential limitation in mind. On a related note, the degree to which existing studies of inversions for basal lubrication have compensated for errors in bed topography has remained largely unquantified.

5.4.1 Uncertainties in bed elevation

We perform an identical-twin experiment, in which the surface elevation and velocity are known for an idealized glacier in steady state. This steady state is found by time-integrating the model in a doubly periodic, 40×40 km domain. A linear sliding law is used, with sliding coefficient β^2 . Bed topography and lubrication are given very simple analytical prescriptions, with single-wavelength variation in both x and y directions:

$$R(x, y) = R_0(x) + 200 \sin\left(\frac{2\pi y}{40}\right) \sin\left(\frac{2\pi x}{40}\right), \quad (20)$$

$$\beta^2(x, y) = 1000 - 800 \cos\left(\frac{2\pi y}{40}\right) \cos\left(\frac{2\pi x}{40}\right), \quad (21)$$

given in units of m and Pa (m/a)⁻¹, respectively, where x and y are in km and $R_0(x)$ gives a constant downward slope of 0.5° in the x direction. The Glen’s law coefficient is as in the previous two experiments. Horizontal resolution is 1 km. Note that R and β^2 are uncorrelated. This is not motivated by any realistic model relating basal lubrication to basal elevation; the physics of basal sliding are far too complicated to be addressed in this simple experiment. Rather, it is a simple experimental setup intended to serve as proof of a concept. To achieve steady state, the surface is allowed to adjust, and the state is presumed steady when the surface elevation rate of change is $O(10^{-4} \text{ m a}^{-1})$. Figure 6 shows R and β^2 , along with the steady-state surface elevation and speed. A hybrid stress balance is used, so the surface speeds do not reflect their depth-averaged values; vertical shear accounts for up to $\sim 30\%$ of the surface velocity.

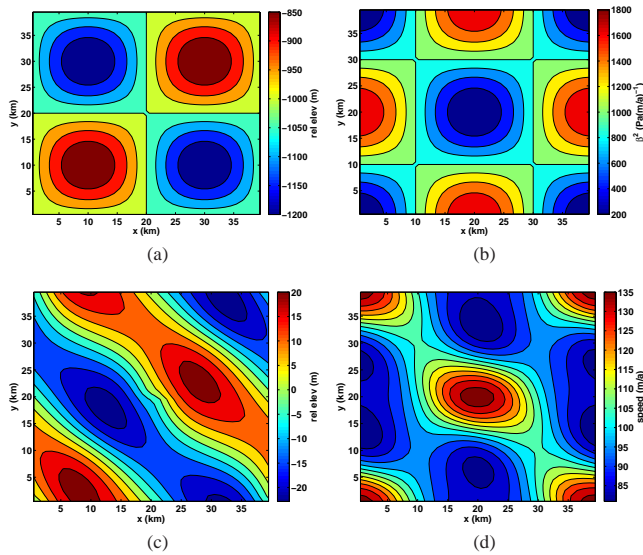


Fig. 6. (a) True basal topography R in Experiment 4, with constant trend in x -direction removed. (b) True β^2 in Experiment 4. (c) Steady-state surface elevation, linear trend removed. (d) Ice surface speed corresponding to steady state.

We set up our identical-twin experiment by assuming perfect knowledge of surface elevation and surface velocity, and little or no information about basal topography and basal lubrication, aside from their general scales. We define a cost function

$$J = \frac{1}{N\sigma_u^2} \sum_i^N |\mathbf{u}_s^*(i) - \mathbf{u}_s(i)|^2 + \frac{\gamma_d}{N\Delta t^2} \sum_i^N (h^{(1)}(i) - h^{(0)}(i))^2 + \frac{\gamma_o}{N} \|\nabla h^{(0)}\|^2 + \gamma_b \sum_i^N \exp\left(\frac{\beta(i)^2}{\beta_{\max}^2}\right), \quad (22)$$

where the summation is over all cells i . The forward model runs for a single time step, and \mathbf{u}_s and $h^{(1)}$ are the model surface velocity and thickness after that time step. \mathbf{u}_s^* is the observed surface velocity shown in Fig. 6d. The first term in Eq. (22) penalizes misfit with observed velocities. σ_u is as in Sect. 5.3. The second term penalizes model drift, i.e., the amount by which thickness changes over the single time step. If we were constraining the rate of thickness change to be close to a nonzero observed rate, the term would be different, but here we are constraining the model to be close to steady state. γ_d is chosen so that the term is order unity when thickness rate of change is on the order of 10^{-4} .

The third and fourth terms in (Eq. 22) pertain to the “model” norm, rather than the misfit norm (Waddington et al., 2007), meaning they deal with a priori knowledge of the unknown parameters. The third term is a simple Tikhonov smoothing term that penalizes large oscillation in thickness (and bed topography; see below). The norm $|\cdot|$ is that induced by summing the square of the centered finite difference approximation to the gradient over all cells. γ_o is chosen so that the term only makes a contribution if thickness gradients are

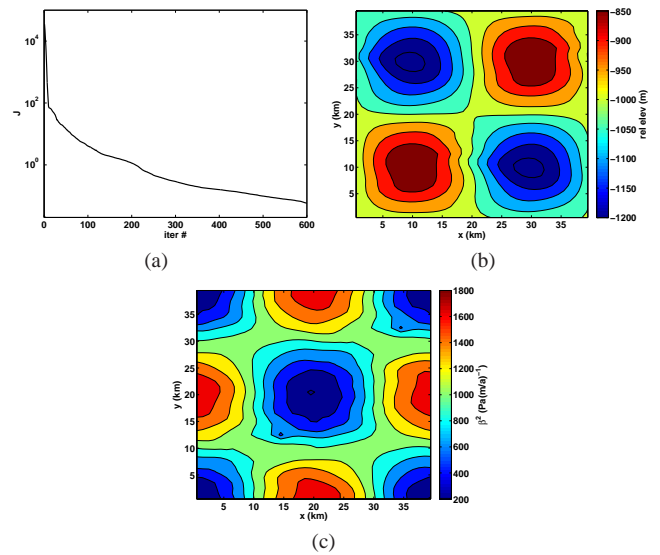


Fig. 7. Inversion for basal topography and lubrication with uniform initial guesses (basal elevation $R \equiv -1000$ m, $\beta^2 \equiv 400$ Pa m a⁻¹). (a) Cost function versus forward and adjoint iteration number. (b) Inverted basal topography; compare with Fig. 6a. rms error is 5.9 m, compared with an initial value of 100 m. (c) Inverted β^2 .

larger than ~ 0.1 on average. The fourth term penalizes basal lubrication if it becomes too large during the minimization. It was found that without this term, the minimization algorithm can sometimes favor extremely large values of β^2 in favor of making adjustments to initial thickness, and this term helps to prevent that. For the experiment shown here this term was found not to be necessary.

The basal lubrication β and the bed elevation R are the control variables in the minimization of Eq. (22). The surface elevation, s , is constrained to be that of the computed steady state, s_{ss} (the same as that shown in Fig. 6c). Rather than penalizing the misfit of s in the cost function, the surface elevation is controlled exactly through the initial guess for ice thickness. That is, the initial guess for thickness, $h_{\text{guess}}^{(0)}$, and topography, R_{guess} , are defined such that

$$h_{\text{guess}}^{(0)} + R_{\text{guess}} = s_{ss}. \quad (23)$$

When the guess for R is updated by a function δR , the topography $h^{(0)}$ is updated by $-\delta R$.

The initial guesses for basal lubrication and topography are

$$\beta_{\text{guess}}^2(x, y) \equiv 400 \text{ Pa (m/a)}^{-1}, \quad (24)$$

$$R_{\text{guess}}(x, y) = R_0(x). \quad (25)$$

The results of this inversion are shown in Fig. 7. While there is significant reduction in J , indicating not only good agreement of model and observed velocities but also small model drift (i.e., steady state is achieved), J does not include

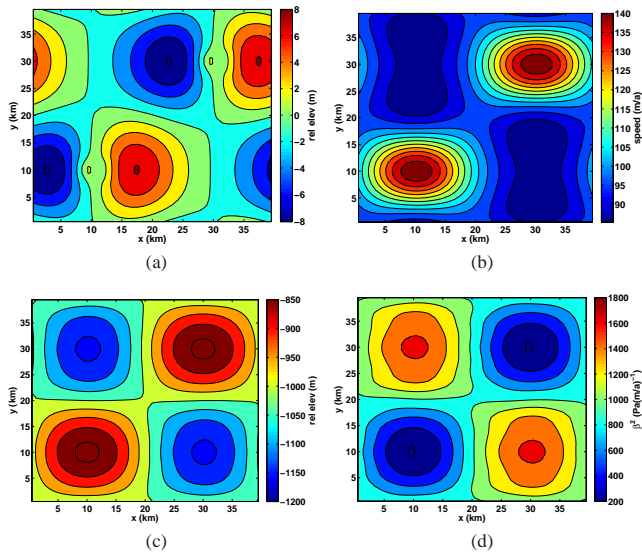


Fig. 8. An inversion similar to Figs. 6–7, but with β^2_{true} “in phase” with R_{true} , such that the sticky spots coincide with the topographic minima. Initial guesses for R and β^2 are the same as before. (a) Steady-state surface elevation with trend removed. (b) Steady-state surface velocity. (c) Inverted basal topography. rms error is 16 m. Inverted topographic minima are smaller in amplitude than the “true” topography by ~ 50 . This is compensated by lower β^2 in those regions (d).

a measure of the misfit of R and β^2 . This can be assessed directly, however, since the true values of these fields are known. The inverted R and β^2 fields compare well with their true values; the root-mean-square (rms) error in the inverted basal topography is ~ 6 m (reduced from 100 m in the initial guess).

Such accuracy cannot always be expected. To demonstrate this an equally simple identical-twin experiment is carried out. It is similar to the experiment discussed above, only the basal traction is in phase with the basal topography, such that the “sticky spots” that can be seen in Fig. 6b are shifted and coincide with the topographic lows in Fig. 6a. Again, there is no specific physical motivation behind the spatial relationship between basal lubrication and topography. Rather, it is to examine the degree of compensation between the two parameters in their inverted fields: since both lubrication and topography control velocity, it is possible that the inversion is nonunique. This type of compensation is similar to the “mixing” referred to by Gudmundsson and Raymond (2008).

The results of this experiment are shown in Fig. 6a. While the reduction in the cost function J is similar (not shown), the inversion is not as accurate, with an rms error in inverted topography of ~ 15 m. The regions of difficulty can be seen to be those of the topographic lows, where the inverted basal topography perturbation is about 50 m less than the true value. At the same time, inverted basal traction is lower than its true value here. The smaller thickness is compensated by the

weaker bed in order to match the observed speeds. Over the topographic bumps, inverted values are more accurate. We draw the tentative conclusion that, when bed strength and topography are uncorrelated, their inversions are more accurate with the procedure used here.

5.4.2 Uncertainties in bed elevation and surface elevation

In the above experiments, the *exact* surface elevation $s_{\text{SS}}(x, y)$ corresponding to the steady state of the model was used as a *hard* constraint, i.e., the initial surface elevation was constrained to be exactly equal to s_{SS} , no matter the guess for bed topography or lubrication. (The surface elevation after a single time step was obviously not directly constrained.) The rationale behind this choice was the relative magnitudes of error in measurements of elevation and ice thickness in general. However, one might ask how sensitive the above results are to uncertainties in surface elevation. We carry out additional experiments of two distinct types: in the first, we still impose a hard constraint on surface elevation, but we constrain to an inexact estimate of s_{SS} . In the second we impose a *soft* constraint on surface, i.e., we add an additional penalty term to the cost function, and allow for an additional control. In these experiments, only runs where the true topography and lubrication are out of phase (as in Fig. 6) are considered.

In the hard-constrained experiments, we perturb s_{SS} to produce our error-prone surface s_0 with a mode 2 pattern:

$$s_0 = s_{\text{SS}} + \eta_s \sin\left(\frac{4\pi y}{40}\right) \sin\left(\frac{4\pi x}{40}\right), \quad (26)$$

where η_s is a specified amplitude. s_0 now takes the place of s_{SS} in Eq. (23). The results are not shown, but when $\eta_s = 0.1$ m (implying accuracy higher than that achieved by satellite or airborne altimetry), the inverted patterns of lubrication and topography are similar to those shown in Fig. 7c and b, and the *rms* topography error is 5.6 m, similar to that achieved when constraining to s_{SS} . When $\eta_s = 0.5$ m, a slightly more realistic value, the patterns are again similar, but with a rms error of 11.4 m still a significant decrease in error relative to the initial guess. However, one can assume the accuracy of inversion will worsen when errors are larger, or when true topography and lubrication are correlated.

In our final experiment we weaken the constraint on surface elevation: we do away with Eq. (23) and instead allow for independent controls of β , R , and h . To the cost function given in Eq. (22) we add the term J_s , defined by

$$J_s = \frac{1}{N\sigma_s^2} \sum_i^N \left(s^{(1)}(i) - s_0(i) \right)^2, \quad (27)$$

where s_0 is as in Eq. (26). Both η and σ_s are set to 0.1 m, implying an accuracy of 0.1 m in surface elevation measurements. Note that the constraint is on the surface *after* a time step is taken, in contrast to the other experiments, but this

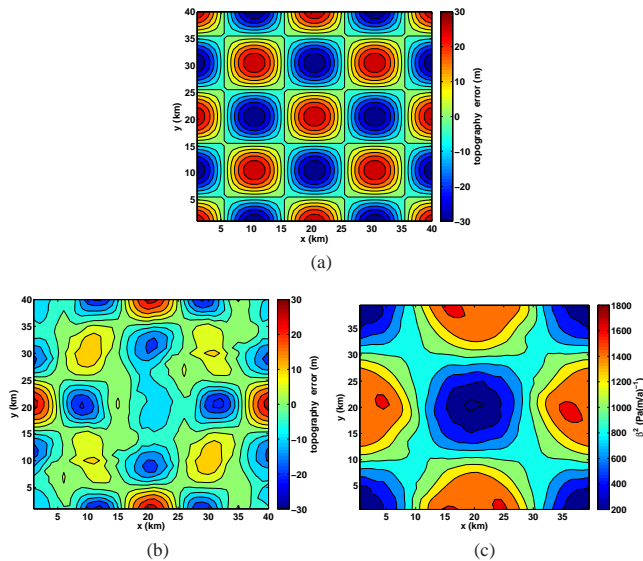


Fig. 9. (a) The error inherent in the initial guess for bed topography, weakly constrained surface elevation. (b) The error in bed topography at the end of the inversion. Note that error is reduced where surface velocities are high or true bed topography is relatively flat. (c) Corresponding inversion for β^2 .

should not matter if the system is in steady state. An experiment was carried out with an initial guess of a flat bed, which did not lead to the recovery of an accurate bed topography (not shown). However, the level of accuracy in the prior bed topography estimation in the above experiments was quite low; in practice bed elevation is known at least somewhat accurately from ice-penetrating radar, or seismic inference, along with interpolation. Therefore we choose an initial guess with an error pattern similar to that of s_0 , and with magnitudes of 30 m rather than 200 m:

$$R_{\text{guess}}(x, y) = R(x, y) + 30 \cos\left(\frac{4\pi y}{40}\right) \cos\left(\frac{4\pi x}{40}\right). \quad (28)$$

The initial guess for β was as in the previous experiments.

The results of this experiment are shown in Fig. 9. Bed topography is depicted in terms of error with respect to the true bed $R(x, y)$, not in absolute (i.e., with respect to sea level). Overall, the reduction in rms bed error was not large (from 15 to 8.3 m.) Nevertheless, the pattern of error is instructive: by comparison between Figs. 6a, b and 9b, it seems that error reduction was significant in areas of high surface velocity or relatively flat bed. Inverted bed lubrication is shown as well, but in an absolute sense, as the initial guess is uniform. The pattern is qualitatively correct, but there are obvious errors, to be expected given the related errors in bed topography.

Here we have examined two approaches that allow for uncertainty in surface elevation measurements; but for a similar order of accuracy, the first (hard-constrained) produces a far more accurate estimate of bed topography and lubrication than the second (soft-constrained). It seems that when the

thickness and bed are independent controls, the minimization scheme is free to find agreement with the imposed surface elevation at the expense of the constraint on elevation change. Since both methods incorporate uncertainties in surface elevation, we therefore suggest the former (hard constraint on surface elevation) when recovering bed topography and lubrication based on surface information. However, one would not be able to use such hard constraints for a time-dependent inversion, such as that of Sect. 5.3.

6 Conclusions and further work

Using a synthesis of AD tools and analytical considerations, we have successfully generated the adjoint of an ice model which accounts for both a time-dependent mass balance and a nonlocal, higher-order stress balance. The adjoint model is capable of providing transient sensitivities, allowing for exploration of very large parameter spaces. Coupled with a large-scale optimization algorithm, the adjoint is able to successfully perform the inversions that are typically carried out in glaciological settings (retrieval of basal lubrication parameters) and some that are not (simultaneous retrieval of basal topography and past thickness). The fact that the adjoint is time-dependent means that an evolving model state can be found consistent with observations, enabling ice model initialization without costly spinups or artificial flux adjustments. Perhaps most importantly, this opens up a fundamentally different approach to the use of ice observations for constraining transient ice flow, in particular with respect to satellite remote sensing data that are resolved in time.

The process of adjoint generation is streamlined: a change to the ice model requires simply another application step of AD tools, rather than the derivation of a new set of adjoint equations. The forward and adjoint models are fully parallel, meaning the costly solves of large linear systems can be spread across a number of processors. Currently the linear solver is a very simple conjugate gradient algorithm with a trivial preconditioner. However, the decoupling of the matrix solve from the AD tools means that the efficacy of different solvers can be investigated without affecting the adjoint model. A possible future development is to replace our solver with an external package for higher efficiency, similarly to other ice models intended for large-scale studies (Bueler and Brown, 2009; Larour et al., 2012b).

A number of developments remain to be made in our model in order to apply it to many relevant glaciological issues. As mentioned in Sect. 5.1, the flotation condition at the grounding line introduces nondifferentiability in the expressions for basal and driving stress. In its present form sensitivity studies cannot account for grounding line migration, and neither can observation-based inversions involving ice thickness and basal topography as controls. In other adjoint studies these types of issues have been approached by “smoothing out” these nondifferentiabilities

(Losch and Heimbach, 2007); we will investigate whether grounding line issues can be dealt with by such smoothing.

An important physical process currently absent in our model is temperature transport. Ice internal temperature, while difficult to observe on large scales, can play an important role in ice flow and in the controls on basal melting and freezing of grounded ice. Many parameter estimation studies assume temperature is steady (Joughin et al., 2009; Morlighem et al., 2010), but there is evidence in certain locations that there is ongoing thermal adjustment due to unsteady transport (Engelhardt, 2004), and complex thermo-mechanical coupling in shear margins (Schoof, 2012). Using the existing tracer transport framework of MITgcm, we plan to implement thermal advection/diffusion in our ice model. This will enable investigations of the adjoint sensitivities of ice evolution to poorly constrained ice internal temperatures, and to even less well-known geothermal fluxes. Such sensitivities are important to constrain, especially for areas undergoing large change, such as Pine Island (Larour et al., 2012a).

Our forward and adjoint model lay the groundwork for a fully coupled land ice–ice shelf–ocean model that are based on the same model infrastructure. Heimbach and Losch (2012) used an ocean model adjoint to investigate the sensitivities of melt rates under Pine Island ice shelf to ocean forcings. Our ice model is not currently coupled to an ocean model, but such coupling studies have been carried out (without the capability of adjoint generation; Goldberg et al., 2012a). Similarly coupling our model to the MITgcm ocean model (for which an adjoint exists) would allow adjoint sensitivities to then be propagated across the ice–ocean interface, and the sensitivities of ice evolution to ocean forcings could be investigated, among other questions. As demonstrated in Sect. 5.1, the structural “weak points” of an ice shelf do not coincide with the places of strongest melting, or necessarily with those areas where melt rates are most sensitive to a given ocean forcing. A coupled ice–ocean adjoint model could reveal which ocean trends are relevant to land ice evolution, which is not currently known.

Lastly, we point out the ubiquity of our approach, as we believe many higher-order ice model codes could potentially be amenable to automated differentiation techniques, as long as the costly linear solvers are hidden from the AD tools (although there may be complicating factors not considered here, such as discrete grid remeshing). Given the utility of being able to generate an adjoint code, we make the suggestion that next-generation ice model code be written with this in mind.

Acknowledgements. D. N. Goldberg was supported by an NSF-OPP postdoctoral fellowship (ANT-1103375). P. Heimbach was supported in part by an NSF “Collaboration in Mathematical Geosciences” (CMG) grant No. 0934404 and DOE/SciDAC grant No. SC0008060 (PISCEES). Valuable comments by reviewers M. Morlighem and M. Nodet, as well as the editor, O. Gagliardini, are gratefully acknowledged.

Edited by: O. Gagliardini

References

- Albrecht, T., Martin, M., Winkelmann, R., Haseloff, M., and Levermann, A.: Parameterization for Subgrid-Scale Motion of Ice-Shelf Calving-Fronts, *The Cryosphere*, 5, 35–44, 2011, <http://www.the-cryosphere-discuss.net/5/35/2011/>.
- Arthern, R. J. and Gudmundsson, G. H.: Initialization of ice-sheet forecasts viewed as an inverse Robin problem, *J. Glaciol.*, 56, 527–533, 2010.
- Blatter, H.: Velocity and stress fields in grounded glaciers: a simple algorithm for including deviatoric stress gradients, *J. Glaciol.*, 41, 333–344, 1995.
- Brinkerhoff, D. J., Meierbachtol, T. W., Johnson, J. V., and Harper, J. T.: Sensitivity of the frozen/melted basal boundary to perturbations of basal traction and geothermal heat flux: Isunguata Sermia, western Greenland, *Ann. Glaciol.*, 52, 43–50, doi:10.3189/172756411799096330, 2011.
- Bueler, E. and Brown, J.: The shallow shelf approximation as a ‘sliding law’ in a thermomechanically coupled ice sheet model, *J. Geophys. Res.-Earth Surf.*, 114, F03008, doi:10.1029/2008JF001179, 2009.
- Colton, D.: *Inverse Acoustic and Electromagnetic Scattering Theory*, Second Edn., 1998.
- Dupont, T. K. and Alley, R.: Assessment of the importance of ice-shelf buttressing to ice-sheet flow, *Geophys. Res. Lett.*, 32, L04503, doi:10.1029/2004GL022024, 2005.
- Durand, G., Gagliardini, O., Favier, L., Zwinger, T., and le Meur, E.: Impact of bedrock description on modeling ice sheet dynamics, *Geophysical Research Letters*, 38, L20501, doi:10.1029/2010GL043334, 2011.
- Engelhardt, H.: Thermal regime and dynamics of the West Antarctic ice sheet, *Ann. Glaciol.*, 39, 85–91, 2004.
- Errico, R. M.: What is an adjoint model?, *BAMS*, 78, 2577–2591, 1997.
- Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barand, 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., Forsberg, R., Fujita, S., Gim, Y., Gogineni, P., Griggs, J. A., Hindmarsh, R. C. A., Holmlund, P., Holt, J. W., Jacobel, R. W., Jenkins, A., Jokat, W., Jordan, T., King, E. C., Kohler, J., Krabill, W., Riger-Kusk, M., Langley, K. A., Leitchenkov, G., Leuschen, C., Luyendyk, B. P., Matsuoka, K., Mouginot, J., Nitsche, F. O., Nogi, Y., Nost, O. A., Popov, S. V., Rignot, E., Rippon, D. M., Rivera, A., Roberts, J., Ross, N., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tinto, B. K., Welch, B. C., Wilson, D., Young, D. A., Xiangbin, C., and Zirizzotti, A.: Bedmap2: improved ice bed, surface and thickness datasets for

- Antarctica, *The Cryosphere*, 7, 375–393, doi:10.5194/tc-7-375-2013, 2013.
- Gagliardini, O., Durand, G., Zwinger, T., Hindmarsh, R. C. A., and Meur, E. L.: Coupling of ice shelf melting and buttressing is a key process in ice sheet dynamics, *Geophys. Res. Lett.*, doi:10.1029/2010GL043334, in press, 2013.
- Giering, R., Kaminski, T., and Slawig, T.: Generating efficient derivative code with TAF adjoint and tangent linear Euler flow around an airfoil, *Future Gener. Comput. Syst.*, 21, 1345–1355, doi:10.1016/j.future.2004.11.003, 2005.
- Gilbert, J. C. and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-Newton algorithms, *Math. Program.*, 45, 407–435, doi:10.1007/BF01589113, 1989.
- 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, doi:10.5194/tc-6-1561-2012, 2012.
- Goldberg, D. N.: A variationally-derived, depth-integrated approximation to a higher-order glaciological flow model, *J. Glaciol.*, 57, 157–170, 2011.
- Goldberg, D. N. and Sergienko, O. V.: Data assimilation using a hybrid ice flow model, *The Cryosphere*, 5, 315–327, doi:10.5194/tc-5-315-2011, 2011.
- Goldberg, D. N., Holland, D. M., and Schoof, C. G.: Grounding line movement and ice shelf buttressing in marine ice sheets, *J. Geophys. Res.-Earth Surf.*, 114, F04026, doi:10.1029/2008JF001227, 2009.
- Goldberg, D. N., Little, C. M., Sergienko, O. V., Gnanadesikan, A., Hallberg, R., and Oppenheimer, M.: Investigation of land ice-ocean interaction with a fully coupled ice-ocean model, Part 1: Model description and behavior, *J. Geophys. Res.-Earth Surf.*, 117, F02037, doi:10.1029/2011JF002247, 2012a.
- Goldberg, D. N., Little, C. M., Sergienko, O. V., Gnanadesikan, A., Hallberg, R., and Oppenheimer, M.: Investigation of land ice-ocean interaction with a fully coupled ice-ocean model, Part 2: Sensitivity to external forcings, *J. Geophys. Res.-Earth Surf.*, 117, F02038, doi:10.1029/2011JF002246, 2012b.
- Greve, R.: A continuum-mechanical formulation for shallow polythermal ice sheets, *Phil. Trans. Royal Soc. of London, Series A*, 355, 921–974, 1997.
- Greve, R. and Blatter, H.: *Dynamics of Ice Sheets and Glaciers*, Springer, Dordrecht, 2009.
- Griewank, A. and Walther, A.: *Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation*, vol. 19 of *Frontiers in Applied Mathematics*, SIAM, Philadelphia, 2nd Edn., 2008.
- Griggs, J. A. and Bamber, J. L.: A new 1 km digital elevation model of Antarctica derived from combined radar and laser data – Part 2: Validation and error estimates, *The Cryosphere Discussions*, 2, 843–872, doi:10.5194/tcd-2-843-2008, 2008.
- Gudmundsson, G. H.: Transmission of basal variability to a glacier surface, *J. Geophys. Res.: Solid Earth*, 108, 2253, doi:10.1029/2002JB002107, 2003.
- Gudmundsson, G. H. and Raymond, M.: On the limit to resolution and information on basal properties obtainable from surface data on ice streams, *The Cryosphere*, 2, 167–178, doi:10.5194/tc-2-167-2008, 2008.
- Heimbach, P. and Bugnion, V.: Greenland ice-sheet volume sensitivity to basal, surface and initial conditions derived from an adjoint model, *Ann. Glaciol.*, 50, 67–80, 2009.
- Heimbach, P. and Losch, M.: Adjoint sensitivities of sub-ice shelf melt rates to ocean circulation under Pine Island Ice Shelf, West Antarctica, *Ann. Glaciol.*, 54, 59–69, doi:10.3189/2012/AoG60A025, 2012.
- Heimbach, P., Hill, C., and Giering, R.: An efficient exact adjoint of the parallel MIT General Circulation Model, generated via automatic differentiation, *Future Generation Computer Systems*, 21, 1356–1371, doi:10.1016/j.future.2004.11.010, 2005.
- Hellmer, H., Jacobs, S., and Jenkins, A.: Oceanic erosion of a floating Antarctic glacier in the Amundsea Sea, in: *Ocean, Ice, and Atmosphere: Interactions at the Antarctic Continental Margin*, AGU Antarctic Research Series Vol 75, edited by: Jacobs, S. S. and Weiss, R. F., AGU, 83–99, 1998.
- Hindmarsh, R. C. A.: A numerical comparison of approximations to the Stokes equations used in ice sheet and glacier modeling, *J. Geophys. Res.*, 109, F01012, doi:10.1029/2003JF000065, 2004.
- Holland, P. R., Jenkins, A., and Holland, D. M.: The Response of Ice Shelf Basal Melting to Variations in Ocean Temperature, *J. Climate*, 21, 2558–2572, 2008.
- Hutter, K.: *Theoretical Glaciology*, Dordrecht, Kluwer Academic Publishers, 1983.
- Jacobs, S. S.: Observations of change in the Southern Ocean, *Phil. Trans. Royal Soc. of London, Series A*, 364, 1657–1681, 2006.
- Jacobs, S. S., Hellmer, H., and Jenkins, A.: Antarctic ice sheet melting in the Southeast Pacific, *Geophys. Res. Lett.*, 23, 957–960, 1996.
- Jacobs, S. S., Jenkins, A., Giulivi, C., and Dutrieux, P.: Stronger ocean circulation and increased melting under Pine Island Glacier ice shelf, *Nat. Geosci.* 4, 519–523, doi:10.1038/NGEO1188, 2011.
- Jay-Allemand, M., Gillet-Chaulet, F., Gagliardini, O., and Nodet, M.: Investigating changes in basal conditions of Variegated Glacier prior to and during its 1982–1983 surge, *The Cryosphere*, 5, 659–672, doi:10.5194/tc-5-659-2011, 2011.
- Jenkins, A.: A one-dimensional model of ice shelf-ocean interaction, *J. Geophys. Res.*, 96, 20671–20677, doi:10.1029/91JC01842, 1991.
- Jenkins, A., Dutrieux, P., Jacobs, S. S., McPhail, S. D., Perrett, J. R., Webb, A. T., and White, D.: Observations beneath Pine Island Glacier in West Antarctica and implications for its retreat, *Nat. Geosci.*, 3, 468–472, 2010.
- 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, *J. Glaciol.*, 55, 245–257, 2009.
- Khazendar, A., Rignot, E., and Larour, E.: Larsen B ice shelf rheology preceding its disintegration inferred by a control method, *Geophys. Res. Lett.*, 34, L19503, doi:10.1029/2007GL030980, 2007.
- Larour, E., Morlighem, M., Seroussi, H., Schiermeier, J., and Rignot, E.: Ice flow sensitivity to geothermal heat flux of Pine Island Glacier, Antarctica, *Journal of Geophysical Research: Earth Surface*, 117, F04023, doi:10.1029/2012JF002371, 2012a.
- 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), *J. Geophys. Res.-Earth Surf.*, 117, F01022, doi:10.1029/2011JF002140, 2012b.
- Little, C. M., Gnanadesikan, A., and Oppenheimer, M.: How ice shelf morphology controls basal melting, *J. Geophys. Res.*, 114, C12007, doi:10.3189/2012JoG11J037, 2009.
- Little, C. M., Goldberg, D. N., Gnanadesikan, A., and Oppenheimer, M.: On the coupled response to ice-shelf basal melting, *J. Glaciol.*, 58, 203–215, 2012.
- Losch, M. and Heimbach, P.: Adjoint sensitivity of an ocean general circulation model to bottom topography, *J. Phys. Ocean.*, 37, 377–393, doi:10.1175/JPO3017.1, 2007.
- MacAyeal, D. R.: Thermohaline circulation below the Ross Ice Shelf: A consequence of tidally induced vertical mixing and basal melting, *J. Geophys. Res.*, 89, 597–606, 1984.
- MacAyeal, D. R.: Large-scale ice flow over a viscous basal sediment: Theory and application to Ice Stream B, Antarctica, *J. Geophys. Res.-Solid Earth Planets*, 94, 4071–4087, 1989.
- MacAyeal, D. R.: The basal stress distribution of Ice Stream E, Antarctica, inferred by control methods, *J. Geophys. Res.*, 97, 595–603, 1992.
- MacAyeal, D. R.: A tutorial on the use of control methods in ice-sheet modeling, *J. Glaciol.*, 39, 91–98, 1993.
- Marshall, J., Hill, C., Perelman, L., and Adcroft, A.: Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling, *J. Geophys. Res.*, 102, 5733–5752, doi:10.1029/96JC02776, 1997.
- Maxwell, D., Truffer, M., Avdonin, S., and Stuefer, M.: An iterative scheme for determining glacier velocities and stresses, *Ann. Glaciol.*, 36, 197–204, 2008.
- Morland, L. W.: Unconfined ice-shelf flow, in: *Dynamics of the West Antarctic Ice Sheet*, edited by de Veen, C. J. V. and Oerlemans, J., 99–116, Reidel Publ. Co, 1987.
- Morlighem, M., Rignot, E., Seroussi, G., 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, *Geophys. Res. Lett.*, 37, L14502, doi:10.1029/2010GL043853, 2010.
- Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., and Aubry, D.: A mass conservation approach for mapping glacier ice thickness, *Geophys. Res. Lett.*, 38, L19503, doi:10.1029/2011GL048659, 2011.
- Pattyn, F.: A new three-dimensional higher-order thermomechanical ice-sheet model: basic sensitivity, ice-stream development and ice flow across subglacial lakes, *J. Geophys. Res.-Solid Earth Planets*, 108, 2382, doi:10.1029/2002JB002329, 2003.
- Pattyn, F., Perichon, L., Aschwanden, A., Breuer, B., de Smedt, B., Gagliardini, O., Gudmundsson, G. H., Hindmarsh, R. C. A., Hubbard, A., Johnson, J. V., Kleiner, T., Kononov, Y., Martin, C., Payne, A. J., Pollard, D., Price, S., Rückamp, M., Saito, F., Soucek, O., Sugiyama, S., and Zwinger, T.: Benchmark experiments for higher-order and full-Stokes ice sheet models (ISMIP-HOM), *The Cryosphere*, 2, 95–108, doi:10.5194/tc-2-95-2008, 2008.
- Petra, N., Zhu, H., Stadler, G., Hughes, T. J., and Ghattas, O.: An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model, *J. Glaciol.*, 58, 889–903, doi:10.3189/2012JoG11J182, 2012.
- Pollard, D. and DeConto, R. M.: Modelling West Antarctic Ice Sheet growth and collapse through the past five million years, *Nature*, 458, 329–332, 2009.
- Raymond, M. J. and Gudmundsson, G. H.: Estimating basal properties of ice streams from surface measurements: a non-linear Bayesian inverse approach applied to synthetic data, *The Cryosphere*, 3, 265–278, 2009, <http://www.the-cryosphere-discuss.net/3/265/2009/>.
- Rignot, E.: Fast Recession of a West Antarctic Glacier, *Science*, 281, 549–551, 1998.
- Rignot, E., Vaughan, D. G., Schmeltz, M., Dupont, T. K., and MacAyeal, D. R.: Acceleration of Pine Island and Thwaites Glaciers, West Antarctica, *Ann. Glaciol.*, 34, 189–194, 2002.
- Schoof, C.: Ice sheet grounding line dynamics: Steady states, stability, and hysteresis, *J. Geophys. Res.*, 112, F03238, doi:10.1029/2006JF000664, 2007.
- Schoof, C.: Thermally driven migration of ice-stream shear margins, *Journal of Fluid Mechanics*, 712, 552–578, doi:10.1017/jfm.2012.438, http://journals.cambridge.org/article_S0022112012004387, 2012.
- Schoof, C. and Hindmarsh, R. C. A.: Thin-Film Flows with Wall Slip: An Asymptotic Analysis of Higher Order Glacier Flow Models, *Quart. J. Mech. Appl. Math.*, 63, 73–114, 2010.
- Sergienko, O. V.: The effects of transverse bed topography variations in ice-flow models, *J. Geophys. Res.-Earth Surf.*, 117, F03011, doi:10.1029/2011JF002203, 2012.
- Sergienko, O. V., Bindshadler, R. A., Vornberger, P. L., and MacAyeal, D. R.: Ice stream basal conditions from block-wise surface data inversion and simple regression models of ice stream flow: Application to Bindshadler Ice Stream, *J. Geophys. Res.*, 113, F04010, doi:10.1029/2008JF001004, 2008.
- Seroussi, H., Morlighem, M., Rignot, E., Larour, E., Aubry, D., Ben Dhia, H., and Kristensen, S. S.: Ice flux divergence anomalies on 79north Glacier, Greenland, *Geophysical Research Letters*, 38, L09501, doi:10.1029/2011GL047338, 2011.
- Shepherd, A., Wingham, D. J., and Mansley, J.: Inland thinning of the Amundsen Sea sector, West Antarctica, *Geophys. Res. Lett.*, 29, L1364, doi:10.1029/2001GL014183, 2002.
- Shepherd, A., Wingham, D. J., and Rignot, E.: Warm ocean is eroding West Antarctic Ice Sheet, *Geophys. Res. Lett.*, 31, L23402, doi:10.1029/2004GL021106, 2004.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C. N., and Marshall, J.: Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model, *J. Geophys. Res.*, 107, 1–27, doi:10.1029/2001JC000888, 2002.
- Thacker, W. C. and Long, R. B.: Fitting dynamics to data, *J. Geophys. Res.*, 93, 1227–1240, doi:10.1029/JC093iC02p01227, 1988.
- Thomas, R. H.: The dynamics of marine ice sheets, *J. Glaciol.*, 31, 347–356, 1979.
- Thorsteinsson, T., Raymond, C. F., Gudmundsson, G. H., Bindshadler, R. A., Vornberger, P., and Joughin, I.: Bed topography and lubrication inferred from surface measurements on fast-flowing ice streams, *J. Glaciol.*, 49, 481–490, 2003.
- Tziperman, E. and Thacker, W. C.: An Optimal-Control/Adjoint-Equations Approach to Studying the Oceanic General Circulation, *J. Phys. Ocean.*, 19, 1471–1485, doi:[http://dx.doi.org/10.1175/1520-0485\(1989\)019<1471:AOCEAT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0485(1989)019<1471:AOCEAT>2.0.CO;2), 1989.
- Utke, J., Naumann, U., Fagan, M., Tallent, N., Strout, M., Heimbach, P., Hill, C., and Wunsch, C.: OpenAD/F: A Modular, Open-

- Source Tool for Automatic Differentiation of Fortran Codes, *ACM Transactions on Mathematical Software*, 34, 18:1–18:36, doi:10.1145/1377596.1377598, 2008.
- Vieli, A. and Payne, A. J.: Application of control methods for modelling the flow of Pine Island Glacier, West Antarctica, *Annals of Glaciol.*, 36, 197–204, 2003.
- Waddington, E., Neumann, T., Koutnik, M., Marshall, H., and Morse, D.: Inference of accumulation-rate patterns from deep layers in glaciers and ice sheets, *J. Glaciol.*, 53, 694–712, 2007.
- Walker, R. T., Dupont, T. K., Parizek, B. R., and Alley, R. B.: Effects of basal melting distribution on the retreat of ice-shelf grounding lines, *Geophys. Res. Lett.*, 35, L17503, doi:10.1029/2008GL034947, 2008.
- Weertman, J.: Deformation of floating ice shelves, *J. Glaciol.*, 3, 38–42, 1957.
- Wunsch, C. and Heimbach, P.: Dynamically and kinematically consistent global ocean circulation and ice state estimates, in: *Ocean circulation and climate: A 21st century perspective*, edited by: Siedler, G., Church, J., Gould, J., and Griffies, S., 553–579, Elsevier, 2013.
- Wunsch, C., Ponte, R., Heimbach, P., and Fukumori, I.: The Global General Circulation of the Ocean Estimated by the ECCO-Consortium, *Oceanography*, 22, 88–103, 2009.