2005

Hierarchical Bayesian modeling of the movement of ice streams

L.M. Berliner  
*Ohio State University*

Noel A. Cressie  
*Ohio State University, ncressie@uow.edu.au*

Kenneth Jezek  
*Ohio University*

Yongku Kim  
*Ohio University*

Calvin Lam  
*Ohio State University*

**Publication Details**
Hierarchical Bayesian modeling of the movement of ice streams

Abstract
Modern studies of the behaviors 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. Though these datasets represent significant advances in terms of spatial coverage and the range of processes we can observe, the physical systems to be modeled are nevertheless incompletely observed. Uncertainties associated with measurement errors are also present. In addition, physical models are also subject to uncertainties. Hence, there is a need for combining observations and models in a fashion that incorporates uncertainty and quantifies its impact on conclusions.

Keywords
bayesian, hierarchical, modeling, ice, movement, streams

Disciplines
Engineering | Science and Technology Studies

Publication Details

This conference paper is available at Research Online: http://ro.uow.edu.au/eispapers/5973
Hierarchical Bayesian modeling of the movement of ice streams

L.M. Berliner¹, N. Cressie¹, K. Jezek², Y. Kim¹, and C.Q. Lam¹

¹ Department of Statistics, The Ohio State University, Columbus, OH 43210
² Byrd Polar Research Center, The Ohio State University, Columbus, OH 43210

1 Introduction

Modern studies of the behaviors 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. Though these datasets represent significant advances in terms of spatial coverage and the range of processes we can observe, the physical systems to be modeled are nevertheless incompletely observed. Uncertainties associated with measurement errors are also present. In addition, physical models are also subject to uncertainties. Hence, there is a need for combining observations and models in a fashion that incorporates uncertainty and quantifies its impact on conclusions.

In this research, we give a fully Bayesian approach that incorporates physical models and data from the Northeast Ice Stream in Greenland. Specifically, we develop statistically enhanced versions of a simple model of driving stress and a familiar model for velocity based on stress, in combination with recently acquired ice-thickness, surface-elevation, and surface-velocity datasets.

The primary output of a Bayesian analysis is a posterior distribution, namely the joint probability distribution for the unknown quantities conditional on the observed data. Even in our simple illustration, we have on the order of 8,000 unknowns, so explicit presentation of the full joint distribution is not feasible. Hence, a key aspect of Bayesian analysis in such high-dimensional settings is the ability to generate realizations or ensembles from the posterior, which is then summarized by statistical analyses of such ensembles.

1.1 Glaciological Motivations

Since glaciers flow under the force of gravity, important factors in determining velocities include quantities such as the ice thickness, in combination
with forces acting along the sides and at the base of the glacier and under the constraints of the constitutive relationship. A simple approximation equates driving stress along the flow (x-direction) to basal shear stress as follows:

$$\tau_{dx} = \tau_{b_x} = -\rho g H \frac{dh}{dx},$$

(1)

where $h$ is ice-surface elevation, $H$ is the ice thickness, $\rho$ is the density of ice, and $g$ is the gravity constant.

Under these assumptions, it is reasonably straightforward to estimate driving stress based on observations of $h$ and $H$. However, a concern in estimating driving stress from geometry data is that the reliance on the slope of the upper ice surface in (1) implies that results are very sensitive to small-scale variations in surface topography. Although there is no theoretical requirement that driving stress be spatially averaged, it is usually calculated over horizontal distances of a few ice thicknesses or so, to eliminate small-scale flow features that are not important to the large-scale flow.

1.2 Hierarchical Bayesian Approach

Many procedures used in the analysis of geophysical data are Bayesian or approximately Bayesian. For example, the Kalman filter is a Bayesian procedure. General presentations of Bayesian analysis intended for geophysical audiences are given in Epstein (1985) and Tarantola (1987). Nevertheless, the full power of the methodology is only recently making inroads in geophysics. Modern modeling approaches, especially hierarchical Bayesian modeling, have become computationally feasible, thus enabling new and richer approaches to analyses. The essence of hierarchical modeling is the formulation of a complex probability model for many variables as the product of a collection of potentially simpler component models. Consider three basic collections of variables to be modeled: data, our observations; process, those physical, state variables of interest (e.g., velocities, stresses, etc.); and parameters, unknown physical constants and parameters introduced in the statistical components of the model. Hierarchical thinking suggests a model with three primary components (e.g., Berliner 1996),

Data model: [data][process, parameters],
Prior process model: [process][parameters],
Prior on parameters: [parameters].

Though the construction of these components is not easy, once the modeling process is complete, Bayes’ Theorem yields the posterior distribution, which is [prior parameters][data].

1.3 Data

We assembled surface-topography and ice-thickness observations for a portion of the Northeast Ice Stream in Greenland. The data were gathered as part of the Program for Arctic Climate Regional Assessments (PARCA).
Surface topography and basal topography were sampled every few hundred meters using equipment mounted on the Wallops Flight Facility P-3 aircraft (basal observations are shown as part of Figure 1). Surface velocity data were calculated by Ian Joughin and provided as part of the PARCA dataset (velocity observations are shown as part of Figure 2).

2 Physical-Statistical Modeling of the NE Ice Stream

Our three datasets are: S, surface observations; B, basal observations; and U, velocity data. The corresponding processes of interest are true surface \( h(x) \), true basal topography \( b(x) \), and true velocities \( u(x) \).

2.1 Physics

We incorporate three physically based models. First, following the discussion leading to (1), we consider the stress,

\[ \tau = \rho g H \frac{dh}{dx}, \]

where the thickness is \( H = h - b \), \( \rho \) is the density of the ice, and recall that \( g \) is the gravity constant. The negative sign present in (1) is omitted here because we model \( \tau \) and velocity in the negative-\( x \) direction. In all computations, we set \( \rho = 911 \ km/m^3 \) and \( g = 9.81 \ m/s^2 \).

Second, under a laminar flow assumption and treating the flow parameter \( A \) as a constant, the surface velocity \( u \) is given by,

\[ u = u_b + \frac{2A}{n+1} H \tau^n, \]

where \( u_b \) is the sliding velocity and \( n \) is a flow parameter (e.g., Paterson, 1994, p. 251, eq. 21).

Finally, as suggested by the analysis given in Paterson (1994, p. 243, eq. 8), we consider the following basic model for the surface:

\[ h = k (L^{1+n-1} - (L - x)^{1+n-1})^{0.50n/(n+1)}. \]

In addition to making explicit use of these three models in our statistical development, we also formulate our models with the intention of addressing issues of smoothing. Specifically, as discussed earlier, there are questions of the degrees to which \( \tau \) and the surface should be smoothed to obtain physically meaningful controls on surface velocity.
2.2 Bayesian Modeling

Recalling the recipe described earlier, our main tasks are the development of the following probability distributions:

Data Model: \[ \mathbf{B}, \mathbf{S}, \mathbf{U} \mid b, h, u, \theta \] = [\mathbf{B} \mid b, \theta_B][\mathbf{S} \mid h, \theta_S][\mathbf{U} \mid u, \theta_U] \quad (5) \\
Process Model: [b, h, u \mid \theta] = [b \mid \theta_b][h \mid \theta_h] \times [u \mid b, h, \theta_b, \theta_h, \theta_u], \quad (6) \\
\text{where } \theta \text{ denotes the collection of all model parameters, and notation such as } \theta_B \text{ is used to indicate those parameters explicitly appearing in the indicated models on the right-hand sides of (5), (6), and (7). The specifications of those quantities along with the prior } \theta \text{ will be described later.}

To assess the degree of smoothing needed to understand velocities modeled in (6) and (7), we assume that the velocity field depends on the base and surface only through their respective smoothed versions \( \theta_b \) and \( \theta_h \). Formally, we assume that

\[ [u \mid b, h, \theta_b, \theta_h, \theta_u] = [u \mid \theta_b, \theta_h, \theta_u]. \quad (8) \]

Once we get this far in the analysis, we assume that the relationship is deterministic (this assumption could be relaxed). That is, the probability distribution on the right-hand side of (8) is degenerate. Alternatively, \( u \) is a function of \( \theta_b \) and \( \theta_h \), albeit with unknown parameters \( \theta_u \). Our choice for that function is based on (3).

2.3 Components of the Hierarchical Model

We now describe the key components (the six models on the right-hand sides of (5), (6), and (7)) of our hierarchical model.

Bed Model We use a Daubechies wavelet-based analysis (e.g., Bruce and Gao, 1996; Vidakovic, 2003). We partition the domain of the data into \( 2^{11} = 2,048 \) bins of equal length (189.5 m).

Let \( \mathbf{b} \) denote the 2048-dimensional vector constructed by averaging \( b \) within each bin. Define the basal data vector \( \mathbf{B} \) of length 2048 with \( i \)-th element given by the simple arithmetic average of those observations lying in bin \( i \). Our data model assumes that the elements of \( \mathbf{B} \) are conditionally independent, each being normally distributed with mean equal to the corresponding element of \( \mathbf{b} \) and variance determined by the measurement-error variability of an individual observation, denoted by \( \sigma^2_\mathbf{b} \), and the number of observations in the corresponding bin.

We form a so-called \textit{multiresolution wavelet decomposition}, which allows us to view the degree of smoothness to be determined by a parameter \( r \). We present results for each of four resolutions, labeled 1, 2, 3, and 4.
After converting to discrete wavelet forms, each model is simply a linear regression model. We write the four models in compact form: For each $r = 1, 2, 3, 4$, we assume a wavelet model,

$$\mathbf{h} \sim N(\mathbf{W}_r \mathbf{C}_r, \sigma^2_r \Sigma(\phi_{1r}, \phi_{2r})),$$

(9)

where $\mathbf{W}_r$ is the $2048 \times k(r)$ matrix of discretized wavelet basis functions ($k(1) = 8, k(2) = 16, k(3) = 32, k(4) = 64$); $\mathbf{C}_r$ is the $k(r) \times 1$ vector of wavelet coefficients; and $\Sigma(\phi_{1r}, \phi_{2r})$ is the correlation matrix of an autoregressive process of order two ($\text{AR}(2)$) with variance $\sigma^2_r$. The selection of an $\text{AR}(2)$ error model to account for spatial dependence among these model errors (i.e., local variations in basal topography) was based on preliminary data analysis and computational practicality.

Of course, we have a mixture model, but as will be clarified in Section 4, we do not need standard mixture model updates in this article.

**Surface Model** Our modeling strategy separates the large- and small-scale behaviors of the surface. We suppose

$$h(x) = h^{\theta_h}(x) + \mathcal{H}(x),$$

(10)

where the large-scale surface is given by a parameterized function $h^{\theta_h}$, assumed known up to a low-dimensional parameter $\theta_h$, and $\mathcal{H}$ is a zero-mean spatial stochastic process. To model $h^{\theta_h}$, we rely on (4) and assume that

$$h^{\theta_h}(x) = \mu + K(L^{1+n^{-1}} - (L - x)^{1+n^{-1}})^{0.5n/(n+1)},$$

(11)

where $\mu$, $K$, and $L$ are treated as unknown parameters. In the analysis here, we set $n = 3$, though we could model $n$ as an unknown as well.

In principle, we could treat $\mathcal{H}$ in a fashion analogous to the treatment of the basal topography. Namely, we anticipate that a smooth representation of $\mathcal{H}$ could be used in developing stresses to explain velocities. As a simplification, we tried ignoring the role of $\mathcal{H}$ in determining stress; specifically, we used only the large-scale component $h^{\theta_h}(x)$ in computing both thickness and the derivative needed for application of (2). The results appear surprisingly reasonable, though enhancements could be explored. Specifically, we assume that for all locations separated by at least 150$m$, the correlation is approximately zero. We then take a subsample of the surface data such that all observations are at least 150$m$ apart. We used 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 ($\mu$, $K$ and $L$). Further, we tried a total of six such subsamples and obtained essentially the same results.
Velocity Model Our data model is again a basic measurement-error model, namely, we assume that conditional on the true velocities, the data vector \( \mathbf{U} \) has a Gaussian distribution with mean equal to the vector of velocities at the corresponding locations, common variance \( \sigma^2_U \), and are independent.

Turning to process modeling, recall that we are to formulate \( [ u | \theta_b, \theta_h, \theta_v ] \) (see (8)) based on (2) and (3). For the fitted surface function and each resolution \( r = 1, 2, 3, 4 \) used in developing the basal models, we consider (i) the corresponding smoothed versions of thickness, defined as

\[
H_r = h^h - W_r C_r, \tag{12}
\]

and (ii) similarly defined smoothed values of \( \tau_r \), defined as

\[
\tau_r = (H_r \cdot \frac{dh^h}{dx}), \tag{13}
\]

where the right-hand side means each coordinate of \( \tau_r \) is obtained as the elementwise product of the corresponding coordinates of the smoothed thickness and the derivative of the surface.

Based on (2) and (3), we model the vector of true velocities at the observation locations, \( \mathbf{u} \), as a linear function of the corresponding coordinates of \( H_r \) times the \( n^{th} \) powers of coordinates of \( \tau_r \). It appears that at least two models (one for small \( x \) and another for large \( x \)) are needed. Hence, we used a statistical analysis based on change-points. We define an unknown change-point, \( c \), and consider different linear regression models on each side of \( c \), as follows:

\[
\mathbf{U} = \left( \begin{array}{c}
u_{0,1} 1_x \\ u_{0,2} 1_x \end{array} \right) + \left( \begin{array}{c}0.50A_1 (H_r \cdot \tau_r^c) \\ 0.50A_2 (H_r \cdot \tau_r^{\hat{c}}) \end{array} \right) + \mathbf{e}_U, \tag{14}
\]

where the subscripts \( c \) and \( \hat{c} \) indicate the varying dimensions of the indicated vectors, depending on the value of the change point \( c \), and \( \mathbf{e}_U \) are errors primarily representing measurement error associated with the velocity data.

3 Bayesian Calculations

We use a Monte Carlo approach that essentially produces an ensemble of realizations from the target posterior. The method relies on Markov Chain Monte Carlo (MCMC). In our case, direct use of MCMC is quite challenging, primarily due to the nonlinearities present in (2) and (3). Hence, we combine MCMC with the technique of Importance Sampling Monte Carlo (ISMC).

An outline of the calculations used here is as follows. We first run separate, independent MCMC algorithms for the each of the four basal models.
and for the surface model. These runs produce ensembles from the posterior distributions \([h, \theta_b, \theta_2, r] \mid B\) and \([\theta_i, S]\). Due to the various conditional-independence assumptions described above, these ensembles are summaries of the posterior distribution of the unknowns conditional on the two datasets \(B\) and \(S\). They can be used in conjunction with the velocity model \([u \mid \theta_b, \theta_1, \theta_2, r]\) to simulate velocities conditional on \(B\) and \(S\) for each value of \(r\). To incorporate the velocity data \(U\), we reweight all of these samples using ISMC theory.

4 Results

Recall (11),

\[
h^{\theta_i}(x) = \mu + k (L^{1+n^{-1}} - (L-x)^{1+n^{-1}})^{0.50n/(n+1)}.
\]

With \(n = 3\), our posterior estimates (i.e., means) for the parameters are \(\mu = -450.53\), \(k = 4.75\), and \(L = 444901\).

For each of the four resolutions, Figure 1 presents 10 realizations of the smoothed base \(W, C\), and the original data. However, due to the size of the basal dataset, that support is overwhelming: The posterior probability \(p(r = 4 \mid B, S)\) is equal to 1 for any plausible choice of prior probabilities \(p(r)\), and hence the other three resolutions have posterior probabilities equal zero. This means that further model-selection analysis incorporating the velocity data must still result in a final posterior probability \(p(r = 4 \mid B, S, U) = 1\). Nevertheless, we shall assess the role of basal smoothing in understanding stress and velocity.

Figure 2 presents plots of the original velocity data, 100 realizations from the posterior distribution of velocity, and the posterior mean based on ensembles of size 2000 for each of the four resolutions. In Figure 3, we plot 50 realizations and the posterior mean, estimated using ensembles of size 2000, of the smoothed \(\tau_r\) (recall (13)) for each of the four resolutions. We can isolate the effect of resolution in predicting velocity by estimating \(\sigma^2_r\) based on each resolution. Specifically, for each of our 2000 simulated ensemble members at a given \(r\), we compute the variance, say \(v^2_{r,m}\), where \(m\) indicates the ensemble member) of the “residuals”, namely the observed velocity data minus the generated velocity model. The average of these then provides a posterior estimate of \(\sigma^2_r\) (again, due to the very large sample sizes, the prior distributions on \(\sigma^2_r\) “wash out”) for the given \(r\).

Figure 4 presents the results. For each \(r\), we have plotted kernel density estimates based on \(\{v^2_{r,m}\}\). For each \(r\), three curves are shown. In each case, the middle curve is based on all data (both sides of the change point). The other two curves result from restricting the computations to each side of the change point. Note that all four resolutions do comparatively poorly to the left of the change point. Also note that we estimated \(\sigma^2_r\) to be about 50 for each resolution. This gives a standard deviation of about 7-8 \(m/yr\), which
Bayesian modeling of ice streams

compares well to measurement errors in velocity data that are expected by
glaciologists to be on the order of 10 m/yr.

5 Conclusions

As one would expect, the highest resolution wavelet model does the best
job of summarizing the basal data. Due to the very large sample size,
formal Bayesian model selection would strongly select the highest resolution
model. Nevertheless, we examined all four basal-model resolutions in terms
of their resulting qualities in modeling velocity based on a very simple
physical model. All four resolutions result in reasonable velocity models.
Indeed, even a very smooth base is surprisingly good at modeling velocity,
at least over much of the study region. Further, there are regions in which
all models are systematically weak, suggesting either breakdowns in the
simple models used or a need for spatially heterogeneous local models.
For example, all models do poorly as we approach the modeled change
point from either side. There appears to be a genuine breakdown in the
physical model here, probably corresponding to the presence of a lineament.
A second region of systematic model error, at a range of 250-390 km, can
be remedied by using a more refined surface model in that region.

Acknowledgments: This research was supported by the National Sci-
ence Foundation (Office of Polar Programs and Probability and Statistics
Program) under grant no. 0229292. We have benefitted considerably from
input from Kees van der Veen.

References

mum Entropy and Bayesian Methods, K. Hanson and R. Silver (Eds.),


Springer-Verlag, New York.

Epstein, E. S. (1985). Statistical Inference and Prediction in Climatology:
A Bayesian Approach. American Meteorological Society, Boston, MA.

Quarterly Journal of the Royal Meteorological Society, 112, 1177-
1194.
