Modeling dynamic controls on ice streams: a Bayesian statistical approach

L.M. BERLINER[1], K. JEZEK[2], N. CRESSIE[1], Y. KIM[1], C.Q. LAM[1], C.J. VAN DER VEEN[3]

[1]Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210-1247, USA
E-mail: mb@stat.osu.edu
[2]Byrd Polar Research Center, The Ohio State University, 1090 Camack Road, Columbus, Ohio 43210-1002, USA
[3]Department of Geography, The University of Kansas, 1475 Jayhawk Boulevard, Lawrence, Kansas 66045, USA

ABSTRACT. Our main goal is to exemplify the study of ice-stream dynamics via Bayesian statistical analysis incorporating physical, though imperfectly known, models using data that are both incomplete and noisy. The physical–statistical models we propose account for these uncertainties in a coherent, hierarchical manner. The initial modeling assumption estimates basal shear stress as equal to driving stress, but subsequently includes a random corrector process to account for model error. The resulting stochastic equation is incorporated into a simple model for surface velocities. Use of Bayes' theorem allows us to make inferences on all unknowns given basal elevation, surface elevation and surface velocity. The result is a posterior distribution of possible values that can be summarized in a number of ways. For example, the posterior mean of the stress field indicates average behavior at any location in the field, and the posterior standard deviations describe associated uncertainties. We analyze data from the 'Northeast Greenland Ice Stream' and illustrate how scientific conclusions may be drawn from our Bayesian analysis.

1. INTRODUCTION

1.1. Background

Modern studies of the behavior of glaciers, ice sheets and ice streams rely heavily on both observations and physically based models. Data acquired via remote sensing provide critical information on geometry and movement of ice over large sections of the Antarctic and Greenland. Although these datasets represent significant advancements in terms of spatial coverage and the variety of processes we can observe, the physical systems to be modeled are nevertheless imperfectly observed. Uncertainties associated with measurement errors are present and physical models are also subject to uncertainties. There is therefore a need for combining observations and models in a fashion that incorporates uncertainty and quantifies its impact on conclusions.

The goal of combining models and observations is hardly new in glaciology (e.g. MacAyeal, 1993; Arthern and Hindmarsh, 2003; Gudmundsson, 2006) nor in the broad areas of the geosciences (e.g. data assimilation as practiced in numerical weather forecasting). We address the goal by formally modeling the uncertainties present, then using Bayes' theorem to deduce information about all unknowns in the data. We focus on the development of statistical models with strong reliance on physical modeling, a strategy Berliner (2003) called 'physical–statistical modeling'. This is different from traditional physical modeling, with perhaps data-based parameter estimates, and traditional statistical modeling which possibly relies upon qualitative physical reasoning.

An in-depth illustration based on data from the 'Northeast Greenland Ice Stream' (NEGIS) is presented. The physical models used are simplified approximations, but are familiar and provide an accessible arena for exemplifying the physical–statistical approach. Specifically, we develop statistically enhanced versions of a simple model of basal shear stress as the sum of driving stress and a *corrector* process representing other effects, such as lateral resistive stress.

Inference for these unobservable stresses is a primary goal here. Bayesian inference is carried out using recently acquired datasets of ice thickness, surface elevation and surface velocity.

The introduction and uncertainty analysis of the corrector process is a crucial feature of our approach. Rather than introducing the correctors, it is natural to suggest that the effects they represent should be formally modeled. This would require more assumptions and/or more data, thereby introducing additional sources of uncertainty. In any case, we do not expect any implementable physical model to remove concerns regarding its applicability nor its uncertainty. Hence, the ability not only to combine models and observations, as discussed by MacAyeal (1989), Blatter (1995) and Joughin and others (2001), but also to analyze indicators of model validity and uncertainty is paramount.

The primary output of a Bayesian analysis is a *posterior distribution*, namely the joint probability distribution for unknown quantities conditional on the observed data. Even in our simple illustration, explicit presentation of their joint distribution is not feasible. Hence, a key aspect of Bayesian analysis in such settings is the ability to generate realizations or ensembles from the posterior distribution; the posterior is then studied through statistical summaries of such ensembles. A key by-product of such techniques is the ability to provide ensembles of initial conditions for use in a dynamic forecasting model. That is, one application of the modeling here is as a method for combining models and observations in support of full dynamical prediction or forecasting.

1.2. Glaciological motivations

That glaciers flow under the force of gravity means that important factors in determining velocities include the constraints of the constitutive relationship (Whillans, 1987) and ice thickness, in combination with resistive forces acting along the sides and at the base of the glacier. The lithostatic
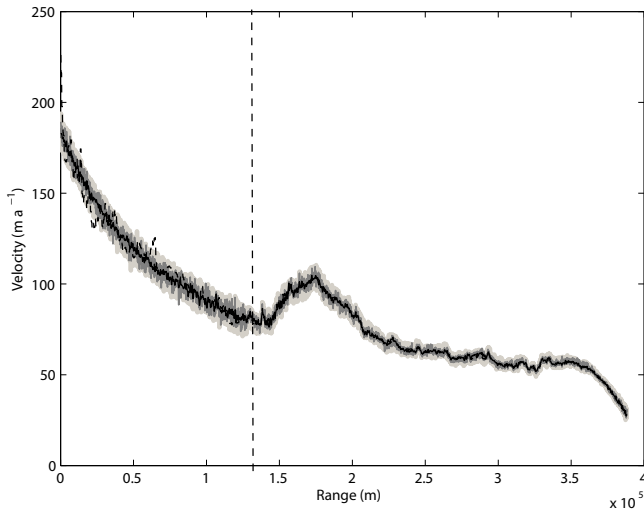
Fig. 1. Fifty posterior realizations of velocity profiles and the posterior mean (black) of velocities based on 2000 realizations. The original velocity data are shown (black dashes). Ensemble members are color-coded by importance-sampling Monte Carlo probability weights α (grey: $\alpha > 0.01$; light grey: $\alpha < 0.01$).

stress at some depth in the ice equals the weight of the ice above that depth, and horizontal gradients in this stress drive ice-stream flow. The flow of the ice stream is impeded by resistive stresses given by the difference between the full stress and the lithostatic component. A separation is therefore made between action or gravitational forces and reaction or resistive forces. After partitioning the full stress into lithostatic and resistive components, substitution into the horizontal balance equation and integration over the full ice thickness gives the following equation for driving stress along the flow (e.g. Van der Veen, 1999):

$$\tau_{dx} = \tau_{bx} + \tau_{Rx}, \tag{1}$$

where τ_{bx} is basal shear stress and τ_{Rx} represents resistive stresses arising from gradients in longitudinal stress and lateral drag. Defining the driving stress as

$$\tau_{dx} = -\rho g H \frac{ds}{dx} \tag{2}$$

where s is ice-surface elevation, H is the ice thickness, ρ is the density of ice and g is the gravity constant, we have

$$\tau_{bx} = -\rho g H \frac{ds}{dx} - \tau_{Rx}. \tag{3}$$

Noting that no direct observations of these stresses are available, the main problem in this paper is estimation of the stresses. A further limitation on such estimation involves the degree of lubrication at the base of the ice. This feature is crucial to understanding glacial dynamics and predicting the glacier's response to climatic controls. However, this feature is not observable directly.

Treating the flow parameter A as a constant, the surface velocity $u(x)$ is approximated by

$$u(x) = u_{bx} + \frac{2A}{n+1} H \tau_{bx}^n, \tag{4}$$

where u_{bx} is sliding velocity and n is a flow parameter (e.g. Paterson, 1994, p. 251, equation 21). We fix $n = 3$ in this

paper. Motivated by Equation (3), we model τ_{bx} as

$$-\tau_{bx} = \rho g H \frac{ds}{dx} + \eta_x, \tag{5}$$

where η_x is a *corrector process* needed to explain the velocity data. To clarify, we substitute Equation (5) into Equation (4) and estimate all quantities in light of the data. If η_x is found to be negative in some region, the velocities in that region are smaller than those expected based on a simple balance between driving and basal shear stresses. Hence, such η_x indicates additional resistive stresses acting on the flow. Alternatively, positive η_x suggests that the glacier is moving faster than expected, perhaps indicating an area of high lubrication at the base.

The simple interpretations above are of course flawed to some degree; we are attempting to use a single quantity η_x to explain the combined effect of at least two phenomena. For example, there may well be regions of both high lubrication and high resistive stress due to lateral effects. These are confounded in the simple approach here. This also suggests a confounding in modeling both sliding velocity u_{bx} and η_x as relatively unconstrained functions of x. We can break this confounding either by introducing very strong priors to constrain these processes or by constraining one of the processes while leaving the other relatively free and responsive to the observations.

In this paper, we compromise between these two notions. First, no matter how the modeling is carried out, we found that the data mandated the use of at least two different models. The estimated value of this *change point* between models is indicated in Figure 1. This estimate is close to an observable lineament. As will be indicated, creep flow appears to be the dominant mechanism upstream of this change point. Near the change point, sliding gains critical importance. We note that Joughin and others (2001) suggest that there is a till plain downstream of where our analysis ends. We hope our results cast light on the nature of the bed between our change point and the start of the till plain.

Based on the previous notions, we specified sliding velocity as a constant upstream of the change point, and used a relatively uninformative prior for η_x (see section 3.4). Downstream of the change point, we used approximations to Weertman-type models for sliding velocity (e.g. Paterson, 1994, ch. 7) of the form

$$u_{bx} = k\tau_{bx}^p(\rho g H)^{-q} \tag{6}$$

for various combinations of p and q with τ_{bx} given in Equation (5), i.e. without a corrector process. We found that the data strongly favored the selection of $p = q = 1$; we assume these values throughout the rest of this paper. Note that with these specifications, Equation (6) reduces to

$$u_{bx} = k \frac{ds}{dx}. \tag{7}$$

We also tried using Equation (4) with no corrector process, no change point and u_{bx} as given in Equation (6) for all x and various combinations of p and q. None of these models was competitive with the model presented here.

Another interesting issue arises in the inference of driving stress based on observations of s and H. Reliance on the slope of the upper ice surface in Equation (2) implies that results are very sensitive to small-scale variations in surface topography and to small-scale, perhaps unimportant, variations in ice thickness. Hence, driving stress is usually
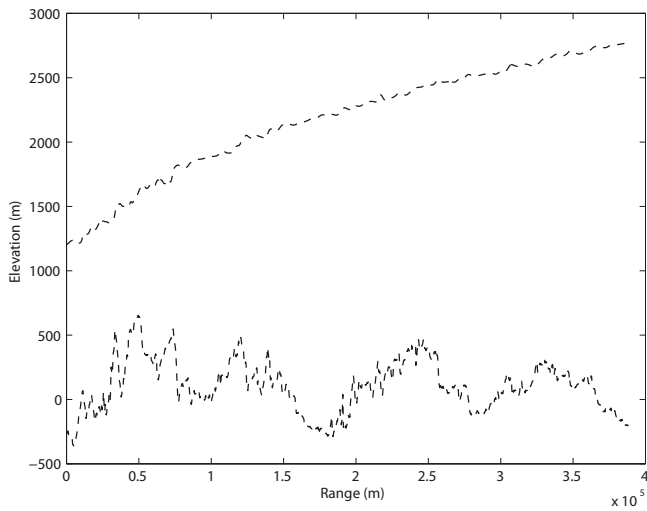
Fig. 2. Observations of surface elevation and basal elevation along a 400 km profile of the NEGIS.

Fig. 3. The basal elevation (black dashes), 50 realizations (grey) of smoothed basal topography from the posterior distribution conditional on those data, and the corresponding posterior mean (black) based on 2000 ensemble members.

estimated based on averaging over horizontal distances of a few ice thicknesses to eliminate small-scale flow features not important to the flow (Kamb and Echelmeyer, 1986; Paterson, 1994). Indeed, if averaging is not done, driving-stress estimates exhibit unreasonably large variations. Hence, we incorporate a statistical smoothing step in our Bayesian model.

1.3. Data

We assembled surface topography and ice-thickness observations for a portion of the NEGIS (Figs 2 and 3). The data were gathered as part of the Program for Arctic Regional Climate Assessment (PARCA). Surface topography and ice thickness were sampled every few hundred meters using equipment mounted on the Wallops Flight Facility P-3 aircraft. Surface velocity data were calculated (personal communication from I. Joughin, 2002) dataset. The three derived datasets are **S** (surface topography), **B** (basal topography), and **U** (surface velocities). Datasets **S** and **B** are shown in Figure 2; **U** in Figure 1.

1.4. Hierarchical Bayesian analysis

One rationale for the Bayesian approach is that it is a mathematically rigorous method for combining information in the presence of uncertainty. Two primary sources of information are available for inferring the unknown quantities of interest: (1) observations or data that convey some information regarding those unknowns, and (2) prior information based on scientific reasoning regarding the unknowns, including physical models as well as past experience and data. Both sources of information are subject to uncertainties. In the Bayesian approach, all such uncertainties are modeled probabilistically. The primary computational tool used is Bayes' theorem, which is simply a result in probability theory relating various conditional distributions. However, the Bayesian view of statistical modeling and analysis involves more than the simple application of probability theory. It is a paradigm that involves the modeling of unknowns as random variables and using observations to update that modeling effort.

Many procedures used in the analysis of geophysical data are Bayesian or approximately Bayesian. For example, the Kalman filter is a Bayesian procedure (e.g. Meinhold and
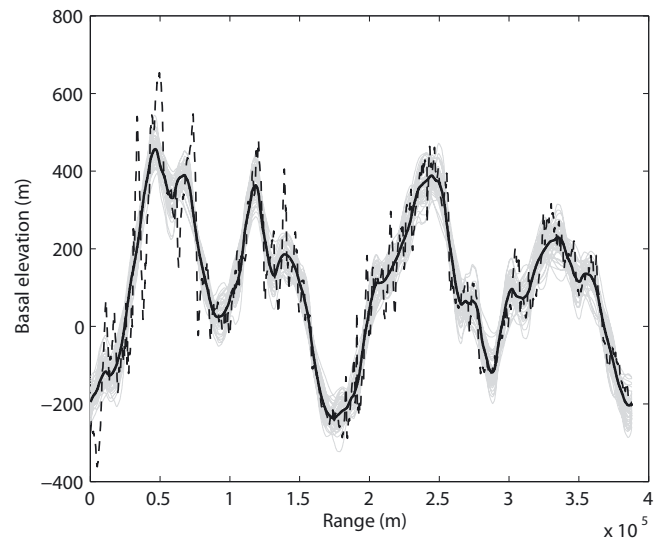
Singpurwalla, 1983). Other examples associated with data assimilation are discussed by Lorenc (1986) and Evensen and Van Leeuwen (2000). A general review of Bayesian analysis intended for geophysical audiences is given by Wikle and Berliner (2006); see also Epstein (1985) and Tarantola (1987).

Nevertheless, the full power of the methodology has only recently been making progress in geophysics (e.g. Berliner and others, 2000; Wikle and others, 2003; Tebaldi and others, 2005). Advances in computation, through variants of Markov chain Monte Carlo (MCMC) algorithms, now enable *hierarchical* Bayesian modeling that is capable of dealing with the complexities in models and data that arise in geophysics. A skeleton of hierarchical reasoning begins with consideration of three basic collections of variables to be modeled: data (our observations) labeled y; process (those physical, state variables of interest (e.g. velocities, stresses, etc.)) labeled x; and parameters labeled θ, including unknown physical constants and parameters introduced in the statistical components of the model. We use the following notation for probability distributions: (1) the distribution of a random variable, say X, is written as $[x]$; and (2) the conditional distribution of Y given $X = x$ is written as $[y|x]$. A review of the basic strategy and associated computations is given by Gelman and others (2004).

Hierarchical thinking suggests a model with three primary components (e.g. Berliner, 1996):

1. Data model: $[y|x, \theta]$;

2. Prior process model: $[x|\theta]$; and

3. Prior on parameters: $[\theta]$.

Although the construction of these components is often not easy, once the modeling process is complete, Bayes' theorem yields the *posterior distribution* $[x, \theta|y]$ given by

$$[x, \theta|y] = \frac{[y|x, \theta][x|\theta][\theta]}{[y]}. \tag{8}$$

The denominator $[y]$ is the marginal distribution of the data. It can be viewed simply as that constant which normalizes the

numerator ensuring that the posterior is indeed a probability distribution, depending upon the observed data but not the unknowns.

In the Bayesian treatment of parameters, even unknown constants are treated as if they are random. This is important since the definitions of some model parameters may be based on approximations. For example, flow parameters are not really unknown physical constants but rather functions of other unmodeled process variables (e.g. temperature or pressure) which introduce uncertainty. Hence, while the Bayesian approach may appear complicated compared to others, its advantage is that it quantifies and manages uncertainties through to final conclusions.

2. PHYSICAL–STATISTICAL MODELING OF THE 'NORTHEAST GREENLAND ICE STREAM'

Recall our three datasets: **S** (surface observations), **B** (basal observations), and **U** (surface velocity). The corresponding processes of interest are true surface topography $s(x)$, true basal topography $b(x)$ and true surface velocities $u(x)$, where x indexes a transect down the middle of the ice stream. There are no observations on the stresses acting on the ice, although physical relations allow inference from modeled stresses.

2.1. Translating physics into statistical models

In response to the issues introduced in section 1.2 regarding averaging of surface and thickness data for the estimation of driving stress, our first step is to model the basal topography b and the surface s as smoothed versions plus local noise, denoted generically by N. We therefore assume

$$b(x) = \psi_b(x) + N_b(x) \ \text{ and } \ s(x) = \psi_s(x) + N_s(x), \quad (9)$$

where $\psi_b(x)$ and $\psi_s(x)$ are smooth versions of the base and surface, respectively.

Applying these definitions in Equation (2), we model smoothed driving stress as

$$\tilde{\tau}_d(x) = -\rho g \tilde{H}(x) \frac{d\psi_s(x)}{dx}, \quad (10)$$

where the smoothed ice thickness is $\tilde{H}(x) = \psi_s(x) - \psi_b(x)$. In all computations, we set $\rho = 911 \, \text{kg m}^{-3}$ and g $= 9.81 \, \text{m s}^{-2}$.

We next define the partially smoothed basal shear stress in the region upstream of the currently undetermined change point, by applying Equation (5) for velocities taken in the negative x direction as

$$\tilde{\tau}_b(x) = \tilde{\tau}_d(x) + \eta(x). \quad (11)$$

We say $\tilde{\tau}_b(x)$ is only partially smoothed because we do not smooth the corrector process. As clarified later, $\eta(x)$ is modeled as an unknown stochastic process. This enables further interpretation of the posterior behavior of $\eta(x)$ as an indication of regions where successful or unsuccessful smoothing has occurred.

We then incorporate the velocity model Equation (4) (recall that we set $n = 3$) and the approximate sliding velocity model Equation (7) into a statistical model:

$$u(x) = k\frac{d\psi_s(x)}{dx} + 0.50 A_1 \tilde{H}(x) \left(\tilde{\tau}_d(x)\right)^3 + N_u(x)$$
$$\text{if } x < c \quad (12)$$
$$= u_b + 0.50 A_2 \tilde{H}(x)(\tilde{\tau}_b(x))^3 + N_u(x)$$
$$\text{if } x \geq c, \quad (13)$$

where k, A_1, u_b, A_2 and c are treated as unknown, random parameters and $N_u(x)$ is an error process.

Before proceeding with the specifics, we emphasize that these formulae motivate probability models that are updated in light of the observations. More directly, we do not produce single estimates of smoothed processes $\psi_s(x)$ and $\psi_b(x)$ based on the data and substitute those estimates into the stress and velocity models. Rather, we produce an ensemble of these quantities simulated from their posterior distributions and propagate them forward into our models. This produces an ensemble of stresses and velocities, which in turn are updated based on the velocity observations.

2.2. Underlying probability theory

Our main task is the development of the following probability distributions:

$$\begin{aligned} \text{Data model:} \quad & [\mathbf{B}, \mathbf{S}, \mathbf{U}|b, s, u, \eta, \boldsymbol{\theta}] \\ \text{Prior process model:} \quad & [b, s, u, \eta, \psi_b, \psi_s|\boldsymbol{\theta}] \\ \text{Prior on parameters:} \quad & [\boldsymbol{\theta}], \end{aligned}$$

where $\boldsymbol{\theta}$ denotes the collection of all model parameters. Our goal is to obtain the posterior distribution $[\psi_b, \psi_s, u, \eta, \boldsymbol{\theta}|\mathbf{B}, \mathbf{S}, \mathbf{U}]$, which can then be used to obtain the posterior distribution of stresses.

Our main assumption regarding the data model is that the three primary datasets are conditionally independent given the true processes they represent and given the model parameters. Notationally, the data model takes the form

$$[\mathbf{B}, \mathbf{S}, \mathbf{U}|b, s, u, \eta, \boldsymbol{\theta}] = [\mathbf{B}|b, \boldsymbol{\theta}_B][\mathbf{S}|s, \boldsymbol{\theta}_S][\mathbf{U}|u, \boldsymbol{\theta}_U], \quad (14)$$

where notation such as $\boldsymbol{\theta}_B$ is used to indicate those parameters (subsets of $\boldsymbol{\theta}$) explicitly appearing in the indicated models. A possible objection to Equation (14) arises since the basal data **B** are actually computed as the difference between surface and thickness observations. The assumed conditional independence therefore may not hold. We checked our posterior results for indications of degrees of departure from this assumption and found none that would affect our results.

Our modeling of the processes begins with a probabilistic equality (i.e. this is not an assumption, but a fact):

$$\begin{aligned} &[b, s, u, \eta, \psi_b, \psi_s|\boldsymbol{\theta}] \\ &\quad = [u, \eta|b, s, \psi_b, \psi_s, \boldsymbol{\theta}][b, s, \psi_b, \psi_s|\boldsymbol{\theta}]. \end{aligned} \quad (15)$$

Similarly, we have

$$\begin{aligned} &[u, \eta|b, s, \psi_b, \psi_s, \boldsymbol{\theta}] \\ &\quad = [u|\eta, b, s, \psi_b, \psi_s, \boldsymbol{\theta}][\eta|b, s, \psi_b, \psi_s, \boldsymbol{\theta}]. \end{aligned} \quad (16)$$

We make the following two assumptions.

1. In formulating the second term on the right side of Equation (15), we assume that the base b and the surface s are independent, *conditional* on their smoothed versions and the model parameters. We then obtain

$$[b, s, \psi_b, \psi_s|\boldsymbol{\theta}] = [b|\psi_b, \boldsymbol{\theta}_b][s|\psi_s, \boldsymbol{\theta}_s], \quad (17)$$

where again notation such as $\boldsymbol{\theta}_b$ is used to indicate appropriate subsets of $\boldsymbol{\theta}$. It is critical to note here that we are not assuming that the base and the surface are independent. Our modeling of both the base and surface is conditional upon smooth processes ψ_b and ψ_s. Our assumption is that the small-scale departures from those large-scale processes are independent.

2. Consider the conditional model for u, namely the first term on the right side of Equation (16). We assume that the velocity profile depends on the base and surface only through their respective smoothed versions ψ_b and ψ_s, i.e.

$$[u|\eta, b, s, \psi_b, \psi_s, \boldsymbol{\theta}] = [u|\eta, \psi_b, \psi_s, \boldsymbol{\theta}]. \qquad (18)$$

The specifications of prior distributions for the parameters are given in Appendix A.

3. COMPONENTS OF THE HIERARCHICAL MODEL

3.1. Basal model

First, the process model is described. We chose to use wavelets to model the smoothed basal topography $\psi_b(x)$ because of their flexibility in representing highly variable processes and because we can easily control their smoothness.

Wavelets work best for equally spaced data where the number of data points is an integer power of 2. Hence, we partition the domain of the data into $2^{11} = 2048$ bins of equal length (189.5 m). Let $\bar{\mathbf{b}}$ denote the 2048-dimensional vector constructed by averaging b within each bin. Note that $\bar{\mathbf{b}}$ is not observed.

We used multi-resolution Daubechies wavelets (e.g. Bruce and Gao, 1996; Vidakovic, 1999). Two collections of wavelets were used. The first set is thought of as the smooth signal, while the second represents the detail signal. We consider a smooth signal of four wavelets. To this component we added a detail signal containing 28 more wavelets. All 32 wavelets are mutually orthogonal.

After converting to a discrete wavelet form, we obtain a linear model for $[\bar{\mathbf{b}}|\psi_b, \boldsymbol{\theta}_b]$:

$$\bar{\mathbf{b}}|\psi_b, \boldsymbol{\theta}_b \sim N(\mathbf{WC}, \sigma_b^2 \Sigma(\phi_1, \phi_2)), \qquad (19)$$

where $\mathbf{W}$ is the 2048×32 matrix of discretized wavelet basis functions, $\mathbf{C}$ is the 32-dimensional vector of wavelet coefficients and $\Sigma(\phi_1, \phi_2)$ is the correlation matrix of an autoregressive process of order 2 (AR(2)) with variance σ_b^2. Notationally, $\psi_b = \mathbf{WC}$ and $\boldsymbol{\theta}_b = (\sigma_b^2, \phi_1, \phi_2)$. The selection of an AR(2) process to account for spatial dependence among the model errors (i.e. local variations in basal topography) was based on preliminary data analysis and practicality; our Bayesian computations require repeated inversion of a 2048×2048 matrix involving the inverse of $\Sigma(\phi_1, \phi_2)$, which is straightforward for an AR(2) process.

We performed analyses for four resolutions, corresponding to 8, 16, 32 and 64 coefficients in Equation (19). We found that the model with $k = 32$ coefficients provided the best results in terms of maintaining fidelity to the thickness data (i.e. not over-smoothing), yet providing good inferences for velocity and therefore stress.

Turning to the data model, we define the basal data vector $\bar{\mathbf{B}}$ of length 2048 with the ith element given by the simple arithmetic average of those basal observations lying in bin i. Let n_i ($i = 1, \ldots, 2048$) be the number of data points lying in bin i. In our case, none of the n_i were equal to zero; most were either 1 or 2.

Our data model (i.e. $[\mathbf{B}|b, \boldsymbol{\theta}_B]$ in Equation (14)) is

$$\bar{\mathbf{B}}|\bar{\mathbf{b}} \sim N(\bar{\mathbf{b}}, \sigma_B^2 \, \text{diag}\{n_i^{-1}\}), \qquad (20)$$

where σ_B^2 represents measurement-error variance and $\text{diag}\{n_i^{-1}\}$ is a 2048×2048 matrix with diagonal elements equal to n_i^{-1} and off-diagonal elements equal to 0 (i.e. the elements of $\bar{\mathbf{B}}$ are assumed to be conditionally independent).

This suggestion arises from a familiar result from elementary statistics. The sample mean of n independent observations with common mean and common variance σ^2 has variance σ^2/n.

Specifications of priors on parameters are given in Appendix A.

3.2. Surface model

Our modeling strategy for $[s|\psi_s, \boldsymbol{\theta}_s]$ in Equation (17) separates the large-scale and small-scale behaviors of the surface. Recalling Equation (9), we suppose

$$s(x) = \psi_s(x) + N_s(x), \qquad (21)$$

where the large-scale surface is given by a function ψ_s, assumed known up to a low-dimensional set of parameters, and N_s is a zero-mean spatial stochastic process described in Appendix A. To model ψ_s, we rely on an analysis given in Paterson (1994, p. 243, equation 8) and assume the basic model for the surface:

$$\psi_s(x) = \mu + K \left[L^{1+n^{-1}} - (L - x)^{1+n^{-1}} \right]^{0.50n/(n+1)}, \qquad (22)$$

where $\boldsymbol{\theta}_s = (\mu, K, L)$ are treated as unknown parameters. Although we set $n = 3$ here, we could also model n as an unknown.

We use only the large-scale surface Equation (22) to compute ice thickness, the surface derivative and hence the stress in Equation (10). Nevertheless, the presence of N_s is important in determining the data model in Equation (14). Under the strategy which uses Equation (22) to obtain the stress, we need a data model $[\mathbf{S}|\psi_s, \boldsymbol{\theta}_s, \boldsymbol{\theta}_S]$, where $\boldsymbol{\theta}_S$ includes measurement-error variances and parameters of the distribution of N_s in Equation (21); our specification is given by Equation (A3) in Appendix A.

3.3. Velocity model

For the data model $[\mathbf{U}|u, \boldsymbol{\theta}_U]$, we assume that, conditional on the true velocities, the elements of $\mathbf{U}$ have independent Gaussian distributions with means equal to the true velocities at the corresponding locations and common variances σ_m^2. That is, at each observation location x, we assume that

$$U(x) = u(x) + m(x), \qquad (23)$$

where $m(x)$ is a zero-mean, Gaussian measurement error with variance σ_m^2. Substituting for $u(x)$ as prescribed in Equations (12) and (13), we have

$$U(x) = k\frac{\mathrm{d}\psi_s(x)}{\mathrm{d}x} + 0.50A_1\tilde{H}(x)(\tilde{\tau}_d(x))^3 + N_u(x) + m(x)$$
$$\text{if } x < c \qquad (24)$$
$$= u_b + 0.50A_2\tilde{H}(x)(\tilde{\tau}_b(x))^3 + N_u(x) + m(x)$$
$$\text{if } x \geq c. \qquad (25)$$

We next combine the model errors N_u and traditional measurement errors m into a single error

$$e_U(x) = N_u(x) + m(x), \qquad (26)$$

assuming that $e_U(x)$ are independent, zero-mean, Gaussian random variables with common variance σ_U^2. (Note that these steps mean that there is no need for further use of the symbol u.) This simplification can be relaxed, but would require additional modeling of the process N_u.
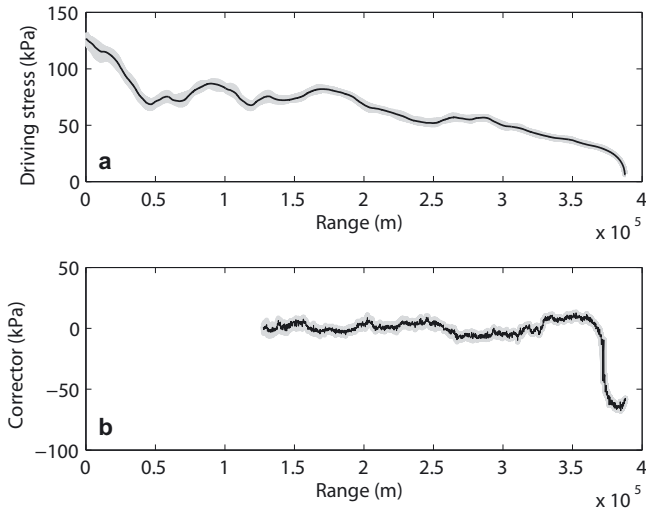
Fig. 4. Fifty posterior realizations of (a) smoothed driving stress (grey) and its posterior mean (black) and (b) the corrector process (grey) and its posterior mean (black), each based on 2000 realizations.

To simplify the presentation of the statistical model, we write our models in vector notation. Recalling Equations (9–11) and (19), we define the vector of smoothed ice thicknesses

$$\tilde{\mathbf{H}} = \boldsymbol{\psi}_s - \mathbf{WC}, \tag{27}$$

where $\boldsymbol{\psi}_s$ is the 2048-dimensional vector of smoothed surface elevations. Similarly, we define the vector of smoothed values of driving stress τ_d as

$$\tilde{\boldsymbol{\tau}}_d = \rho g \tilde{\mathbf{H}} * \mathbf{vec}\left(\frac{\mathrm{d}\psi_s}{\mathrm{d}x}\right), \tag{28}$$

where $*$ indicates the Hadamard product, i.e. we compute the vector of element-wise products of the coordinates of the smoothed thickness and the derivative of the smoothed surface.

Next, we let $\boldsymbol{\eta}$ denote the vector of stress corrections at the locations of velocity observation and set

$$\boldsymbol{\tau}_b = \tilde{\boldsymbol{\tau}}_d + \boldsymbol{\eta}. \tag{29}$$

From Equation (12) we model $\mathbf{u}$, the vector of true velocities at the observation locations, as a linear function of the corresponding coordinates of smoothed thickness multiplied by the third power of coordinates of $\boldsymbol{\tau}_b$.

In preliminary data analyses, we noted that at least two models (one for small x and another for large x) are needed. Let $x = c$ be an unknown change point and consider different linear functions above and below the change point. Finally, the model for the velocity data vector $\mathbf{U}$ is

$$\mathbf{U} = \left(\begin{array}{c} k \, \mathbf{vec}\left(\frac{\mathrm{d}\psi_s}{\mathrm{d}x}\right) \\ u_b \mathbf{1}_2 \end{array}\right) + \left(\begin{array}{c} 0.50 A_1 (\mathbf{H} * \boldsymbol{\tau}_d^3)_1 \\ 0.50 A_2 (\mathbf{H} * \boldsymbol{\tau}_b^3)_2 \end{array}\right) + \mathbf{e}_U, \tag{30}$$

where subscripts 1 and 2 indicate the varying dimensions of the vectors $\mathbf{H} * \boldsymbol{\tau}_d^3$, $\mathbf{H} * \boldsymbol{\tau}_b^3$ and $\mathbf{1}_2$ (a vector with all elements equal to 1) depending on the value of the change point c and $\mathbf{e}_U$ which are measurement and model errors, respectively (recall Equation (26)).

Table 1. Prior and posterior results for model parameters

	Error variance	
	Prior mean	Posterior mean
	m^2	m^2
Basal measurement	50	3.6748
Basal process	2000	4115
Velocity measurement	9.000	3.850

3.4. Stress corrector process

Since the corrector process arises from unmodeled and unobserved effects (e.g. longitudinal and lateral drag, basal lubrication, local under- or over-smoothing, etc.), we assume a relatively uninformed prior for $\boldsymbol{\eta}$. Specifically, our prior $[\boldsymbol{\eta}]$ indicates that each element has a uniform distribution on the interval (–100 kPa, 100 kPa) and that all elements are mutually independent.

3.5. Bayesian calculations

Although we can write down Bayes' theorem for the posterior distribution of all unknowns conditional on the observations, the result is typically not computable in closed form. We use a Monte Carlo approach that produces an ensemble of realizations from the target posterior distribution.

The method relies on the emerging technology of MCMC. The idea of MCMC is to simulate a Markov chain that has been carefully designed so that its stationary distribution coincides with the target posterior distribution. It follows that, after a burn-in or transience period, the generated realizations of the chain comprise a simulated sample from the posterior. Data analysis (often known as 'output analysis') is performed on this sample to produce the desired inferences.

In our case, direct use of MCMC is possible but challenging, primarily due to the non-linearities present in Equations (3) and (12). Hence, we combine MCMC with the technique of importance-sampling Monte Carlo (ISMC). Consider a setting in which direct simulation from a target distribution is difficult or inefficient. In ISMC, one generates an ensemble from another more manageable distribution. The theory of ISMC provides formulae for the calculation of *weights* that are used to re-weight the ensemble, permitting inferences relative to the original target. General introductions to both MCMC and ISMC can be found in Robert and Casella (1999). An illustration of these technologies in a geophysical problem is given in Berliner and others (2003).

An outline of the calculations used here is as follows. We first run separate, independent MCMC algorithms (simple versions known as Gibbs samplers) for the basal model and the surface model. These runs produce ensembles from the posterior distributions $[\psi_b, \boldsymbol{\theta}_b, \boldsymbol{\theta}_B | \bar{\mathbf{B}}]$ and $[\psi_s, \boldsymbol{\theta}_s, \boldsymbol{\theta}_S | \mathbf{S}]$. They are then used in conjunction with the velocity model $[u | \eta, \psi_b, \psi_s, \boldsymbol{\theta}_b, \boldsymbol{\theta}_s, \boldsymbol{\theta}_u]$ (recall Equation (18)) to simulate velocities conditional on $\bar{\mathbf{B}}$ and $\mathbf{S}$. (As described in Appendix A, we actually use a portion of $\mathbf{S}$.)

To incorporate the velocity data $\mathbf{U}$, we re-weight all of these samples using ISMC results, yielding the ensemble used to summarize the full posterior distribution. Technical details regarding the actual implementation are presented in Appendix B.

Table 2. Prior and posterior results for model parameters, before and after change point

Parameter	Prior mean	Posterior mean	Posterior SD
Before change point			
k (m s^{-1})	7×10^{-4}	5.497×10^{-4}	6.50×10^{4}
A_1 (s^{-1} kPa^{-3})	5×10^{-16}	1.257×10^{-16}	1.03×10^{-17}
After change point			
u_b (m a^{-1})	35	46.07	0.34
A_2 (s^{-1} kPa^{-3})	10^{-16}	1.43×10^{-15}	0.033×10^{-15}

Table 3. Posterior credible intervals (95%) for velocity model parameters

Parameter	Interval
Before change point	
k (m s^{-1})	$(5.37, 5.62) \times 10^{-4}$
A_1 (s^{-1} kPa^{-3})	$(1.05, 1.48) \times 10^{-16}$
After change point	
u_b (m a^{-1})	$(45.39, 46.72)$
A_2 (s^{-1} kPa^{-3})	$(1.33, 1.45) \times 10^{-15}$

4. POSTERIOR RESULTS

Figure 3 presents 50 realizations of the smoothed base $\psi_b = \mathbf{WC}$ (recall Equation (19)) from the posterior distribution conditional on the basal elevation data. The corresponding posterior mean basal topography, estimated using an ensemble of size 2000 and the original data, is also depicted. We see that the posterior distribution of the smoothed base is relatively faithful to the basal data, i.e. the wavelets are able to capture much of the variation of the basal data with 32 coefficients.

Figure 4a presents 50 realizations and the posterior mean, conditional on the basal and surface-elevation datasets, of the smoothed driving stresses ψ_d (recall Equation (28)). Figure 4b presents 50 realizations and the posterior mean of the correctors η, conditional on all three datasets (recall that we only modeled the corrector process to the right of the change point). In both cases, the posterior means were estimated using ensembles of size 2000. We note that except for the upper end of the data extent, the correctors are relatively small compared to the values of the driving stress. However, we also note that there is spatial structure indicated in the posterior for the correctors, although none was imposed by the prior.

For example, we see a run of positive correctors (perhaps due to lubrication) over the approximate range 100–120 km and a run of negative values (perhaps due to additional resistive stresses beyond those explained by the model) from ∼260 km to 330 km. The large negative correctors at the top end of the range are probably the combined result of a poor surface model in that region and a failure of the approximate physics so far upstream.

Figure 1 presents 50 realizations from the posterior distribution of velocity and the posterior mean velocity, estimated using ensembles of size 2000; the original velocity data are also shown. As explained in Appendix A, our estimate of the change point is $x = 127.3$ km; this point is also indicated on the plot. Interestingly, this location corresponds roughly to that of an apparent lineament.

As mentioned in section 3.5 and developed in Appendix B, ensemble members shown in Figures 1 and 4b are to be weighted according to importance-sampling theory. Although not indicated in these figures, the correct weights were used in producing posterior mean estimates.

The estimates (i.e. posterior means) for the other parameters in the model of the large-scale surface elevation (Equation (22)) are $\hat{\mu} = -450.53$, $\hat{K} = 4.75$ and $\hat{L} = 444\,901$. Posterior results for other key model parameters are presented in Tables 1–3.

5. CONCLUSIONS

Tables 1 and 2 list various prior and posterior means and posterior standard deviations of selected model parameters. Table 3 provides 95% posterior credible intervals (Bayesian analogues of confidence intervals) for the approximate Weertman coefficient, sliding velocity and flow parameters of the velocity model.

We note that while our prior estimate of sliding velocity upstream of the change point was fairly small (35 m a^{-1}), the posterior value is substantially larger. Next, we find that the posterior results for the flow parameters A on either side of the change point are plausible, although the results upstream from the change point may suggest warmer conditions than expected. Our treatment of A is of course linked to the specification of $\rho = 911$ kg m^{-3}. It is very easy to determine the results of other choices since it enters the model as a multiplicative constant.

We proposed smoothed surface and basal elevation models and showed that their use, in combination with simple physical models coupled with a corrector process, resulted in good predictions of velocity over much of the study region. There are regions to the right of the change point where the fitted corrector process indicates interesting interpretations regarding unmodeled dynamic controls, including lubrication at the base and additional resistive stresses. To the left of the change point, sliding modeled via a Weertman-like approximation with $p = q = 1$ dominates our model. Indeed, even if we removed the stress component of the model in the region, we would obtain very similar results.

REFERENCES

Arthern, R.J. and R.C.A. Hindmarsh. 2003. Optimal estimation of changes in the mass of ice sheets. *J. Geophys. Res.*, **108**(F1), 6007. (10.1029/2003JF000021.)

Berger, J.O. 1985. *Statistical decision theory and Bayesian analysis.* New York, Springer-Verlag.

Berliner, L.M. 1996. Hierarchical Bayesian time series models. *In* Hanson, K. and R. Silver, *eds. Maximum entropy and Bayesian methods.* Dordrecht, etc., Kluwer Academic Publishers, 15–22.

Berliner, L.M. 2003. Physical–statistical modelling in geophysics. *J. Geophys. Res.,* **108**(D24), 8776. (10.1029/2002JD002865.)

Berliner, L.M. and C.K. Wikle. 2006. Approximate importance sampling Monte Carlo for data assimilation. *Physica D,* **230**(1–2), 37–49.

Berliner, L.M., C.K. Wikle and N. Cressie. 2000. Long-lead prediction of Pacific SSTs via Bayesian dynamic modelling. *J. Climate,* **13**(13), 3953–3968.

Berliner, L.M., R.F. Milliff and C.K. Wikle. 2003. Bayesian hierarchical modelling of air–sea interaction. *J. Geophys. Res.,* **108**(C4), 3104. (10.1029/2002JC001413.)

Blatter, H. 1995. Velocity and stress fields in grounded glaciers: a simple algorithm for including deviatoric stress gradients. *J. Glaciol.,* **41**(138), 333–344.

Bruce, A. and H.Y. Gao. 1996. *Applied wavelet analysis with S-PLUS.* New York, Springer-Verlag.

DeGroot, M.H. 1970. *Optimal statistical decisions.* New York, McGraw-Hill.

Epstein, E.S. 1985. *Statistical inference and prediction in climatology: a Bayesian approach.* Boston, MA, American Meteorological Society.

Evensen, G. and P.J. van Leeuwen. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.,* **128**(6), 1852–1867.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin. 2004. *Bayesian data analysis. Second edition.* Boca Raton, FL, Chapman and Hall/CRC.

Goldstein, R.M., H. Engelhardt, B. Kamb and R.M. Frolich. 1993. Satellite radar interferometry for monitoring ice sheet motion: application to an Antarctic ice stream. *Science,* **262**(5139), 1525–1530.

Gudmundsson, G.H. 2006. Case study: estimating basal properties of glaciers from surface measurements. *In* Knight, P.G., *ed. Glacier science and environmental change.* Oxford, Blackwell, 415–417.

Joughin, I., M. Fahnestock, D. MacAyeal, J.L. Bamber and P. Gogineni. 2001. Observation and analysis of ice flow in the largest Greenland ice stream. *J. Geophys. Res.,* **106**(D24), 34,021–34,034.

Kamb, B. and K.A. Echelmeyer. 1986. Stress-gradient coupling in glacier flow: I. Longitudinal averaging of the influence of ice thickness and surface slope. *J. Glaciol.,* **32**(111), 267–284.

Lorenc, A.C. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.,* **112**(474), 1177–1194.

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

MacAyeal, D.R. 1993. A tutorial on the use of control methods in ice-sheet modeling. *J. Glaciol.,* **39**(131), 91–98.

Meinhold, J. and N. Singpurwalla. 1983. Understanding the Kalman filter. *Am. Stat.,* **37**(2), 123–127.

Paterson, W.S.B. 1994. *The physics of glaciers. Third edition.* Oxford, etc., Elsevier.

Robert, C.P. and G. Casella. 1999. *Monte Carlo statistical methods.* New York, Springer-Verlag.

Tarantola, A. 1987. *Inverse problem theory: methods for data fitting and model parameter estimation.* New York, Elsevier.

Tebaldi, C., R.L. Smith, D. Nychka and L.O. Mearns. 2005. Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. *J. Climate,* **18**(10), 1524–1540.

Van der Veen, C.J. 1999. *Fundamentals of glacier dynamics.* Rotterdam, A.A. Balkema.

Vidakovic, B. 1999. *Statistical modeling by wavelets.* New York, etc., Wiley.

Whillans, I.M. 1987. Force budget of ice sheets. *In* Van der Veen, C.J. and J. Oerlemans, *eds. Dynamics of the West Antarctic ice sheet.* Dordrecht, etc., D. Reidel Publishing Co., 17–36.

Wikle, C.K. and L.M. Berliner. 2006. A Bayesian tutorial for data assimilation. *Physica D,* **230**(1–2), 1–16.

Wikle, C.K., L.M. Berliner and R.F. Milliff. 2003. Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. *Mon. Weather Rev.,* **131**(6), 1051–1062.

APPENDIX A

DETAILS OF THE HIERARCHICAL MODEL

Basal model: parameter priors

Our prior for the measurement-error variance σ_B^2 is that it has an inverse gamma distribution with mean 50 and standard deviation 15. We also assumed an inverse gamma prior distribution for σ_b^2 with mean 2000 and standard deviation 200.

While we could treat ϕ_1 and ϕ_2 as unknown parameters, they are only of minor interest here and small variations in their values do not affect our results. Hence, we considered ordinary least-squares fits to the data and used the resulting residuals to estimate ϕ_1 and ϕ_2, resulting in

$$\phi_1 = 1.603 \quad \text{and} \quad \phi_2 = -0.638.$$

Next, we describe our prior for the wavelet coefficients. Regarding the four coefficients $\mathbf{C}_s$ for the smooth signal,

$$\mathbf{C}_s \sim N(\boldsymbol{\mu}, \sigma_{sc}^2 \mathbf{I}_4), \qquad (A1)$$

where $\boldsymbol{\mu}$ is the vector of conventional, ordinary least-squares estimates of Haar-wavelet coefficients. Regarding the coefficients for the detail signal, we assume their prior means to be zero and that they are independent.

The model is completed by forming priors for the variances of the wavelet coefficients. We assumed these variances to be independent and to have inverse gamma distributions. For σ_{sc}^2, we used prior mean 200^2 and standard deviation 40. For the variance σ_c^2 of the detail signals, we assumed prior mean 1000^2 and prior standard deviation 100.

Surface modeling

Recall that we assume $s(x) = \psi_s(x) + N_s(x)$ (Equation (21)), where the large-scale parameterized surface is given by ψ_s as defined in Equation (22). Assume N_s is a zero-mean, Gaussian spatial stochastic process with constant variance $\sigma_{N_s}^2$ and stationary correlation function $\gamma(x, x') = \gamma(|x - x'|)$.

To produce a data model readily usable to isolate the large-scale surface, we first assumed that the individual observations were conditionally independent with means equal to the true surface values. Next, we formally eliminated the small-scale process N_s from the model via a probabilistic method. The result then leads to observations whose means are equal to the large-scale model but whose values are correlated. Informally, they inherit the correlation structure of N_s.

Rather than dealing with the overhead of modeling that correlation structure, we make a simple assumption. We assume that for all locations separated by at least 150 m, the correlation is approximately zero. (We arrived at this value by fitting the surface by conventional least-squares analysis and then inspected the correlation function of the residuals of the fitted surface. Residuals at locations separated by more than 150 m were essentially uncorrelated.)

We then took a subsample of the surface data such that all observations were at least 150 m apart, which resulted in a subsample of 600 observations. We should have very little loss in efficiency in basing our analysis of the surface on such

a subsample: 600 is a very large sample size for estimating three parameters (μ, K and L). Further, we tried a total of six such subsamples and obtained essentially the same results. Of course, the analysis hinges on the approximation that the decorrelation length scale of small-scale variations is 150 m.

Data model
We assume that, conditional on the true surface and the measurement-error variance σ_S^2,

$$\mathbf{S}|\mathbf{s}, \sigma_S^2 \sim N(\mathbf{s}, \sigma_S^2\, \mathbf{I}), \tag{A2}$$

where $\mathbf{s}$ is the vector of true surface values at the observation locations. Standard results from probability theory imply that if we integrate out N_s in Equation (21), the resulting data model is

$$\mathbf{S}|\boldsymbol{\theta}_s, \sigma_S^2, \sigma_{N_s}^2, \boldsymbol{\Gamma} \sim N(\boldsymbol{\psi}_s, \sigma_S^2\, \mathbf{I} + \sigma_{N_s}^2\, \boldsymbol{\Gamma}), \tag{A3}$$

where $\boldsymbol{\psi}_s$ is the vector of the large-scale parametric surface model evaluated at the observation locations and $\boldsymbol{\Gamma}$ is the matrix of correlations of N_s at pairs of those locations, as implied by the correlation function γ.

Process model
A Bayesian or a non-Bayesian analysis (e.g. generalized least squares) based on Equation (A3) requires modeling and estimation of $\boldsymbol{\Gamma}$ (in parallel with the estimation of $\boldsymbol{\Sigma}(\phi_1, \phi_2)$ in Equation (19)). Assuming a decorrelation length of 150 m, a subsample $\mathbf{S}_l$ such that all observations are at least 150 m apart has distribution

$$\mathbf{S}_l|\boldsymbol{\psi}_s, \sigma_S^2, \sigma_{N_s}^2 \sim N(\boldsymbol{\psi}_{s,l}, \sigma_{S,l}^2\, \mathbf{I}_l), \tag{A4}$$

where $\sigma_{S,l}^2 = \sigma_S^2 + \sigma_{N_s}^2$ and $\boldsymbol{\psi}_{s,l}$ is the vector of large-scale surface values at the subsample locations.

Parameter model
Under the assumptions and simplifications outlined above, we need only consider a prior for the parameters μ, K, L and $\sigma_{S,l}^2$. Probability theory greatly simplifies our task. Having reduced the original problem to that of only 4 parameters and 600 conditionally independent observations, we can approximate the desired posterior. For all but the most highly concentrated priors, the posterior is approximately Gaussian, with means given by the conventional least-squares estimates and covariance matrix given by the estimated inverse information matrix for our setting (Berger, 1985, section 4.9).

Velocity modeling
Data model
We assume

$$\mathbf{U}|\mathbf{u}, \sigma_U^2 \sim N(\mathbf{u}, \sigma_U^2\mathbf{I}), \tag{A5}$$

where $\mathbf{u}$ is the vector of true velocities at the observation locations and σ_U^2 is the measurement-error variance.

Process model
Based on the corresponding smoothed versions of thickness Equation (27) and stress Equation (28), we consider the change-point models

$$\mathbf{u} = \begin{pmatrix} k\, \mathbf{vec}\left(\frac{\mathrm{d}\boldsymbol{\psi}_s}{\mathrm{d}x}\right) \\ u_b\, \mathbf{1}_2 \end{pmatrix} + \begin{pmatrix} 0.50A_1\, (\mathbf{H} * \boldsymbol{\tau}^3)_1 \\ 0.50A_2\, (\mathbf{H} * \boldsymbol{\tau}^3)_2 \end{pmatrix} + \mathbf{e}_U. \tag{A6}$$

This leads to the data models

$$\mathbf{U} = \mathbf{u} + \mathbf{e}_U \tag{A7}$$

as given in Equation (30).

Parameter model
We next describe our prior distribution for the unknown parameters $k, u_b, A_1, A_2, \sigma_U^2$ and c.

Our prior for k, u_b, A_1, A_2 and σ_U^2 is the following so-called *conjugate multivariate normal-inverse gamma* distribution. The prior is described hierarchically as

$$[k, u_b, A_1, A_2, \sigma_U^2] = [k, u_b, A_1, A_2|\sigma_U^2][\sigma_U^2], \tag{A8}$$

where

$$\begin{pmatrix} k \\ 0.50A_1 \\ u_b \\ 0.50A_2 \end{pmatrix} \Big| \sigma_U^2 \sim N(\boldsymbol{\mu}, \sigma_U^2\mathbf{W}) \tag{A9}$$

and

$$\sigma_U^2 \sim IG(a, b), \tag{A10}$$

where all off-diagonal elements of $\mathbf{W}$ are zero and

$$\boldsymbol{\mu} = (7 \times 10^{-4}, 5 \times 10^{-16}, 35, 10^{-16})^T \tag{A11}$$
$$\mathbf{W} = \mathrm{diag}(10^{-6}, 10^{-32}, 14^2, 0.50 \times 10^{-30}) \tag{A12}$$
$$(a, b) = (11, 90). \tag{A13}$$

Under these specifications, the prior means of the regression coefficients are given in Equation (A11), the prior variances of the regression coefficients are

$$9 \times (10^{-6}, 2 \times 10^{-32}, 14^2, 10^{-30})^T$$

and the prior mean and variance of σ_U^2 are both equal to 9 (Goldstein and others, 1983).

We remark that the prior above may seem cumbersome and strange in that we model the regression-model parameters conditioned on the error variance σ_U^2. However, this strategy leads to a computable posterior distribution for the regression parameters and σ_U^2. Specifically,

$$[k, u_b, A_1, A_2, \sigma_U^2|\mathbf{U}] = [k, u_b, A_1, A_2|\sigma_U^2, \mathbf{U}][\sigma_U^2|\mathbf{U}], \tag{A14}$$

where the first distribution is the usual multivariate normal posterior holding σ_U^2 fixed, and $[\sigma_U^2|\mathbf{U}]$ is an inverse gamma distribution with updated parameters. Hence, it was a popular prior model before the advent of MCMC. A complete discussion and explanation is given by DeGroot (1970, ch. 9); see also Berger (1985, p. 288). We use the prior in this case because it readily adapts to our use of ISMC.

To closely mimic an uninformative analysis regarding the change point, we performed a least-squares piecewise linear regression with an unknown change point. We found very strong agreement on the value of c and, hence, treated c as known and equal to 127.3 km.

APPENDIX B
MCMC–ISMC APPROACH

Our computational approach combines MCMC and ISMC. The following derivations are motivated by a need to make the ISMC steps efficient. To achieve efficiency, we seek to generate ensembles of unknown quantities that are reasonably similar to ensembles generated directly from the posterior distribution. To do so, we apply Bayes' theorem to portions of the hierarchical model in an attempt to allow

the observational data to strongly influence the simulations. The details of these derivations involve interplay of our modeling assumptions and repeated use of two basic facts from probability theory. First, for the random quantities A and B, we have $[A|B][B] = [B|A][A]$ (Bayes' theorem). Second, such expressions hold conditionally. That is, if C is also some random quantity, we have $[A|B, C][B|C] = [B|A, C][A|C]$.

We may write the full posterior distribution as

$$[\text{unknowns} \mid \mathbf{U}, \bar{\mathbf{B}}, \mathbf{S}_l]$$
$$\propto [\mathbf{U}|\boldsymbol{\theta}_u, \sigma_U^2, \boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\boldsymbol{\theta}_u, \sigma_U^2][\boldsymbol{\eta}]$$
$$\cdot [\boldsymbol{\psi}_b, \boldsymbol{\psi}_s, \boldsymbol{\theta}_b, \boldsymbol{\theta}_B, \boldsymbol{\theta}_s, \boldsymbol{\theta}_S|\bar{\mathbf{B}}, \mathbf{S}_l]. \qquad (B1)$$

Note that we have incorporated the basal and surface data-sets into a partial posterior distribution $[\boldsymbol{\psi}_b, \boldsymbol{\psi}_s, \boldsymbol{\theta}_b, \boldsymbol{\theta}_B, \boldsymbol{\theta}_s, \boldsymbol{\theta}_S|\bar{\mathbf{B}}, \mathbf{S}_l]$ for the basal and surface processes and parameters. This model is amenable to simple MCMC. Further, as indicated in Figure 2, simulated smoothed basal topographies from this partial posterior are clearly responsive to the observations. Next, we rewrite the first two terms on the right side of Equation (B1) as

$$[\mathbf{U} \mid \boldsymbol{\theta}_u, \sigma_U^2, \boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\boldsymbol{\theta}_u, \sigma_U^2]$$
$$= [\boldsymbol{\theta}_u, \sigma_U^2|\mathbf{U}, \boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\mathbf{U}|\boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s], \qquad (B2)$$

and then note that

$$[\mathbf{U}|\boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\boldsymbol{\eta}] = [\boldsymbol{\eta}|\mathbf{U}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\mathbf{U}|\boldsymbol{\psi}_b, \boldsymbol{\psi}_s]. \qquad (B3)$$

In summary, we have

$$[\text{unknowns} \mid \mathbf{U}, \bar{\mathbf{B}}, \mathbf{S}_l]$$
$$\propto [\boldsymbol{\theta}_u, \sigma_U^2|\mathbf{U}, \boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\boldsymbol{\eta}|\mathbf{U}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s][\mathbf{U}|\boldsymbol{\psi}_b, \boldsymbol{\psi}_s]$$
$$\cdot [\boldsymbol{\psi}_b, \boldsymbol{\psi}_s, \boldsymbol{\theta}_b, \boldsymbol{\theta}_B, \boldsymbol{\theta}_s, \boldsymbol{\theta}_S|\bar{\mathbf{B}}, \mathbf{S}_l]. \qquad (B4)$$

Our basic algorithm is as follows.

1. Via MCMC, produce an ensemble of size M from the partial posterior distribution $[\boldsymbol{\psi}_b, \boldsymbol{\psi}_s, \boldsymbol{\theta}_b, \boldsymbol{\theta}_B, \boldsymbol{\theta}_s, \boldsymbol{\theta}_S|\bar{\mathbf{B}}, \mathbf{S}_l]$ (recall Equation (B1)). The result includes an ensemble

$$\{(\boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m) : m = 1, \dots, M\}$$

of smoothed basal and surface elevations.

2. For each $(\boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m)$, generate $\boldsymbol{\eta}^m$ from $[\boldsymbol{\eta}|\mathbf{U}, \boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m]$.

3. Similarly, generate $(\boldsymbol{\theta}_u^m, \sigma_U^{2,m})$ from

$$[\boldsymbol{\theta}_u, \sigma_U^2|\mathbf{U}, \boldsymbol{\eta}^m, \boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m].$$

Finally, define ISMC weights:

$$\alpha^m \propto [\mathbf{U}|\boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m]. \qquad (B5)$$

Coupling these weights with the ensemble members yields a weighted ensemble from the full posterior distribution.

We remark that the selection of the prior described in Appendix A for $\boldsymbol{\theta}_u$ and σ_U^2 enables the exact calculation of the partial posterior distribution needed in step 3 (Equation (A14)). However, there are two difficulties with

implementation in our example. First, $[\mathbf{U}|\boldsymbol{\psi}_b, \boldsymbol{\psi}_s]$ is not easily obtained, disabling our ability to compute the weights in Equation (B5). To deal with this problem, we apply the general reasoning suggested in Berliner and Wikle (2006) to claim that the earlier steps in the analysis, particularly step 3, have sufficiently accounted for the information in the velocity data regarding the unknowns. Therefore, a good approximate analysis arises if we ignore these weights.

Similarly, step 2 is problematic in that the required distribution $[\boldsymbol{\eta}|\mathbf{U}, \boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m]$ is not obtainable in closed form. Recall that our prior for the elements of $\boldsymbol{\eta}$ is that they are independent, uniform random variables on the interval $(-100, 100)$. Hence,

$$[\boldsymbol{\eta}|\mathbf{U}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s] \propto [\mathbf{U}|\boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s] \qquad (B6)$$

for vectors $\boldsymbol{\eta}$ lying in the 2880-dimensional cube with sides given by $(-100, 100)$. We can obtain the exact form of $[\mathbf{U}|\boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s]$. Specifically, it is a multivariate t density (Berger, 1985, p. 561), i.e.

$$\mathbf{U}|\boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s \sim T(2a, \mathbf{Z}\boldsymbol{\mu}, (a/b)^{-1}(\mathbf{I} + \mathbf{ZWZ}')), \qquad (B7)$$

where $\boldsymbol{\mu}$, $\mathbf{W}$ and (a, b) are defined in Equations (A11–A13) and $\mathbf{Z}=\mathbf{Z}(\boldsymbol{\eta}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s)$ is a 2880×4 matrix obtained by rewriting Equation (30) as

$$\mathbf{U} = \mathbf{Z} \begin{pmatrix} u_{b,1} \\ 0.50A_1 \\ u_{b,2} \\ 0.50A_2 \end{pmatrix} + \mathbf{e}_U. \qquad (B8)$$

Although we have Equation (B7), it is intractable viewed as a probability density function for $\boldsymbol{\eta}$. To deal with this problem, we appeal to importance sampling. Specifically, rather than sampling from $[\boldsymbol{\eta}|\mathbf{U}, \boldsymbol{\psi}_b, \boldsymbol{\psi}_s]$, we sample from another distribution, $[\boldsymbol{\eta}]_{\text{IS}}$, chosen to mimic it. Specifically, we used a multivariate t distribution, with mean specified by inverting the regression model and degrees of freedom parameter equal to 10. By inverting the regression model, we mean the following. For each observation of velocity $U(x)$ at a location x after the change point, and each ensemble member, our regression model (Equation (30)) is

$$U(x) = u_b + 0.50A_2 \, \tilde{H}(x)[\tilde{\tau}_d(x) + \eta(x)]^3 + e_U(x), \qquad (B9)$$

where we have suppressed notation indicating dependence on the ensemble member. We then simply approximate e_U by 0, yielding the approximation

$$\eta(x) \approx \left(\frac{U(x) - u_b}{0.50A_2 \, \tilde{H}(x)} \right)^{1/3} - \tilde{\tau}_d(x). \qquad (B10)$$

The right side of this expression is then used to specify the mean of $[\boldsymbol{\eta}]_{\text{IS}}$. This may appear to be cheating, but it is perfectly legal. One may construct importance-sampling distributions in any required fashion, including using the data, as long as ISMC weights are used. These weights are readily approximated by

$$\tilde{\alpha}^m \propto [\boldsymbol{\eta}^m|\mathbf{U}, \boldsymbol{\psi}_b^m, \boldsymbol{\psi}_s^m]/[\boldsymbol{\eta}^m]_{\text{IS}}. \qquad (B11)$$