BAYESIAN METHODS IN GLACIOLOGY

By

Douglas Brinkerhoff, B.S., M.S.

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Geophysics

University of Alaska Fairbanks

December 2017

APPROVED:

Martin Truffer, Committee Chair

Andy Aschwanden, Committee Chair

Carl Tape, Committee Member

Ed Bueler, Committee Member

Paul McCarthy, Chair

Department of Geosciences

Paul Layer, Dean

College of Natural Science and Mathematics

Michael Castellini, Dean

Graduate School

Abstract

The problem of inferring the value of unobservable model parameters given a set of observations is ubiquitous in glaciology, as are large measurement errors. Bayes' theorem provides a unified framework for addressing such problems in a rigorous and robust way through Monte Carlo sampling of posterior distributions, which provides not only the optimal solution for a given inverse problem, but also the uncertainty. We apply these methods to three glaciological problems. First, we use Markov Chain Monte Carlo sampling to infer the importance of different glacier hydrological processes from observations of terminus water flux and surface speed. We find that the opening of sub-glacial cavities due to sliding over asperities at the glacier bed is of a similar magnitude to the opening of channels due to turbulent melt during periods of large input flux, but also that the processes of turbulent melting is the greatest source of uncertainty in hydrological modelling. Storage of water in both englacial void spaces and exchange of water between the englacial and subglacial systems are both necessary to explain observations. We next use Markov Chain Monte Carlo sampling to determine distributed glacier thickness from dense observations of surface velocity and mass balance coupled with sparse direct observations of thickness. These three variables are related through the principle of mass conservation. We develop a new framework for modelling observational uncertainty, then apply the method to three test cases. We find a strong relationship between measurement uncertainty, measurement spacing, and the resulting uncertainty in thickness estimates. We also find that in order to minimize uncertainty, measurement spacing should be 1–2 times the characteristic length scale of variations in subglacial topography. Finally, we apply the method of particle filtering to compute robust estimates of ice surface velocity and uncertainty from oblique time-lapse photos for the rapidly retreating Columbia Glacier. The resulting velocity fields, when averaged over suitable time scales, agree well with velocity measurements derived from satellites. At higher temporal resolution, our results suggest that seasonal evolution of the subglacial drainage system is responsible for observed changes in ice velocity at seasonal scales, and that this changing configuration produces varying degrees of glacier flow sensitivity to changes in external water input.

Table of Contents

List of Figures

Page

List of Tables

Acknowledgements

Throughout the course of conceiving the research and writing this thesis, my co-advisors Martin Truffer and Andy Aschwanden have been inexplicably tolerant of my multitude of bad ideas and judiciously supportive of the occasional good one. In each case, they have been generous with their time and insight in matters variously intellectual, logistical, and philosophical. Getting to know both Martin and Andy has also been a lot of fun, from chasing Caribou on the north slope, to flying around the Alaska Range in a Cessna, to eating Raclette in a sleet-storm on a spring trip to Taku Glacier. Here's to many more projects together.

I am also grateful to the remainder of my committee. Ed Bueler has kept me mathematically honest and has been both patient and effective in teaching me the proper way to think about numerical modelling. Carl Tape has been a font of good ideas and his perspective from outside glaciology helped me to recognize when I'm talking in jargon and also in providing an additional literature from which to draw potential solutions to glaciological problems. Shad O'Neel was kind enough to host me for a year at the USGS Alaska Science Center for an NSF internship and became a *de facto* committee member, providing sound scientific feedback and invaluable ideas on how to convince people that they should actually care about this stuff.

While it might not be evident from the contents of this thesis, throughout my PhD I was lucky enough to get to go on a variety of field expeditions to locations in Alaska and abroad. While visiting glaciated places is always beautiful and exotic, each trip was defined by its companions. Throughout several visits to Taku Glacier in Southeast Alaska, I had the pleasure of working with a large and diverse group, but I especially want to thank Roman Motyka, Jason Amundson, Franklyn Dunbar, Jenna Zechmann, Aurora Roth, and Dale Pomraning (whose mechanical aptitude I can only dream of one day emulating). On a trip to Wolverine Glacier in the Kenai Mountains, Louis Sass, Chris McNeil, and Emily Baker of the USGS taught me about how to measure surface mass balance and the finer points of snow machine travel in crevassed terrain. I'll bring more beer next time, guys. Finally, I had the great fortune of moonlighting on a research cruise to the Antarctic Peninsula, where I learned a little bit about the ocean while joining Peter Winsor, Hank Statscewich, and Peter Shipton (among others) on a variety of misadventures, including a late night operation to recover a missing AUV by Zodiac, braving the stench of penguin effluent to repair a corroded seismometer, and yet another late night operation to drink all of the pisco in Punta Arenas.

Back in the office, thanks to all of my colleagues in the GI's glaciers group who have provided commentary and encouragement, and who have made the GI such a pleasant place to work. These

include Regine Hock, Mark Fahnestock, Christian Kienholz, Joanna Young, Mat Debolskiy, Jessica Zimmerman, and many others. I particularly want to thank my friend and colleague Christina Carr, who has been the detail-oriented foil to my decidedly big-picture nature. I would not be here without her, if only because I would not have filled out the correct forms. Christina has consistently provided valuable scientific and stylistic feedback, encouragement, and a floor to sleep on when necessary. She has also been an effective advocate for all of the GI's graduate students. I am in her debt.

My parents, Molly and Tom Brinkerhoff, have been a constant source of support since the day I was born, and I attribute any success that I've had to the values that they hold and were careful to try to teach me: intellectual honesty, an open mind, and the belief that I should always do my best, regardless of the task. While I know that I have rarely made their lives easy, I hope that the credit that I share with them for this thesis is at least a partial repayment for the efforts that they have always made on my behalf.

One adventure always leads to another; Molly Staats has been there for most of them, and for me, throughout my academic career. She moved to Fairbanks, Alaska sight-unseen and immediately embraced the Alaskan lifestyle, from dipnetting salmon and butchering caribou to skiing alpine lines above Valdez. She was also more patient with me than I deserved through late nights of frenzied coding and months away in the field. As we finish this stage of our lives, we move on to a new one: in a little over two weeks from when I write these acknowledgements, I will have the privilege of marrying her. It goes without saying that I am and always will be deeply grateful for her presence and her support.

"I have no satisfaction in formulas unless I feel their arithmetical magnitude."

—Lord Kelvin

"Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones."

—Donald Rumsfeld

Chapter 1

Introduction

Observations rarely provide certainty. As a rule, mistakes are made in the process of making observations, or the model (conceptual or otherwise) that the observations are intended to inform may be inadequate to explain them. The best that we can hope for then is that an observation will provide evidence for or against a particular hypothesis, updating our degree of belief in it. This is not necessarily a problem: absolute certainty is not a requirement in many problems of practical importance (Should I board this airplane? How much are humans contributing to climate change?). Indeed, such a requirement would be paralyzing. Yet, even in the absence of certainty, whatever method is used to evaluate belief in a particular hypothesis needs to be both rigorous and consistent, else we risk a descent into nihilism.

Bayes' theorem provides the means to quantitatively update belief in a hypothesis [*Tarantola*, 2005]. Suppose we have some model of the world which is governed by a number of parameters $\mathbf{m}$, and we wish to update our hypothesis of $\mathbf{m}$ based on some observations $\mathbf{d}$. We use the notation $P(\mathbf{m}|\mathbf{d})$ to represent a conditional probability density function, which integrates to unity and quantifies proportional belief that a given model state $\mathbf{m}$ is true given $\mathbf{d}$. A simple manipulation of an identity gives Bayes' Theorem

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}. \tag{1.1}$$

Each component has a distinct interpretation. The left hand side is known as the *posterior distribution*, which is the probability distribution of model states after our belief has been updated with observations. It is the state of knowledge about $\mathbf{m}$ after all information has been considered. The first term in the numerator is known as the *likelihood*, which is the likelihood of having measured $\mathbf{d}$, assuming a certain state $\mathbf{m}$ to be true. This also encodes the entire process of observation. Such a function can be very complicated when the effects of parameters $\mathbf{m}$ are filtered through a complex model before being observed. The denominator $P(\mathbf{d})$ is the marginal probability of the observations, which simply acts as normalizing constant. Finally, the second term in the numerator is known as the *prior probability*, or simply the prior, which is the probability distribution of model states before having considered observations. The prior can encode many presuppositions, such as positivity, smoothness, or regularity [*Neal*, 1997]. It may also enforce dynamical considerations, such as when the model at the next instant in time is random but depends strongly on the model at the current instant [*Kalman*, 1960]. Finally, the prior can be viewed recursively, such that when new information becomes available, the prior may be the posterior from having considered previous observation. Stated plainly, Bayes' theorem says that belief in a hypothesis about the model state updated with observations is proportional to the probability of having made those observations assuming the hypothesis to be true multiplied by *a priori* belief in the hypothesis. Either the likelihood or the prior can dominate the update of belief. If prior belief in a hypothesis is very strong, then no amount of evidence can change this belief. On the contrary, if little is already known about a hypothesis, then belief in it defers to observations even if they are error-prone.

Bayes' theorem provides a means to make inferences about arbitrary unobservable quantities given observable ones, which is to say that it is a mechanism for solving inverse problems. This makes it a powerful tool in geophysics and particularly in understanding glacier dynamics, where a primary concern is using quantities that can be observed at the relatively accessible glacier surface to make inferences about governing processes that are mostly occurring in unobservable locations such as the glacier interior or bed. Such problems are ubiquitous in glaciology, but two examples have been of particular interest in the literature. First is the problem of determining the thickness of a glacier given an estimate of its distributed specific mass balance, its surface geometry and rate of change, and perhaps its surface velocity. Often, pointwise measurements of thickness supplement these distributed estimates, but these are usually sparse. A variety of deterministic methods have been proposed to solve this problem, from artificial neural networks [*Clarke et al.*, 2009] to partial differential equation-constrained optimization [*Morlighem et al.*, 2011; *Huss and Farinotti*, 2012]. However, a recent intercomparison showed a large spread in the solutions produced by these methods [*Farinotti et al.*, 2017]. Unsurprisingly, the most consistent results were found for synthetic examples where the surface mass balance and surface velocity estimates could actually be error-free (an impossibility in real life). In general the sparser the measurements, the larger the discrepancy between individual methods. A second common problem is to estimate the traction exerted by a glacier's bed on the overlying ice, given a physical model of ice flow and estimates of ice geometry and surface velocity [*MacAyeal*, 1993]. While most models are able to reproduce the data with arbitrary precision, the resulting estimates of basal traction show extreme sensitivity to errors in prescribed ice thickness, choice of physical model, and the spatial smoothness imposed [*Morlighem et al.*, 2013; *Habermann et al.*, 2013; *Brinkerhoff and Johnson*, 2013; *Sergienko et al.*, 2014]. This non-uniqueness has led to uncertainty regarding whether such inversions are even useful, particularly towards the goal of prognostic glacier modelling at the continental scale.

A reasonable question to ask when confronted with multiple answers to the same problem is: which one of these is right? The only reasonable response is that none of them are. With some exceptions [*Petra et al.*, 2012; *Raymond and Gudmundsson*, 2009], glaciologists have been concerned with finding the most probable value of $\mathbf{m}$ given observations, because the problems are high dimensional and can be most efficiently solved with deterministic iterative methods. However, finding the most probable state only gives half the story, and without an estimate of the complete posterior distribution, we have no knowledge of important factors such as uncertainty, correlation, and local versus global optima. In particular, considering only the maximum a posteriori solution provides little information on the degree of uniqueness for a particular inverse model solution: is a given solution clearly the best, or are there many other options that are just as good? It is easy to see how the sensitivity of different methods for inferring ice thickness or basal traction could be the result of such non-uniqueness. Finding the complete posterior distribution solves each of these problems. Parameter uncertainty and covariance are a fundamental part of the posterior. Local and global maxima are evident because the entire distribution is computed. The posterior distribution gives all probable solutions, not just the most probable, which allows us to assess uniqueness. The advantages conferred by producing the complete posterior distribution come with a price, however: they are computationally expensive. Because of the curse of dimensionality, the effort required to even enumerate the potential states of a model increases exponentially in the number of parameters. This onerous requirement can be partially circumvented by carefully constructed sampling schemes, so-called Monte Carlo methods that produce finite samples of the complete posterior distribution that converge asymptotically to the true distribution as the number of samples grows large [*Brooks et al.*, 2011]. Inferences are thus made from these samples, rather than an analytical expression for the posterior distribution. Still, the number of samples necessary to adequately represent the underlying distribution may be large, and generating them most efficiently is an area of active research.

As an example, consider the problem of inferring the (statistical) state of a glacier's hydrologic system (i.e. englacial and subglacial storage and transmissivity) from observations of surface velocity. The hydrologic system, in conjunction with imperfect estimates of influx from the glacier surface and outflux through the glacier terminus, determines the water pressure at the glacier base, which in turn affects the amount of drag exerted by the bed. This traction resists the driving and membrane stresses induced by glacier geometry, all/owing a certain amount of slip between the two surfaces to occur, but the specific relationship between pressure, stress, and slip is imprecisely known. Basal velocity is translated up through the body of the ice, while being attenuated by temperature-dependent viscous flow. A GPS receiver installed in the ice surface is then advected along with the resulting complex superposition of basal velocity and internal deformation (with additional noise from small motions due to melting and wind), and while doing so records the position and transmission time of a multitude of satellites in orbit around the Earth. Thus the likelihood for this model is a function that tells us the probability of the GPS receiver reading a certain set of times from orbital satellites, were the subglacial drainage system organized in a hypothesized way. As a prior, we can specify that it is extremely improbable for water flux or basal shear stress to be inversely proportional to water pressure (among other things). At each step in this process errors accrue from imprecision in the measurement device, from random fluctuations, and from inadequacies in the physical models of the atmosphere, ice flow, and basal physics. Given all of this uncertainty, is it even possible to say something useful about a glacier's hydrology? If so, how sure are we about our results? More broadly, what limits exist to the amount of information that can be inferred about unobservable glacial processes from those that can be observed? The latter is the central question addressed by this dissertation.

In Chapter 2, we apply Monte Carlo sampling to the specific problem discussed in the previous paragraph, using Kennicott Glacier, AK as a test case. We borrow the observations and hydrologic model from *Bartholomaus et al.* [2011], who is a co-author on the paper, along with Colin Meyer, Ed Bueler, and Martin Truffer. The work appears in Annals of Glaciology as 'Inversion of a subglacial hydrology model.' The mathematical underpinnings were mostly conceived just steps from the glacier in question, at UAF's Glaciology Summer School. We first make the problem tractable by performing formal non-dimensionalization, which isolates the true model parameters that control the relative magnitude of a variety of hydrologic processes such as melting by turbulent dissipation and the characteristic small-scale topography beneath Kennicott Glacier, before using the Metropolis-Hastings algorithm to find their probability distributions. We find that the opening of linked cavities due to sliding over a rough bed and turbulent melting of cavity walls are of similar importance in determining the evolution of the drainage system and that significant water is stored both subglacially and in englacial voids. The primary source of uncertainty in the model results is driven by the system's sensitivity to melting from turbulent dissipation.

Chapter 3 appears in Frontiers in Earth Science as 'Bayesian inference of subglacial topography using mass conservation,' and is co-authored by Andy Aschwanden and Martin Truffer. We again use Monte Carlo sampling to find the approximate posterior distribution of an ice thickness field, given dense observations of surface velocity and specific mass balance and sparse, pointwise measurements of ice thickness. We link these quantities through mass conservation: the glacier is constrained such that the divergence of ice flux is equal to the surface mass balance, a requirement for the glacier to maintain a (quasi-)steady-state. For this situation (which is a common one in glaciology), we elucidate the various sources of uncertainty that must be considered, including that associated with non-contemporaneous measurements. We apply the method to three test cases: a synthetic glacier, a small Swedish mountain glacier, and Greenland's largest outlet. In each case we provide evidence for the convergence of the sampling procedure to the posterior distribution, and present the posterior distribution of ice thickness for each of the cases considered. We then explore the way in which uncertainty in both the velocity and surface mass balance propagate into uncertainty in the ice thickness. Finally, we examine the influence that the thickness correlation length (specified through the prior covariance) has on the resulting posterior distribution, and make a recommendation for the appropriate spacing between thickness measurements for the optimal trade-off between effort and return. Additionally, the method developed in this paper was applied to the ice thickness model intercomparison project [*Farinotti et al.*, 2017], which compared solutions of multiple methods for estimating ice thickness from surface observations.

In Chapter 4, we take a step back from the direct inference of bed properties, and instead focus on the observation of glacier surface velocities from oblique time-lapse images. This is difficult problem for several reasons. First, glaciers tend to show up in places with bad weather, and so occlusions are common. Second, many characteristic features of a glacier surface, such as crevasses, are quasi-periodic, and differentiating between them is error-prone. Third, the procedure of projecting from two-dimensional image coordinates to three-dimensional spatial coordinates is ill-posed. We circumvent these issues by using a particle filter, a Bayesian sampling technique that sequentially culls and replicates samples from a large number of potential solutions generated by a model of the underlying physics. We specify a likelihood function by matching characteristic glacier features between images while simultaneously correcting for small camera motions. Using this technique, we construct a daily-resolution estimate of glacier surface velocities for the terminus of Columbia Glacier, a rapidly retreating and highly dynamic glacier in Alaska's Chugach Mountains. We find significant seasonal changes in velocity in line with previous work, and also short-term fluctuations driven by periods of intense melt and rainfall. This work was performed in collaboration with Shad O'Neel of the USGS, and is in preparation for submission to Journal of Glaciology.

Finally, in Chapter 5, we take stock of the primary conclusions of this work, and discuss ways in which the techniques developed here can be used to further improve glaciological inverse modelling.

Bibliography

Bartholomaus, T. C., R. S. Anderson, and S. P. Anderson (2011), Growth and collapse of the distributed subglacial hydrologic system of Kennicott Glacier, Alaska, USA, and its effects on basal motion, *Journal of Glaciology*, *57*(206), 985–1002.

Brinkerhoff, D. J., and J. V. Johnson (2013), Data assimilation and prognostic whole ice sheet modelling with the variationally derived, higher order, open source, and fully parallel ice sheet model VarGlaS, *The Cryosphere*, *7*(4), 1161–1184, doi:10.5194/tc-7-1161-2013.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng (2011), *Handbook of Markov Chain Monte Carlo*, CRC press.

Clarke, G. K. C., E. Berthier, C. G. Schoof, and A. H. Jarosch (2009), Neural networks applied to estimating subglacial topography and glacier volume, *Journal of Climate*, *22*(8), 2146–2160, doi:10.1175/2008JCLI2572.1.

Farinotti, D., D. J. Brinkerhoff, G. K. C. Clarke, J. J. Fürst, H. Frey, P. Gantayat, F. Gillet-Chaulet, C. Girard, M. Huss, P. W. Leclercq, A. Linsbauer, H. Machguth, C. Martin, F. Maussion, M. Morlighem, C. Mosbeux, A. Pandit, A. Portmann, A. Rabatel, R. Ramsankaran, T. J. Reerink, O. Sanchez, P. A. Stentoft, S. Singh Kumari, W. J. J. van Pelt, B. Anderson, T. Benham, D. Binder, J. A. Dowdeswell, A. Fischer, K. Helfricht, S. Kutuzov, I. Lavrentiev, R. McNabb, G. H. Gudmundsson, H. Li, and L. M. Andreassen (2017), How accurate are estimates of glacier ice thickness? results from ITMIX, the Ice Thickness Models Intercomparison eXperiment, *The Cryosphere*, *11*(2), 949–970, doi:10.5194/tc-11-949-2017.

Habermann, M., M. Truffer, and D. Maxwell (2013), Changing basal conditions during the speed-up of Jakobshavn Isbræ, Greenland, *The Cryosphere*, *7*(6), 1679–1692, doi:10.5194/tc-7-1679-2013.

Huss, M., and D. Farinotti (2012), Distributed ice thickness and volume of all glaciers around the globe, *Journal of Geophysical Research: Earth Surface*, *117*(F4), doi:10.1029/2012JF002523, f04010.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, *82*(1), 35–45, doi:10.1115/1.3662552.

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

Morlighem, M., E. Rignot, H. Seroussi, E. Larour, H. Ben Dhia, and D. Aubry (2011), A mass conservation approach for mapping glacier ice thickness, *Geophysical Research Letters*, *38*(19), doi:10.1029/2011GL048659, l19503.

Morlighem, M., H. Seroussi, E. Larour, and E. Rignot (2013), Inversion of basal friction in Antarctica using exact and incomplete adjoints of a higher-order model, *Journal of Geophysical Research: Earth Surface*, *118*(3), 1746–1753, doi:10.1002/jgrf.20125.

Neal, R. M. (1997), Monte Carlo implementation of Gaussian process models for Bayesian regression and classification, *arXiv preprint physics/9701026*.

Petra, N., H. Zhu, G. Stadler, T. J. Hughes, and O. Ghattas (2012), An inexact Gauss–Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model, *Journal of Glaciology*, *58*(211), 889–903.

Raymond, M. J., and G. H. Gudmundsson (2009), Estimating basal properties of ice streams from surface measurements: a non-linear Bayesian inverse approach applied to synthetic data, *The Cryosphere*, *3*(2), 265–278, doi:10.5194/tc-3-265-2009.

Sergienko, O. V., T. T. Creyts, and R. C. A. Hindmarsh (2014), Similarity of organized patterns in driving and basal stresses of Antarctic and Greenland ice sheets beneath extensive areas of basal sliding, *Geophysical Research Letters*, *41*(11), 3925–3932, doi:10.1002/2014GL059976.

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, SIAM.

Chapter 2

Inversion of a glacier hydrology model[1]

Abstract

The subglacial hydrologic system exerts strong controls on the dynamics of the overlying ice, yet the parameters which govern the evolution of this system are not widely known or observable. To gain a better understanding of these parameters, we invert a spatially-averaged model of subglacial hydrology from observations of ice surface velocity and outlet stream discharge at Kennicott Glacier, Wrangell Mountains, AK. To identify independent parameters, we formally non-dimensionalize the forward model. After specifying suitable prior distributions, we use a Markov Chain Monte Carlo algorithm to sample from the distribution of parameter values conditioned on the available data. This procedure gives us not only the most probable parameter values, but also a rigorous estimate of their covariance structure. We find that the opening of cavities due to sliding over basal topography and turbulent melting are of a similar magnitude during periods of large input flux, though turbulent melting also exhibits the greatest uncertainty. We also find that both the storage of water in the englacial system and the exchange of water between englacial and subglacial systems are necessary in order to explain both surface velocity observations and the relative attenuation in the amplitude of diurnal signals between input and output flux observations.

2.1 Introduction

The sub- and englacial hydrologic systems strongly influence the dynamics of glaciers [e.g *Iken et al.*, 1983; *Jansson*, 1995; *Fischer and Clarke*, 1997]. Observations of this linkage span a continuum from small mountain glaciers [*Flowers et al.*, 2002; *Jansson*, 1995; *Harper et al.*, 2007], to much larger valley glaciers [*Bartholomaus et al.*, 2008; *Truffer and Harrison*, 2006], to the large polar Isbræ and ice streams of Greenland and Antarctica [*Zwally et al.*, 2002; *Engelhardt and Kamb*, 1998]. In particular, multiple studies have indicated that subglacial water pressure, particularly effective pressure (the difference between ice overburden and water pressure), is dominant in explaining glacier sliding velocities [*Iken and Bindschadler*, 1986; *Willis et al.*, 1995; *Iken and Truffer*, 1997]. Theoretical analysis of ice flow over bedrock asperities in the presence of liquid water also suggest a dependence between sliding velocity and effective pressure [*Lliboutry*, 1968; *Fowler*, 1986; *Schoof*, 2005; *Gagliardini et al.*, 2007]. This dependence is primarily due to the filling of subglacial cavities (and drowning of bedrock asperities) with pressurized water, leading to a relative decrease in

---

[1] Published as Brinkerhoff, D. J., C. M. Meyer, E. Bueler, M. Truffer, and T. Bartholomaus (2016), Inversion of a glacier hydrology model, *Journal of Glaciology*, 57 (72), 84–95.

frictional forces between the ice and bedrock.

Despite the available theoretical and observational conclusions, predictive models of sliding due to subglacial hydrology are difficult to formulate, and more difficult to validate. The specific relationship between effective pressure and sliding is not fully understood. Initial theoretical results postulated an inverse power law relationship [*Fowler*, 1986]. This type of sliding law has been shown to produce good correspondence with observations, often with model parameters being remarkably consistent between locations [*Jansson*, 1995; *Sugiyama and Gudmundsson*, 2004]. However, such a model is singular at vanishing effective pressures, a state which is observed rather frequently [e.g *Amundson et al.*, 2006; *Harper et al.*, 2007]. Additional theoretical work by *Schoof* [2005] and *Gagliardini et al.* [2007] suggests a phenomenological sliding law which satisfies 'Iken's bound' [*Iken*, 1981], which predicts a maximum finite shear stress in the limit of vanishing effective pressure. Nonetheless, identifying the appropriate parameters for such a law, which generally depend upon the specifics of small scale bed geometry, remains problematic.

Simultaneously, the estimation of effective pressure presents difficulties. Far from being static in time, it is widely believed that the configuration of the subglacial drainage system can evolve rapidly and on sub-seasonal scales, often changing its qualitative configuration from an inefficient network of linked cavities and perhaps sheet flow [*Kamb*, 1987; *Hubbard and Sharp*, 1995], to an efficient channelized system more analogous to fluvial systems [*Nienow et al.*, 1998; *Sugiyama and Gudmundsson*, 2004]. The combination of these interesting dynamics have inspired several models which have sought to capture the evolution of the subglacial hydrologic system. *Arnold et al.* [1998] presented a semi-distributed subglacial hydrology model which explicitly allowed the evolution of the channelized drainage system. *Flowers and Clarke* [2002a] developed a model which accounted for exchanges between the surface, englacial, subglacial, and groundwater systems to Trapridge Glacier [*Flowers and Clarke*, 2002b], successfully capturing seasonal evolution of subglacial routing efficiency, and also suggested the possible importance of englacial storage. *Schoof* [2010] more recently modelled a spontaneously-generated channelized subglacial system, and showed that the transition between the dominant modes of cavities and channels exhibits hysteresis. Importantly, he also showed that an increase in mean water supply does not necessarily translate into an increase in glacier flux and that the rate at which input flux changes is a more dominant control on glacier dynamics. More recently, *Schoof et al.* [2012] addressed the issue of bounds on water pressure, both above and below, by casting the problem as a variational inequality. *Werder et al.* [2013] extended the model of *Schoof* [2010] by introducing a novel unstructured edge-based discretization scheme, which allowed for an unbiased simulation of channel orientation.

Despite advances in hydrologic modelling, validation against observations of hydrologic variables such as water pressure or subglacial drainage geometry remains difficult due both to a sparsity in observations of sub- and englacial hydrology and also uncertainty in the many controlling model parameters which are generally unknown, ranging from englacial porosity and bedrock asperity size to factors controlling the semi-empirical relationships used to specify water fluxes. The lack of *in situ* observations of these controlling parameters is a limiting factor in the operational use of subglacial hydrology models in predicting water pressures and in predicting the influence of water on glacial dynamics.

In this work, we take an inverse approach, and seek to quantify parameter values in a coupled subglacial hydrology and sliding model from observations of surface velocity and terminal discharge. To simplify matters, we utilize the simplest possible model of subglacial and englacial hydrology that can still be reasonably expected to capture observed dynamics. In particular, we adopt the 'lumped' model of *Bartholomaus et al.* [2011], who proposed to treat both englacial storage and subglacial storage in a spatially-averaged sense, thus collapsing the model into a set of coupled ordinary differential equations. The relationship between the lumped model and a spatially distributed one was demonstrated by *Bueler* [2014]. The linkage between the spatially-averaged model used in this work and several other contemporary models which do not explicitly account for channel processes was demonstrated by *Bueler and van Pelt* [2015] We extend the model both by non-dimensionalization to identify the true parameter ratios controlling model dynamics, and also by introducing a Manning relation to predict output flux, a feature that the model previously lacked.

As a test case we adopt the Kennicott Glacier, Wrangell Mountains, AK, which is the same glacier for which the *Bartholomaus et al.* [2011] model was initially developed. Kennicott Glacier is approximately 43km long, and covers around 400km^2. The maximum thickness is not precisely known, but is believed to be at least 450m. It experiences an annual flood due to the outburst of a marginal lake, which reorganizes the subglacial hydrologic system. For summer of 2006, both discharge and surface velocity measurements are available, as well as a well-constrained estimate of input flux resulting from both surface melt and marginal lake drainage [*Bartholomaus et al.*, 2008].

As an inversion strategy, we adopt a Bayesian perspective as a means to estimate distributions of unknown model parameters [*Tarantola*, 2005]. In particular, we formulate a likelihood function by considering the misfit between modelled and observed surface velocities and output fluxes, subject to assumed observational and model uncertainties. Using explicit prior assumptions about

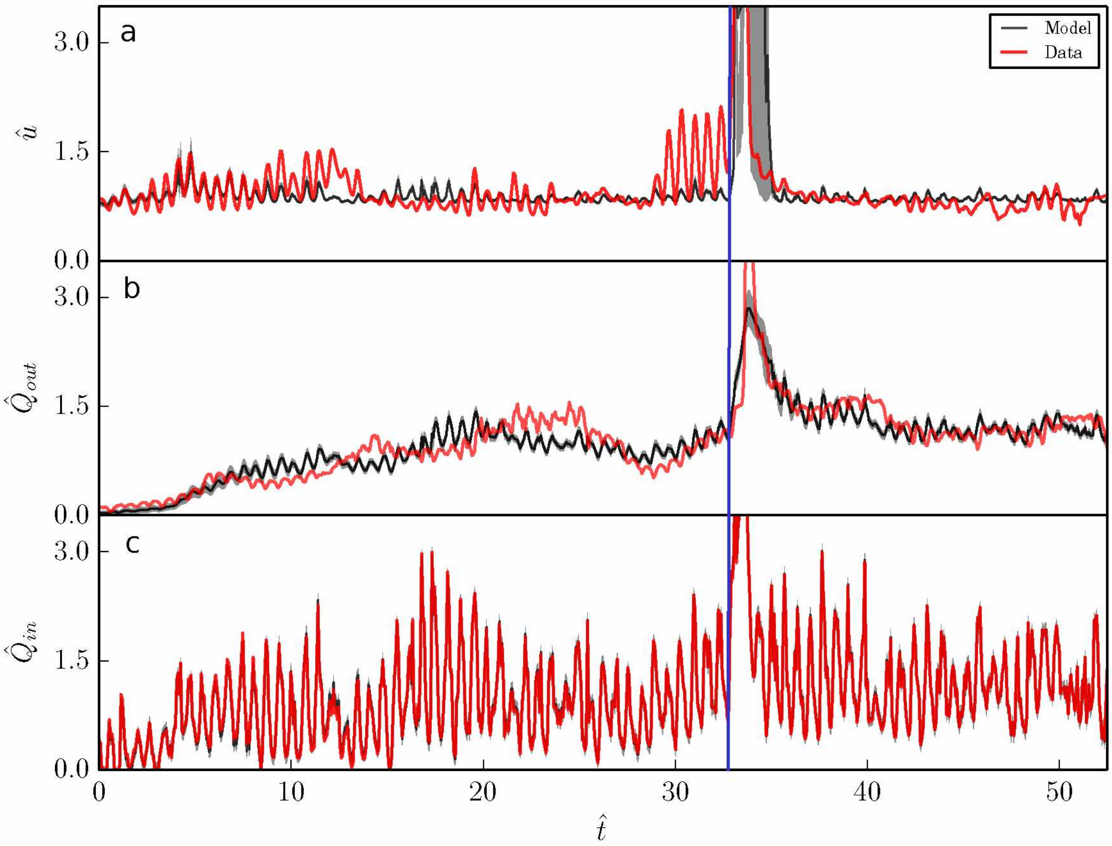

Figure 2.1. Observed (red) and modelled (black) non-dimensionalized velocity, input flux, and output flux. Gray envelopes correspond to the 1σ credibility interval. Blue line indicates approximate time at which marginal lake outburst flood occurred.

parameters, we can then draw samples from the posterior probability distribution of each model parameter. Not only does this procedure provides estimates of the most probable set of parameters, it also provides the covariance structure of the joint distribution of model parameters, which is necessary for assessing both the uncertainty and uniqueness of each parameter.

2.2 Data

The three data sets utilized in this work are shown in red in Figure 2.1. They span of approximately 75 days, beginning in mid-May 2006, and ending at the end of July 2006. At Julian day 185 (corresponding to non-dimensional time $\hat{t} = 32.5$ in Fig. 1), a marginal lake 15km from the glacier terminus drained, producing both an outburst flood and a large speed-up event. Velocity data were derived from a differential GPS located near the glacier center line approximately 14km from the glacier terminus. Output flux was assessed with sonic ranger measurements of stage on the Kennicott River, the primary outlet stream of the glacier. Input flux was computed using a positive degree day model [e.g. *Hock*, 2005] calibrated using temperatures and specific mass balance measurements at five stake locations located near the glacier centerline at various elevations. See *Bartholomaus et al.* [2011] for further details regarding the specifics of each of these data sets.

2.3 Model Description

The model of the subglacial hydrologic system considered here treats the subglacial-englacial hydrologic system as area-averaged quantities over the extent of the glacier. It is fashioned after that proposed by *Bartholomaus et al.* [2011], with a few adaptations. Additionally, we have adopted the notation used by *Bueler* [2014]. It consists of two linked elements. First, storage of water in the englacial system is parameterized by its proxy water pressure $P(t)$. In particular, under the assumption of an englacial drainage system that is macroporous and well-connected to the subglacial system, storage in the englacial system defines a water table which corresponds directly to subglacial water pressure. Thus, the water pressure in the englacial system is given by a simple mass budget

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{\rho_w g}{LW\phi}\left(Q_{in}(t) - Q_{out}(t) - \frac{fLW}{h}\frac{\mathrm{d}A_c}{\mathrm{d}t}\right), \tag{2.1}$$

where L and W are the glacier length and width, respectively, ϕ the englacial porosity, h an average bedrock bump height, and f a geometric factor related to the geometry of bedrock bumps. The essential statement of this equation is that the change in water pressure is governed by the flux into the system from the surface minus the flux out the terminus, less the change in capacity of the subglacial system.

Average cavity size $A_c(t)$ (a proxy for subglacial storage) is governed by the opening of cavities due to sliding, the melting of cavities due to turbulent heat generation, and creep closure [*Nye*, 1976]

$$\begin{aligned} \frac{\mathrm{d}A_c}{\mathrm{d}t} &= u_b h \\ &+ \left(\frac{1-\gamma_r}{\rho_i L_f}\right)\left(\frac{\lambda_x}{W}Q_{out}(t)\right)\nabla P \\ &- C_c A_c (P_0 - P)^n, \end{aligned} \tag{2.2}$$

where u_b is the sliding velocity, γ_r the Rothlisberger constant [*Röthlisberger*, 1972], $C_c = 2/(Bn)^n$ an effective ice softness, and ∇P the hydraulic gradient, which we henceforth approximate as $\nabla P \approx \frac{P}{\ell}$ where ℓ is a water pressure gradient length scale. P_0 denotes the ice overburden pressure. $A_c(t)$ is constrained to be positive, which reflects the fact that a cavity cannot possess negative cross-sectional area. $P(t)$ is constrained to lie between zero and P_0. The lower bound reflects that

Table 2.1. Dimensional constants and their values as used by *Bartholomaus et al.* [2011].

Param.	Value	Units	Description
ρ_w	1000	kg m^{-3}	Water density
ρ_i	917	kg m^{-3}	Ice density
g	9.81	m s^{-2}	Gravitational acceleration
L	15	km	Glacier length
W	4	km	Glacier width
H	400	m	Glacier thickness
ℓ	-	m	Pressure gradient length scale
ϕ	2×10^{-3}		Englacial porosity
f	0.05		Bedrock form factor
h	2		Bedrock bump height
λ_x	10	m	Bedrock bump wavelength
γ_r	0.303		Röthlisberger constant
L_f	3.35×10^5	J kg^{-1}	Latent heat of fusion
n	3		Glen's flow law exponent
C_c	1.77×10^{-25}	(Pa $s)^{-n}$	Creep parameter
α	$\frac{5}{4}$		Cavity flux exponent
β	$\frac{3}{2}$		Pressure flux exponent
r	-	$Pa^{1-\beta}$ s^{-1} $m^{\beta+2(1-\alpha)}$	Flux constant
$k\tau_b^n$	8×10^{-4}	m a $^{-1}$ Pa^{γ}	Scaled driving stress
P_0	3.6	MPa	Ice overburden pressure

it is unlikely for a subglacial cavity to hold a vacuum [e.g *Schoof et al.*, 2012]. Properly, the upper bound on pressure should be $(\rho_w/\rho_i)P_0$, where ρ_w and ρ_i are water and ice density, respectively. This bound is the limit at which the englacial water table would overtop the glacier itself, assuming a well-connected englacial cavity network, as is thought to be the case at Kennicott Glacier [*Bartholomaus et al.*, 2011]. Nonetheless, our chosen sliding law becomes singular at pressures greater than overburden (which reflects the entire glacier reaching flotation), as it was developed without the consideration of either well-connectedness or more exotic mechanisms by which overburden pressure can be exceeded such as hydrofracturing [*Tsai and Rice*, 2010]. As such, we impose this more restrictive constraint.

The forcing function $Q_{in}(t)$ is specified by data. *Bartholomaus et al.* [2011] also consider $Q_{out}(t)$ as known, but this amounts to the specification of both an influx and outflux boundary condition. This is problematic from a physical perspective and amounts to the odd conceptual situation of a fixed volume pump being attached to *both* ends of the hydrologic system. Such a configuration implies that the model is permanently sensitive to the initial pressure, as there is no mechanism for the self-correction of over- and underpressure, which has serious implications for the interpretation of recovered parameter values.

Here, we take an alternative approach: rather than specify the output flux, we use a generalized Manning relationship relating water pressure and channel size to output flux [e.g. *Walder*, 1986], namely

$$Q_{out}(t) = rA_c^{\alpha}\left(\frac{P}{\ell}\right)^{\beta-1}, \tag{2.3}$$

where the hydraulic gradient has again been approximated as proportional to water pressure over a gradient length scale ℓ.

To close the model, we require a constitutive relationship for basal velocity. After *Bartholomaus et al.* [2011], we adopt the commonly used sliding law

$$u_b = \frac{k\tau_b^n}{(P_0 - P)^{\gamma}}, \tag{2.4}$$

where k is a constant, and τ_b is the basal shear stress [e.g *Bindschadler*, 1983; *Jansson*, 1995]. Despite notable limitations, this sliding law is simple and has been used extensively in the literature, allowing the specification of reasonable prior information about the value of its flow exponent.

Table 2.2. Parameters computed using the constants defined in Table 1 (Column 2) and the maximum a posteriori probability (MAP) parameter estimates determined using the methods described in this paper (Column 3). Parameters above the doubled line are parameters directly controlling model dynamics, while below parameters represent additional sources of uncertainty in model input or output data.

Param.	Bartholomaus (2011)	MAP	Description
$\hat{k}$	-	0.44	Basal traction coef.
γ	0.22	0.4	Pressure sliding exp.
Ψ	0.018	0.61	Channel melt coef.
$\hat{r}$	-	0.02	Linear flux coef.
χ	0.11	3.41	Englacial storage coef.
Π	0.09	0.44	Sub-/englacial exchange coef.
α	$\frac{5}{4}$	1.98	Cavity/flux exp.
β	$\frac{3}{2}$	1.54	Pressure/flux exp.
$\hat{Q}_{in}$	-	-	Input flux
u_d	-	0.281	Deformational velocity
$\hat{P}_0$	-	0.2	Initial pressure
$\hat{A}_0$	-	0.9	Initial cavity size

2.3.1 Non-dimensionalization

The model described above has 23 parameters. Many of these are well constrained (e.g. physical constants like g). However, many are not. For example, the value of the geometric factor f, which describes the ratio of asperity size to spacing, is not practically observable. Simultaneously, it multiplies another underconstrained parameter ϕ, the englacial porosity. This is problematic: these two parameters could take on vastly different values, but so long as their product remained the same, the model would produce the same result. Stated another way, these parameters are not independent: the value of one depends on the other. Many such dependent pairings of parameters exist in models of subglacial hydrology. Given that their behavior cannot be independently deduced, it makes sense to treat them as a single parameter.

A formal process of non-dimensionalization identifies independent parameters, while simultaneously scaling the value of the state variables to be $\mathcal{O}(1)$. We denote non-dimensional parameters with a hat, and their associated scaling factor with a tilde. We introduce the following relations

$$\begin{aligned} A_c &= \widetilde{A}\hat{A}, \quad P = \widetilde{P}\hat{P}, \quad t = \widetilde{t}\hat{t}, \quad Q_i = \widetilde{Q}\hat{Q}_i, \\ u_b &= \widetilde{u}\hat{u}_b, \quad k = \widetilde{k}\hat{k}, \quad r = \widetilde{r}\hat{r}. \end{aligned} \tag{2.5}$$

Some scaling factors emerge naturally from data. We define $\widetilde{P} = P_0$, such that pressure scales from 0 at atmospheric pressure to unity at overburden. We scale the flux terms by $\widetilde{Q} = \mathbf{E}[Q_{obs}]$, where $\mathbf{E}[Q_{obs}]$ is the mean of the available output flux data (and due to mass conservation, quite near the mean of the input flux data). Similarly, we scale basal velocity by choosing $\widetilde{u}$ to be the mean of the observed surface velocities. With these scales fixed, we choose the remaining ones such that the number of parameters remaining in the model is minimized. Although several parameterizations exist, we choose to use the following scaling

$$\begin{aligned} \widetilde{t} &= \frac{1}{C_c P_0^n} \\ \widetilde{A} &= \widetilde{t}\widetilde{u}h \\ \widetilde{r} &= \frac{\widetilde{Q}}{\widetilde{A}^{\alpha}\left(\frac{P_0}{l}\right)^{\beta-1}} \\ \widetilde{k} &= P_0^{\gamma}\widetilde{u}\tau_b^{-n}. \end{aligned} \tag{2.6}$$

Substituting these non-dimensional parameters and substituting the constitutive relationships for flux and velocity into the state equations produces the following non-dimensional model

$$\frac{\mathrm{d}\hat{A}}{\mathrm{d}\hat{t}} = \frac{\hat{k}}{(1-\hat{P})^{\gamma}} + \Psi\hat{r}\hat{A}^{\alpha}\hat{P}^{\beta} - \hat{A}(1-\hat{P})^{n} \tag{2.7}$$

$$\frac{\mathrm{d}\hat{P}}{\mathrm{d}\hat{t}} = \chi\left(\hat{Q}_{in}(\hat{t}) - \hat{r}\hat{A}^{\alpha}\hat{P}^{\beta-1} - \Pi\frac{\mathrm{d}\hat{A}}{\mathrm{d}\hat{t}}\right), \tag{2.8}$$

where we have introduced the non-dimensional parameters

$$\Psi = (1-\gamma_r)\left(\frac{P_0}{\rho_i \ell L_f}\right)\left(\frac{\lambda_x}{W}\right)\frac{\widetilde{Q}\widetilde{t}}{\widetilde{A}} \tag{2.9}$$

$$\chi = \left(\frac{\rho_w g}{LW\phi}\right)\frac{\widetilde{Q}\widetilde{t}}{P_0} \tag{2.10}$$

$$\Pi = \left(\frac{fLW}{h}\right)\frac{\widetilde{A}}{\widetilde{t}\widetilde{Q}}. \tag{2.11}$$

Each of these groups acts to scale a particular term: Ψ serves to determine the relative importance of turbulent heat generation on cavity opening, χ gives the degree to which changes in englacial storage modulate fluxes, and Π governs the exchange rate between englacial and subglacial systems.

In addition to the parameters derived from non-dimensionalization, the model has two constitutive parameters $\hat{k}$ and $\hat{r}$, which linearly scale the velocity and flux terms, respectively. $\hat{k}$ in particular can be thought of as a non-dimensional basal traction. Note that it would be possible to choose a scaling that eliminates $\hat{k}$, and reduces the number of parameters by one. However, in so doing we would lose the means to redimensionalize velocity, which is critical for inverse methods involving velocity data.

Finally, the model depends on three exponents. γ is the non-linear dependence of sliding speed on effective pressure, while α and β relate average cavity size and pressure to discharge via the Manning relation.

2.4 Inversion

As discussed in 2.3, our chosen glacier hydrology model has eight parameters describing its dynamical evolution, and we seek to estimate these parameters given the observations described in 2.2. We approach the matter of parameter estimation from the perspective of Bayesian inference. Bayes' theorem [e.g. *Tarantola*, 2005] provides a means to compute the distribution of parameter values, given prior assumptions about parameter values coupled with data according to the following formula

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m}), \tag{2.12}$$

where $\mathbf{m}$ is the vector of model parameters, $\mathbf{d}$ is the vector of observations, and $P(\cdot)$ denotes a probability density. The first term on the right hand side is known as the likelihood. This quantifies the probability of observing the data given a particular value of $\mathbf{m}$. The second term is a prior probability (or simply 'prior'), or the pre-supposed distribution of parameter values prior to consideration of the data. The left hand side of Bayes' theorem is referred to as the posterior probability, or the probability distribution of a given parameter after having considered the data. With this probability distribution in hand, it is trivial to determine the most likely parameters, and under certain assumptions this procedure is equivalent to minimizing a least squares misfit function. However, possessing the complete distribution also gives us a rigorous assessment of parameter covariance.

To specify a likelihood model, we assume that observations of both discharge and surface velocity are independent and normally distributed around the true value at each data point with pointwise variances $\sigma^2_{Q,i}$ and $\sigma^2_{u,j}$. Thus the likelihood is

$$\begin{aligned} P(\mathbf{d}|\mathbf{m}) \quad &\propto \quad \prod_{i \in n} \exp\left(-\frac{(\hat{Q}_{out}(\hat{t}_i, \mathbf{m}) - \hat{Q}_{obs,i})^2}{2\sigma^2_{Q,i}}\right) \\ &\times \quad \prod_{j \in k} \exp\left(-\frac{(\hat{u}_s(\hat{t}_j, \mathbf{m}) - \hat{u}_{obs,j})^2}{2\sigma^2_{u,j}}\right) \end{aligned} \tag{2.13}$$

where n and k are the number of discharge and velocity measurements.

Selection of the data variances is somewhat subjective. While the measurements themselves are sufficiently precise to be considered nearly error-free, we note that these measurements are necessarily point measurements, and that the model produces only area averaged quantities. As such, the specified variances should be interpreted as including the uncertainty induced by extrapolating from a point measurement to the area average over the glacier. We suppose uncertainties similar to the magnitude of the diurnal fluctuations. For the velocity estimate this is approximately $\sigma_u = 0.4$, and for discharge $\sigma_Q = 0.6$. We also note here that we do not consider the velocity data during the flood and the enhanced diurnal signals leading up to it, as we would not expect the assumed sliding law to remain valid during such an event, where pressures are likely at or exceeding overburden [*Bartholomaus et al.*, 2011].

We also need to specify priors. For the non-dimensional groups Ψ, χ, and Π, we adopt a positively constrained uniform distribution, which effectively contributes no prior information aside from positivity. We truncate the prior distribution at ten. Experimentation has shown that this choice of upper bound does not have an effect on the computed posterior distributions. We use the same prior for the velocity and flux scaling factors $\hat{k}$ and $\hat{r}$.

Conversely, the value of the constitutive exponent γ is fairly well constrained by previous studies [*Jansson*, 1995; *Sugiyama and Gudmundsson*, 2004]. We choose to model γ as

$$\gamma \sim \ln\mathcal{N}(\mu = -0.95, \sigma = 0.3), \tag{2.14}$$

where the Log-normal distribution is given by

$$\ln\mathcal{N}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] \tag{2.15}$$

. This distribution has a mean of $\mathbf{E}[\gamma] \approx 0.4$ and variance $\mathrm{Var}(\gamma) \approx \frac{2}{125}$. α, and β are similarly constrained by theory, and we model these as Log-normally distributed as well

$$\alpha \sim \ln\mathcal{N}(0.35, 0.32) \tag{2.16}$$

$$\beta - 1 \sim \ln\mathcal{N}(-0.78, 0.43) \tag{2.17}$$

$$\tag{2.18}$$

which have expected values of $\mathbf{E}[\alpha] \approx \frac{3}{2}$ and $\mathbf{E}[\beta] \approx \frac{3}{2}$ respectively, corresponding to common literature values [e.g *Fowler*, 1986; *Werder et al.*, 2013]. The variance for each of these distributions is $\mathrm{Var}(\alpha) \approx \frac{1}{4}$ and $\mathrm{Var}(\beta) \approx \frac{1}{20}$. We note that while the means of each of these distributions was chosen to correspond to published values, the variances were specified heuristically; in an absence of any rigorous empirical estimates of variance, we chose values that seemed to provide a plausible estimate of values that these parameters might assume.

In addition to the eight governing parameters, we must also include uncertainty in the input function $\hat{Q}_{in}$ and in converting from basal velocity $\hat{u}_b$, which is modelled, to surface velocity $\hat{u}_s$, which is observed, by estimating the deformational velocity $\hat{u}_d$. Effectively, this means that we must assign prior distributions to each of these quantities and include them as additional parameters in the inversion.

We model uncertainty in $\hat{Q}_{in}(\hat{t})$ as a multivariate Normal distribution with a mean given by value from a positive degree day model and a Gaussian covariance with correlation time scale $\tau = 1$ (i.e. random variations occur smoothly and over the characteristic time scale of the differential equation), which in dimensional time corresponds to $\tau \approx 1.4$d. We assume a standard deviation of $\sigma_{in} = 0.2$.

Basal velocities produced by the forward model are not directly comparable to the surface velocity data, and we do not know *a priori* the proportion of the surface velocity accounted for by deformational velocity $\hat{u}_d$. To account for this, we assume that

$$\hat{u}_s(\hat{t}) = \hat{u}_b(\hat{t}) + \hat{u}_d \tag{2.19}$$

where $\hat{u}_d$ is given by

$$\hat{u}_d \sim \text{Unif}(0, \min(\hat{u}_{obs})). \tag{2.20}$$

This implies that the surface velocity is the sum of a modelled time-varying basal velocity, and a constant but unknown deformational velocity. The deformational velocity may, as end-member cases, account for either none of the surface velocity or all of the velocity signal minus the time-varying component. This is a more conservative assumption than modelling deformational velocities; the shallow ice approximation would not yield an accurate result in this context, and we have insufficient geometric information to model higher order stresses.

Finally, we also specify Log-normally distributed priors for the initial conditions on $\hat{P}$ and $\hat{A}_c$

$$\hat{P}(\hat{t} = 0) \sim \ln \mathcal{N}(-1.12, 0.64) \tag{2.21}$$

$$\hat{A}_c(\hat{t} = 0) \sim \ln \mathcal{N}(0.49, 0.64), \tag{2.22}$$

which enforce non-negativity, but otherwise provide relatively little information. Experimentation has shown that the choice of initial conditions has a very minimal effect on model results, and that after a short ($\approx \hat{t} = 1$) equilibration period even extremely improbable initial conditions yield very similar solutions to more reasonable ones.

2.4.1 Sampling

The posterior distribution cannot be computed directly, and must be characterized with samples instead. With a large number of these in hand, we can then evaluate the statistical properties of the samples as a proxy for the joint posterior distribution of the parameters.

There are many choices of sampling algorithm, but we chose the Adaptive Metropolis Algorithm (AMA) [*Haario et al.*, 2001], which is a variant of the classic Metropolis-Hastings (MH) algorithm [*Hastings*, 1970]. The MH algorithm works by travelling through parameter space, sequentially updating each parameter independently by randomly drawing a jump from a proposal distribution. If the posterior probability is greater at the new location than at the present, the new parameter value is accepted. Otherwise, it is accepted with probability proportional to the ratio of the current posterior probability and that of the proposed value.

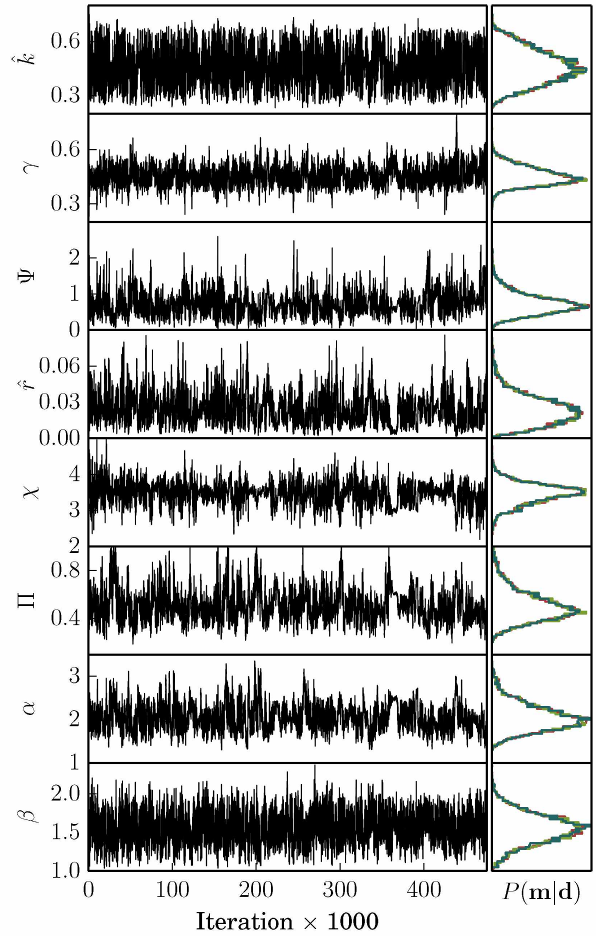

Figure 2.2. (Left) Sampling history of each state parameter. The fuzzy appearance is an indicator of the algorithm efficiently and fully exploring the feasible parameter space. (Right) Histograms of each parameter's marginal distribution for three independently sampled parameter populations. The identical posterior distributions strongly indicate population convergence.

The AMA functions identically, except that the proposal distribution is updated by iteratively constructing a covariance matrix from the previous samples, and proposing a new parameter set en masse, rather than individually. This variant is particularly suited for this problem, where parameters tend to be strongly correlated and the covariance matrix can help to identify suitable proposal steps. Simultaneously, by block updating the parameters, the number of forward model evaluations is reduced, increasing efficiency. We use the implementation of these algorithms available in the PyMC package [*Patil et al.*, 2010].

We drew 5×10^5 samples from the joint posterior distribution of the parameters given in Table 2, repeated three times in order to assess consistency between sample populations. In order to assess convergence, we relied upon a heuristic examination of the history of each parameter as it was sampled; well-converged sample populations typically traverse the feasible parameter space many times, as shown in Figure 2.2a. Additionally, we compared sample histograms between populations to assess whether the posterior distribution was unique, which are shown in Figure 2.2b. As a numerical convergence test, we computed the Gelman-Rubin statistic [*Gelman and Rubin*, 1992], which compares the intra-population variance to the inter-population variance. In the limit as the sample size goes to infinity, the Gelman-Rubin statistic R converges to unity. A well converged sample set that exhibits limiting statistical behavior should also have an R value near one. In our case, for each parameter tested $R \ll 1.1$, which is typically taken to indicate adequate convergence.

2.5 Results and Discussion

2.5.1 State variables

Figure 2.1a shows the posterior distribution of velocity along with velocity observations. It is immediately evident that the model reproduces the long time scale variability in velocity observations well. Additionally, we see that the model is also capable of reproducing diurnal variability in both magnitude and duration. The good fidelity to diurnal timing should not be surprising: these features are forced by the input data, and the input data exhibits the same structure, as seen in Figure 2.1c. Adequately reproducing the magnitude of velocity fluctuations is more difficult and more important, as this imposes strong constraints on pressure.

One case where the model does not reproduce the velocity observations well is during and before the flood that occurred at $\hat{t} \approx 32$. Although it is not well shown in Figure 2.1a, the model predicts velocities of approximately twice those observed, with a broader peak. Two observations seem relevant here. First, this misfit is likely due to inadequacies in the sliding law both in the sense

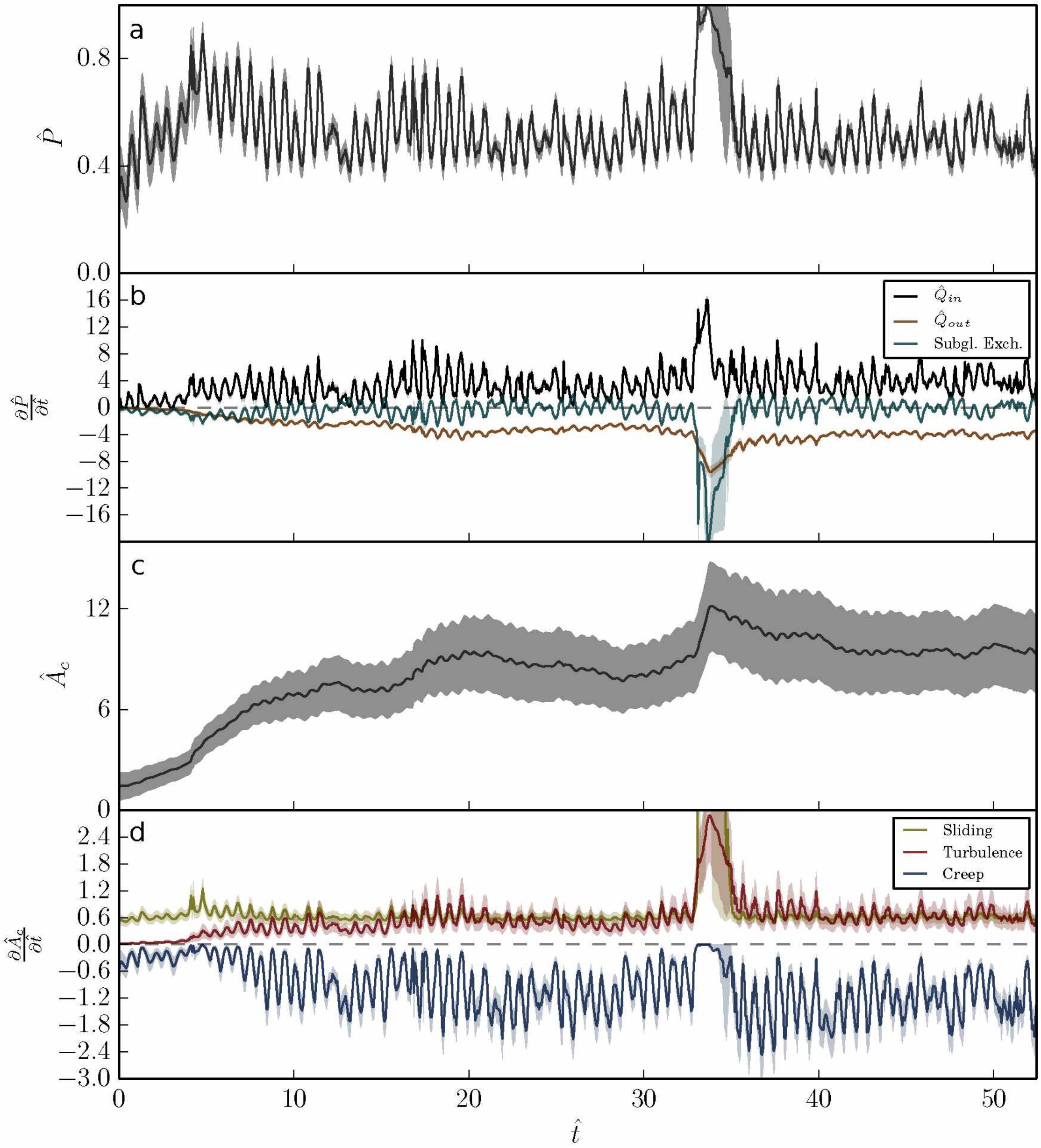

Figure 2.3. Posterior distributions and observations of non-dimensional pressure (a) and average cavity size (c), and magnitude of model terms (b and d). Envelopes correspond to the 1σ credibility interval.

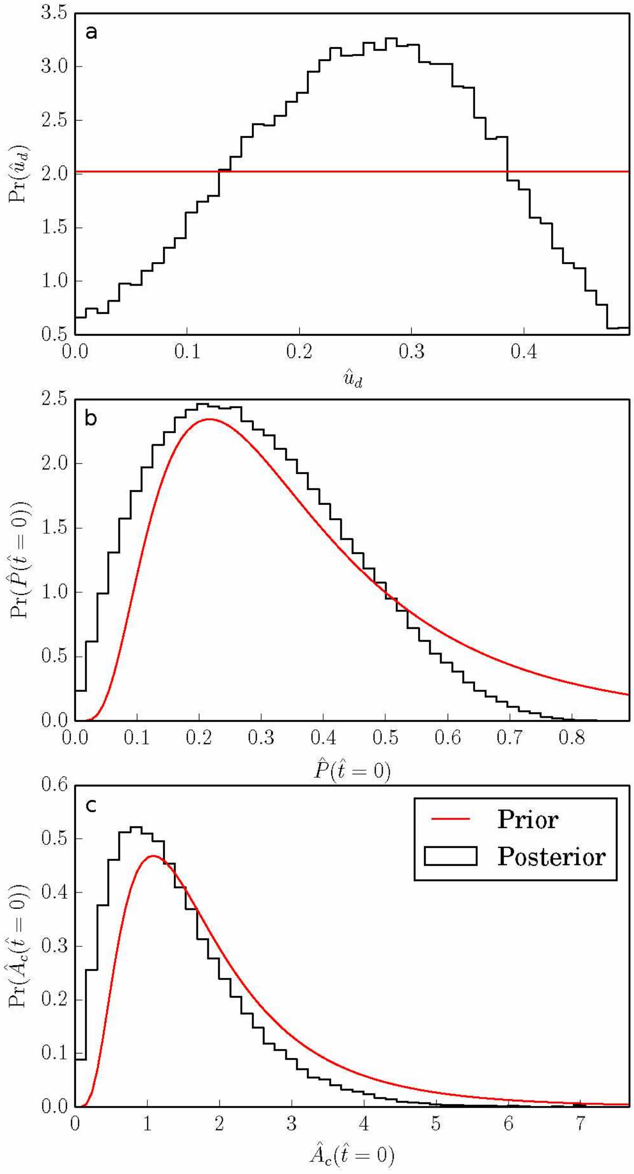

Figure 2.4. Posterior distributions of peripheral variables deformational velocity u_d, pressure initial condition $\hat{P}(\hat{t}=0)$, and cavity area initial condition $\hat{A}_c(\hat{t}=0)$.

that it is not well-equipped to handle pressures very near overburden, and also because it is unable to include effects which would serve to mitigate the velocity increase such as longitudinal and transverse stresses. Secondly, the large diurnal fluctuations evident before the main velocity peak indicate that processes occur which are either not captured by the input data or are not captured by the model physics. An example of the former would be an additional pre-emptive lake drainage or other anomalous source of extra water. This seems unlikely, as such an input would appear in the output flux data, which it does not. An example of the latter could be the inability of the model to capture local hydraulic jacking effects due to the lake partially draining and subsequently refilling, without overcoming the necessary pressure barrier to route excess water into the greater subglacial drainage system.

Figure 2.1b show the posterior distribution of modelled output flux along with observations. Once again, the model effectively captures both the long term variability in output flux, as well as the frequency and magnitude of diurnal variability. However, as with the modelled velocity, some limitations are evident. First, the modelled output fluxes are offset by around half a diurnal period. This is a result of the spatially averaged nature of the model, particularly the fact that pressure is assumed to propagate instantaneously through the system, and output flux responds immediately to variations in input flux. In reality, pressure changes induced in the upper reaches of the system would take some amount of time to propagate down-gradient, and this propagation time would be dependent upon the state of the hydrologic system.

Another notable instance where the model fails to reproduce observations occurs just prior to the flood event. Here, input and output fluxes are out of phase with one another by around ten diurnal cycles. This too could be a result of the lack of spatially heterogeneous storage. This misfit is less pronounced or non-existent following the flood. An explanation for this is that the drainage system has grown more efficient, causing the implicit assumption of uniform spatial response times to become a better approximation to reality. This modelled increase in cavity size is evident in Figure 2.3c, which shows a marked increase in average cavity size following the flood. As cavities become larger (and presumably better connected), we would expect the lumped model to capture the temporal variability of the system with more fidelity.

The modelled englacial water pressure as a fraction of overburden is shown in Figure 2.3a. While no direct observations of water pressure are available, we observe that the model reproduces the expected diurnal fluctuations, and that under non-flood conditions, the glacier tends to oscillate between around 40% and 80% of overburden. The magnitude of the modelled fluctuations are of a similar magnitude to observations in many mountain glaciers [*Harper et al.*, 2007; *Amundson et al.*, 2006], lending support to the reliability of predictions. During the flood event, the pressure increases to overburden (which it is constrained not to exceed), though the uncertainty during this event is high.

We can also examine the relative magnitude of each mechanism contributing to the evolution of the water pressure on a termwise basis. Figure 2.3b shows the individual contributions of the input flux, output flux, and subglacial exchange terms through time. For dynamic equilibrium to occur (i.e. no long term storage change), the three terms, on average, must sum to zero, which they do less the water moving out of subglacial storage and into englacial storage. Input flux and output flux are both uniformly and respectively positive and negative (by definition). The more interesting contributor to the evolution of the pressure state is the subglacial/englacial exchange term Π, which acts as a buffer for the large swings in input flux while maintaining an approximately zero mean. This short term fluctuation in subglacial storage is responsible for the evident attenuation in flux magnitude between input and output. Note that in Figure 2.3b, a positive value of the subglacial exchange term implies that water is moving into englacial storage from subglacial storage. Predictably, this state occurs when water pressure is low, and creep closure acts to drive water out of cavities.

Figure 2.3b shows the non-dimensional cavity size. While we cannot compute the dimensional constant $\widetilde{A}$ that would be necessary to re-dimensionalize the modelled cavity size, we note that the (known) time and velocity scales imply that it is of the same order as the bedrock bump height. As such, $\hat{A}$ is similar to A_c so long as average bedrock asperities are on the order of meters in height. We first note that the variance in cavity size distribution is higher than the other state variables. This implies that the parameters controlling it cannot be precisely determined given the available data. Alternatively, it would appear that the specifics of cavity formation play a less critical role in explaining surface velocity and output flux than does the pressure, commensurately limiting the amount of information that can be used to constrain its governing parameters. Nonetheless, it remains possible to make a qualitative assessment of cavity evolution over the modelled time period. At the beginning of the simulation, cavities are relatively small. The cavities grow during pressure-driven speed-up events, but the increasingly well-developed cavities tend to damp this response as time goes on. During the lake drainage event, the average cavity size increases significantly. This increase in size decays over the course of a few days, as the reduced water pressure following drainage is no longer capable of sustaining large cavities and creep closure becomes the dominant mechanism of channel evolution.

Once again, it is useful to examine each modelled mechanism's contribution to cavity evolution, as shown in Figure 2.3d. Near the beginning of the simulation when cavities are small, sliding is the dominant mechanism of cavity opening. During this stage, there is insufficient flux through the cavities to support the production of much turbulent heat. Simultaneously, creep closure has yet to act strongly, because the cavities are still relatively small and the closure rate scales linearly with cavity area. As the cavities grow, opening due to bedrock sliding remains relatively constant, while both turbulent dissipation and creep closure grow in magnitude. Creep closure in particular exhibits strong diurnal variations as the large variations in water pressure are amplified by the nonlinearity in ice rheology, with nearly no closure occurring when effective pressures are near zero. In particular, during the highly pressurized flood event, creep closure effectively shuts down for several days while both enhanced basal sliding and increased output flux rapidly enlarge subglacial cavities.

Evaluating the uncertainties associated with each term in Equation 2.7 gives us an understanding of the source of the relatively large degree of uncertainty associated with cavity size. In particular, we see that the cavity opening rate is subject to a much larger relative degree of uncertainty than other terms in Figure 2.3d, and it is this uncertainty that generates the large spread evident in Figure 2.3c. The greater degree of uncertainty in the magnitude of this term is to be expected; like opening due to sliding, dissipative heating is directly associated with one of the observed quantities (namely output flux). However, unlike opening due to sliding, dissipative heating is associated with the additional free parameter Ψ apart from the constitutive flux relation.

Finally, we can look at the posterior distribution of the deformation velocity and initial conditions, as shown in Figure 2.4. In the case of deformational velocities, the model has a slight tendency towards predicting deformational velocities near the center of the admissible range. Due to the insensitivity of model dynamics to initial conditions, the posterior distributions of both initial conditions are very similar to their prior distributions. In general, the inversion procedure contributes little useful information towards these parameters, and in this context they should be viewed primarily as sources of additional uncertainty with respect to the other model parameters.

2.6 Configuration Stability

Because we have in our model included turbulent heat generation as a mechanism for melting cavity walls, we must also assess whether or not our envisaged drainage configuration is compatible with the parameter values that we have recovered from a stability perspective. A well known result from *Kamb* [1987] shows that for water pressures above a threshold value and a given hydraulic gradient, a linked cavity system undergoes runaway evolution towards a channelized system. We note these results, and the similar ones found in, for example, *Schoof* [2010] and *Hewitt* [2011] are based on linear stability analysis, and only strictly valid for autonomous systems of equations. The inclusion of a time-varying influx term clearly makes this system of equations non-autonomous, and the analysis of the stability of such systems is beyond the scope of this paper. Nonetheless, we can find model steady states for a prescribed steady influx (i.e. a fixed $\hat{Q}_{in}$ as a opposed to a time varying one) in order to assess the stable states of the model.

In steady state, Equations 2.7 and 2.8 reduces to

$$\hat{A}_c = (1-\hat{P})^n \left[\frac{\hat{k}}{(1-\hat{P})^\gamma} + \Psi \hat{Q}_{in} \hat{P} \right] \tag{2.23}$$

$$0 = \hat{Q}_{in} - \hat{r} \hat{A}_c^\alpha \hat{P}^{\beta-1}, \tag{2.24}$$

the numerical solution of which is straightforward to compute. Furthermore, we can compute the stability of each of these points by evaluating the eigenvalues of the Jacobian matrix of Equations 2.7 and 2.8. Doing so for the parameters computed through inversion yields the interesting result that regardless of the chosen flux, the system has a unique and stable steady state. This is held in contrast to the results of *Kamb* [1987], *Schoof* [2010], and *Hewitt* [2011], which would predict runaway channelization due to the magnitude of the turbulent melting terms included here. The reason that this model differs is that flux is fundamentally limited by the requirement of global mass conservation; runaway channel growth is not possible because such growth efficiently evacuates the water necessary to maintain the low effective pressures that limit creep closure. This occurs because the lumped structure of the model requires that this channelization occur everywhere simultaneously. In a spatially explicit model (and presumably in reality), spatial heterogeneity allows distinct channels to remain pressurized in a self-sustaining way by drawing upon local rather than global reservoirs. The suppression of the channelizing instability by the model structure begs the question of whether the dynamics induced by such effects can credibly be neglected. This remains an unresolved issue, and we do not claim to have the answer here. Our approach is qualitatively supported by observations at the terminus of Kennicott Glacier, where there exist no obvious subglacial channels. The Kennicott terminus is characterized by a linked series of terminal lakes, none of which have visible subglacial input. Nonetheless, we imagine that during the lake outburst flood, development of large and efficient channels becomes dominant over a distributed

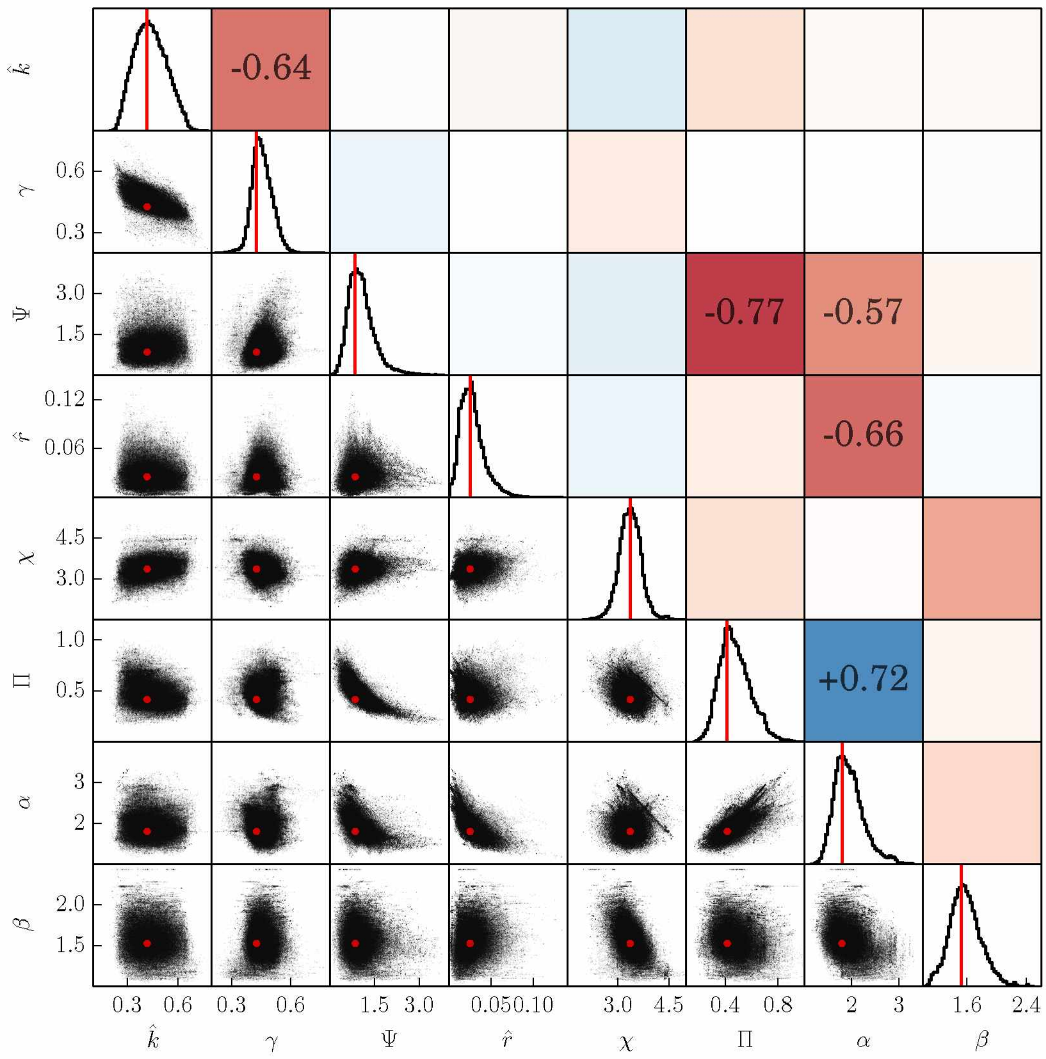

Figure 2.5. Histograms on the diagonal show the marginal posterior distribution of each parameter. The red line indicates the maximum posterior probability. Scatter plots below the diagonal indicate the relationship between each parameter set. Colored boxes above the diagonal indicate the correlation coefficient for each pair of parameters: red indicates a negative correlation, while blue indicates a positive one.

system due to the high water pressures involved, and the suppression of creep closure as seen in Figure 2.3b. During this period the model tends to overestimate the length of the perturbation in velocity (and so probably water pressure as well) due to the flood. During the lake drainage, we speculate that a channel developed which evacuated water much more quickly than this model can account for without explicit channelization.

2.6.1 Parameter covariance

In addition to assessing the feasible model states, possessing the joint posterior distribution allows us to examine the covariance structure of the model parameters as well. Figure 2.5 shows the joint sample distribution of each parameter pair, as well as the histogram of each.

The first set of strongly correlated parameters are the two which govern the sliding speed $\hat{k}$ and γ. This result is not surprising: given data uncertainty and the flexibility granted by their priors, there are a variety of combinations that these two parameters can feasibly adopt due to the inversely correlated relationship built in to the sliding law. Nonetheless, it is worth noting that both parameters have smaller variance than their priors, indicating that the data provides information about both (recall for example that the prior on $\hat{k}$ is uniform). $\hat{r}$ and α, which control the magnitude and cavity size dependence of the flux, are correlated in a similar fashion.

A more interesting negative correlation exists between Π, the parameter scaling the importance of englacial-subglacial exchange, and Ψ, the parameter controlling the rate of turbulent melting. In particular, this correlation implies that decreases in the geometric capacity for the subglacial system to store water can be offset by increases in the rate at which turbulent melting can occur. Since the magnitude of cavity opening due to bedrock sliding and closing due to creep are both constrained by the scaling of the problem, only Π and Ψ can covary. These parameters are positively and negatively correlated with α, respectively. An increase in either has the tendency to produce a greater amount of subglacial storage, but this correlation tells us that an increase in the flux dependence on $\hat{A}_c$ is more easily offset by an increase in turbulent melt rates, than by an increase in the sub-/englacial transfer rate as a whole.

A final observation is that the value of χ is of $\mathcal{O}(1)$. This parameter relates the size of the flux terms on the right hand side of Equation 2.8 to the rate of change of the pressure, and is effectively a proxy for englacial porosity. Previous work has suggested that this term is limitingly large, which is to say that englacial porosity is close to zero. This effectively assumes that changes in englacial storage occur instantaneously [*Schoof et al.*, 2012], and that the time derivative appearing in Equation 2.8 is zero. Other models have retained this term under the auspices of a non-negligible englacial porosity as a means to regularize a distributed model of subglacial cavity evolution [*Werder et al.*, 2013; *Bueler and van Pelt*, 2015]. *Clarke* [2003], though citing an alternative physical mechanism of compressibility, also included an analogous term to improve numerical stability. Our results suggest that despite the initial numerical motivation, retaining the time derivative in Equation 2.8 may also be physically correct, and also that the inclusion of significant englacial storage is required to simultaneously explain observed velocity and output fluxes. This additional storage could take the form of englacial void space [*Fountain et al.*, 2005], basal crevasses [*Harper et al.*, 2010], or a combination thereof. This further begs the question, what is the englacial macroporosity corresponding to a χ of $\mathcal{O}(1)$? Assuming the length and width scales of *Bartholomaus et al.* [2011] (See Table 2), we find that porosity must be in the range $\phi = [10^{-4}, 10^{-3}]$. The scaling analysis of Section 2.3.1 shows us that even this modest amount of porosity provides a sufficient amount of storage to provide an englacial reservoir of equivalent magnitude to the subglacial reservoir.

2.7 Conclusions

We have extended the subglacial hydrology model of *Bartholomaus et al.* [2011] in a few ways, which is 'lumped' in the sense that it treats the whole of the subglacial system using area averaged quantities. An advantage of this treatment is that it allows a simple numerical treatment, as the model consists of a pair of non-homogeneous, non-linear ordinary differential equations. Furthermore, the simplicity of the model allows for the straightforward identification of the governing parameters.

First, we have discarded the simultaneous specification of both input and output fluxes in favor of a Manning flow relation which relates output flux to both average cavity size and water pressure. Second, we have formally non-dimensionalized the model to determine the specific parameter ratios that govern model dynamics. In so doing, we identified eight parameters. Three are non-dimensional groups controlling the relative importance of turbulent melting of cavity walls (Ψ), the exchange rate between the englacial and subglacial hydrologic systems (Π), and the rate at which the englacial water system can accommodate flux imbalances (χ). The remaining five parameters describe constitutive relationships describing both the relationship between effective pressure and basal ice velocity and the relationship between average cavity size, water pressure, and outflux.

The values of these parameters are known *a priori* with varying degrees of certainty. We sought to improve these estimates through inverse modelling. Using flux and velocity observations from *Bartholomaus et al.* [2011] to construct a likelihood function, and in conjunction with prior parameter estimates, we used a Markov Chain Monte Carlo method to sample from the joint posterior probability distribution of the model parameters, conditioned upon velocity and flux data from Kennicott Glacier. Not only did this allow us to determine the most probable parameter values, but also to characterize the covariance within and between parameters.

Despite the simplicity of the lumped modelling approach, we were able to reproduce observations with a reasonable degree of fidelity. The model captures both the magnitude and timing of diurnal variability in velocity and output flux, even passably capturing the dynamics of a lake-related flooding event. The model predicted diurnal water pressure variations between 40% and 80% of overburden, which corresponds well to prior observations of borehole pressure records in similar systems. The model also produced reasonable estimates of average cavity size. Nonetheless, the limitations due to the assumption of spatial uniformity were also evident, as longer scale temporal variability, particularly in the less efficient pre-flood configuration, was not captured. It is possible that adding spatial dimensionality to the model could help to reduce some of the remaining misfit, and the way forward in doing this is clear [*Bueler*, 2014]. However, this would drastically increase the computational cost of the forward model, making the rigorous estimation of parameter covariance through Monte Carlo methods less practical. Nonetheless, it would be extremely valuable to determine whether the conclusions suggested by this work hold in the presence of more advanced physics, and if not, the reason for the inconsistency.

The parameter estimates produced by the inverse modelling procedure suggested that all of the mechanisms included in the model were important in explaining observations. Cavity opening due to basal sliding seems to dominate the evolution of the subglacial system until cavities grow large enough to support turbulent heat generation, at which point the interplay between turbulent melting and creep closure become important as well. Our results also suggest that transfer of water into the subglacial hydrologic system from the englacial system acts to attenuate the input flux signal, leading to the observed relative reductions in magnitude in the diurnal variability of output flux.

Finally, we find that the assumption of negligible englacial porosity is not compatible with observations under the assumptions of the model used here, and thus we suppose that englacial storage plays an important role in the hydrologic systems of glaciers similar to the one examined here, even for relatively low absolute values of glacier macroporosity. Addressing this supposition further will require direct measurements of englacial porosity and connectivity in more glaciers, as well as further numerical investigation through inverse modelling. If englacial porosity indeed has a ubiquitous influence on glacier dynamics, then further effort towards quantifying and predicting this value will be required in order to credibly model the effect of hydrology on glacier dynamics.

2.8 Acknowledgements

This paper was initially conceived and developed at the International Glaciology Summer School, held in McCarthy, AK. We acknowledge the summer school's organizer Regine Hock and the support of the National Science Foundation ARC-1204202, GlacioEx (Glaciology Exchange program funded by Norwegian Centre for International Cooperation in Higher Education, Partnership Program for North America), the International Association of Cryospheric Sciences, and the International Glaciological Society. Many thanks to Patricia Eugster for insightful conversations. We acknowledge the helpful comments of the scientific editor Christian Schoof and two anonymous reviewers which greatly improved the quality of this manuscript. DJB was supported by the National Science Foundation Graduate Research Fellowship grant number DGE1242789. CRM was supported by the National Science Foundation Graduate Research Fellowship grant number DGE1144152. EB was supported by NASA grant NNX13AM16G.

Bibliography

Amundson, J. M., M. Truffer, and M. P. Lüthi (2006), Time-dependent basal stress conditions beneath Black Rapids Glacier, Alaska, USA, inferred from measurements of ice deformation and surface motion, *Journal of Glaciology*, *52*(178), 347–357.

Arnold, N., K. Richards, I. Willis, and M. Sharp (1998), Initial results from a distributed, physically based model of glacier hydrology, *Hydrological Processes*, *12*(2), 191–219.

Bartholomaus, T. C., R. Anderson, and S. Anderson (2008), Response of glacier basal motion to transient water storage, *Nature Geoscience*, *1*, 33–37, doi:10.1038/ngeo.2007.52.

Bartholomaus, T. C., R. S. Anderson, and S. P. Anderson (2011), Growth and collapse of the distributed subglacial hydrologic system of Kennicott Glacier, Alaska, USA, and its effects on basal motion, *Journal of Glaciology*, *57*(206), 985–1002.

Bindschadler, R. (1983), The importance of pressurized subglacial water in separation and sliding at the glacier bed, *Journal of Glaciology*, *29*(101), 3–19.

Bueler, E. (2014), Extending the lumped subglacial–englacial hydrology model of Bartholomaus and others, *Journal of Glaciology*, *60*(222).

Bueler, E., and W. van Pelt (2015), Mass-conserving subglacial hydrology in the Parallel Ice

Sheet Model version 0.6, *Geoscientific Model Development*, *8*(6), 1613–1635, doi:10.5194/gmd-8-1613-2015.

Clarke, G. K. (2003), Hydraulics of subglacial outburst floods: new insights from the spring-hutter formulation, *Journal of Glaciology*, *49*(165), 299–313.

Engelhardt, H., and B. Kamb (1998), Basal sliding of Ice Stream B, West Antarctica, *Journal of Glaciology*, *44*(147).

Fischer, U. H., and G. K. C. Clarke (1997), Stick-slip sliding behaviour at the base of a glacier, *Annals of Glaciology*, *24*, 390–396.

Flowers, G. E., and G. K. Clarke (2002a), A multicomponent coupled model of glacier hydrology 1. Theory and synthetic examples, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *107*(B11), ECV–9.

Flowers, G. E., and G. K. Clarke (2002b), A multicomponent coupled model of glacier hydrology 2. Application to Trapridge Glacier, Yukon, Canada, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *107*(B11), ECV–10.

Flowers, G. E., H. Björnsson, and F. Pàlsson (2002), New insights into the subglacial and periglacial hydrology of Vatnajökull, Iceland, from a distributed physical model, *Journal of Glaciology*, *48*, 467–477.

Fountain, A. G., R. W. Jacobel, R. Schlichting, and P. Jansson (2005), Fractures as the main pathways of water flow in temperate glaciers, *Nature*, *433*, 618–621.

Fowler, A. C. (1986), A sliding law for glaciers of constant viscosity in the presence of subglacial cavitation, *Proceedings of the Royal Society of London A*, *407*, 147–170.

Gagliardini, O., D. Cohen, P. Råback, and T. Zwinger (2007), Finite-element modeling of subglacial cavities and related friction law, *Journal of Geophysical Research: Earth Surface (2003–2012)*, *112*(F2).

Gelman, A., and D. Rubin (1992), *A single series from the Gibbs sampler provides a false sense of security*, pp. 625–631, Oxford University Press.

Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, pp. 223–242.

Harper, J. T., N. F. Humphrey, W. T. Pfeffer, and B. Lazar (2007), Two modes of accelerated glacier sliding related to water, *Geophysical Research Letters*, *34*(12), doi:10.1029/2007GL030233, l12503.

Harper, J. T., J. H. Bradford, N. F. Humphrey, and T. W. Meierbachtol (2010), Vertical extension of the subglacial drainage system into basal crevasses, *Nature*, *467*, 579–582, doi: 10.1038/nature09398.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*(1), pp. 97–109.

Hewitt, I. J. (2011), Modelling distributed and channelized subglacial drainage: the spacing of channels, *Journal of Glaciology*, *57*(202), 302–314, doi:doi:10.3189/002214311796405951.

Hock, R. (2005), Glacier melt: a review of processes and their modelling, *Progress in physical geography*, *29*(3), 362–391.

Hubbard, B., and M. Sharp (1995), Basal ice formation in the western Alps, *Arctic and Alpine Research*, *27*(4), 301–310.

Iken, A. (1981), The effect of the subglacial water pressure on the sliding velocity of a glacier in an idealized numerical model, *Journal of Glaciology*, *27*(97), 407–421.

Iken, A., and R. A. Bindschadler (1986), Combined measurements of subglacial water pressure and surface velocity of Findelengletscher, Switzerland: conclusions about drainage system and sliding mechanism, *Journal of Glaciology*, *32*(110), 101–119.

Iken, A., and M. Truffer (1997), The relationship between subglacial water pressure and velocity of Findelengletscher, Switzerland, during its advance and retreat, *Journal of Glaciology*, *43*(144), 328–338.

Iken, A., H. Röthlisberger, A. Flotron, and W. Haeberli (1983), The uplift of unteraargletscher at the beginning of the melt season - a consequence of water storage at the bed?, *Journal of Glaciology*, *29*(101), 28–47.

Jansson, P. (1995), Effective pressure and basal sliding, Storglaciaren, northern Sweden, *Journal of Glaciology*, *41*(138), 232–240.

Kamb, B. (1987), Glacier surge mechanism based on linked cavity configuration of the basal water conduit system, *Journal of Geophysical Research*, *92*(B9), 9083–9100.

Lliboutry, L. (1968), General theory of subglacial cavitation and sliding of temperate glaciers, *Journal of Glaciology*, *7*(49), 21–58.

Nienow, P., M. Sharp, and I. Willis (1998), Seasonal changes in the morphology of the subglacial drainage system, Haut Glacier d'Arolla, Switzerland, *Earth Surface Processes and Landforms*, *23*(9), 825–843.

Nye, J. (1976), Water flow in glaciers: jökulhlaups, tunnels, and veins, *Journal of Glaciology*, *17*, 181–207.

Patil, A., D. Huard, and C. J. Fonnesbeck (2010), PyMC: Bayesian stochastic modelling in Python, *J. Stat. Softw*, pp. 1–81.

Röthlisberger, H. (1972), Water pressure in intra- and subglacial channels, *Journal of Glaciology*, *11*, 177–203.

Schoof, C. (2005), The effect of cavitation on glacier sliding, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, *461*(2055), 609–627.

Schoof, C. (2010), Ice-sheet acceleration driven by melt supply variability, *Nature*, *468*(7325), 803–806.

Schoof, C., I. J. Hewitt, and M. A. Werder (2012), Flotation and free surface flow in a model for subglacial drainage. Part 1. Distributed drainage, *Journal of Fluid Mechanics*, *702*, 126–156.

Sugiyama, S., and G. Gudmundsson (2004), Short-term variations in glacier flow controlled by subglacial water pressure at Lauteraargletscher, Bernese Alps, Switzerland, *Journal of Glaciology*, *50*(170), 353–362.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, doi:10.1137/1.9780898717921.

Truffer, M., and W. D. Harrison (2006), In situ measurements of till deformation and water pressure, *Journal of Glaciology*, *52*(177), 175–182.

Tsai, V. C., and J. R. Rice (2010), A model for turbulent hydraulic fracture and application to crack propagation at glacier beds, *Journal of Geophysical Research: Earth Surface*, *115*(F3), doi:10.1029/2009JF001474.

Walder, J. S. (1986), Hydraulics of subglacial cavities, *J. Glaciol*, *32*(112), 439–445.

Werder, M. A., I. J. Hewitt, C. G. Schoof, and G. E. Flowers (2013), Modeling channelized and distributed subglacial drainage in two dimensions, *Journal of Geophysical Research: Earth Surface*, *118*(4), 2140–2158.

Willis, I., M. Sharp, M. Tranter, B. Hubbars, M. Nielsen, C. Smart, G. Brown, R. Arnold, R. Hodgkins, H. Lamb, S. Gordon, and B. Ketterling (1995), Water storage, drainage evolution and water quality in alpine glacial environments - Final report on NERC grant GR3/8114, *Tech. rep.*

Zwally, H. J., W. Abdalati, T. Herring, K. Larson, J. Saba, and K. Steffen (2002), Surface melt-induced acceleration of Greenland Ice Sheet flow., *Science*, *297*, 218–222.

Chapter 3

Bayesian inference of subglacial topography using mass conservation[1]

Abstract

We develop a Bayesian model for estimating ice thickness given sparse observations coupled with estimates of surface mass balance, surface elevation change, and surface velocity. These fields are related through mass conservation. We use the Metropolis-Hastings algorithm to sample from the posterior probability distribution of ice thickness for three cases: a synthetic mountain glacier, Storglaciären, and Jakobshavn Isbræ. Use of continuity in interpolation improves thickness estimates where relative velocity and surface mass balance errors are small, a condition difficult to maintain in regions of slow flow and surface mass balance near zero. Estimates of thickness uncertainty depend sensitively on spatial correlation. When this structure is known, we suggest a thickness measurement spacing of one to two times the correlation length to take best advantage of continuity based interpolation techniques. To determine ideal measurement spacing, the structure of spatial correlation must be better quantified.

3.1 Introduction

Bed elevation is required to model glacier dynamics. Measurements are typically performed with spatially localized radar soundings [e.g. *Allen et al.*, 2015] and tend to be precise and dense along a line. However, because modeling often requires a thickness field, reliable methods of interpolation between observations are valuable.

Many widely used digital elevation models (DEMs) of subglacial topography are based upon classical geostatistical techniques such as Kriging [*Bamber et al.*, 2013]. Such DEMs tend to induce immediate modeled surface elevation changes from dynamical models that are implausibly larger than observations of surface elevation change [*Seroussi et al.*, 2011; *Bindschadler et al.*, 2013]. In an effort to minimize these spurious model transients, contemporary DEMs incorporate physical constraints on interpolated fields. The procedure is conceptually simple and fits neatly into the general framework of geophysical inversion theory: formulate a cost functional that quantifies the misfit between the field of interest and observations, subject to the constraint that the field be compatible with a forward model.

Inverse methods are widely used in glaciology. *MacAyeal* [1993] presented a method for using the momentum conservation equations to invert for basal shear stress using an observed surface

---

[1] Published as Brinkerhoff, D. J., A. Aschwanden, M. Truffer (2016), Bayesian inference of subglacial topography using mass conservation, *Frontiers in Earth Science*, 4.

velocity field. The same method has been used over the last twenty years mostly unchanged, though with advances in forward model sophistication and data availability [e.g. *Brinkerhoff and Johnson*, 2013; *Morlighem et al.*, 2010; *Sergienko et al.*, 2014]. *Gudmundsson and Raymond* [2008] developed a Bayesian approach that used surface elevation, surface elevation trend, and surface velocity, along with strong prior information about the bed elevation to compute the maximum *a posteriori* estimate of basal shear stress and basal topography. *Perego et al.* [2014] solved for thickness and basal traction by simultaneously inverting the mass and momentum conservation relations. *McNabb et al.* [2012] and *Morlighem et al.* [2014a] used optimal control methods and mass conservation (without an associated momentum conservation model) to infer ice thickness. *Brinkerhoff and Johnson* [2015] used a similar approach to produce seamless velocity maps of Greenland, filling data gaps in satellite-based velocity estimates with balance velocities. *Huss and Farinotti* [2012] and *Li et al.* [2011] used surface geometry with mass conservation to estimate global glacier volume, even in the absence of velocity data. However, *Bahr et al.* [2014] suggests caution in using mass conservation unconstrained by observations since uncertainty grows exponentially as resolution increases unless short wavelength topography is suppressed.

An important limitation exists in all of the above studies: none rigorously quantify the error bounds of their solutions in the sense that uncertainty estimates are linearized around the optimal solution or estimated empirically by comparison with independent data. Viewed in a Bayesian context, these methods report the maximum *a posteriori* probability, but not the associated probability distribution. Uncertainty, when reported, is subject to the assumption of a local Gaussian approximation based on an approximately computed Hessian [e.g. *Tarantola*, 2005, Chapter 3]. However the posterior distribution of an unknown variable often displays richer behavior due to nonlinearities in the governing physics and associated covariance structure. If we are to use fields inferred from model inversion to predict glacier evolution, then we must know whether two equally likely instances of these fields can produce qualitatively different conclusions.

Monte Carlo sampling techniques can provide distributions of model parameters. Several examples of the application of Monte Carlo techniques to glaciological problems exist. *Petra et al.* [2013] developed a Markov Chain Monte Carlo method that generated samples from the posterior basal traction distribution using an efficiently computed approximation to the true Hessian to steer the sampling algorithm. *Chandler et al.* [2006] randomly perturbed measured surface velocities of Glacier de Tsanfleuron and repeatedly inverted for the associated basal properties. *Colgan et al.* [2012] used a Monte Carlo technique to characterize the retreat regime of Columbia Glacier over a wide range of unknown model input parameters, using a heuristic filter to eliminate improbable simulations. All of these cases produced samples from probability distributions of unobserved glaciological variables without the assumption of linearization around a fixed point.

In this paper, we use the Metropolis-Hastings (MH) algorithm [*Hastings*, 1970] to construct probability distributions of ice thickness (and hence bed elevation when surface elevation is know), subject to observations and prior estimates of surface velocity, specific surface mass balance rates (hereafter referred to simply as mass balance), and surface elevation change, as well as sparse pointwise observations of thickness. This situation is common with the recent availability of spatially distributed climate model output and satellite-derived velocity fields. We assume that depth-averaged velocity, thickness, and apparent mass balance (defined here as mass balance minus surface rate of change, *Farinotti et al.* [2009]) are related through mass conservation, and that depth averaged velocities relate to surface velocities in a known albeit unobserved fashion. Each dataset is subject to an assumed covariance structure. This is an analogous problem as that of *Brinkerhoff and Johnson* [2015], *McNabb et al.* [2012], and *Morlighem et al.* [2014a], using mass conservation to interpolate ice thickness estimates while satisfying observational and physical constraints. Contrary to those works, we view the problem from a Bayesian perspective, which allows construction of the full posterior probability density for each model variable. It also serves as an example of how to propagate uncertainty through non-linear glaciological models.

We apply our method to three cases. We begin by considering a synthetically generated glacier, where the simulated velocity, mass balance, and thickness measurements (or prior estimates) are corrupted with noise, and the thickness field is recovered under several assumptions regarding covariance and measurement spacing. We then examine the degree of uncertainty induced by the choice of error structure and physical properties of the glacier. Next, we apply the method to Storglaciären, a ∼3 km long alpine glacier where dense measurements of thickness, velocity, and mass balance are available. This provides an interesting test case with which to gauge the uncertainties induced in topographic estimates due to real-world uncertainties. Finally, we apply the method to Jakobshavn Isbræ, the largest outlet glacier on the Greenland Ice Sheet, and assess resulting uncertainty estimates in the context of making a recommendation for flightline spacing during future airborne radar campaigns.

3.2 Methods

Bayes' theorem states that

$$P(\mathbf{m}|\hat{\mathbf{d}}) \propto P(\hat{\mathbf{d}}|\mathbf{m})P(\mathbf{m}), \tag{3.1}$$

where $\mathbf{m} \in \mathcal{R}^m$ is a vector of unobserved model parameters, and $\hat{\mathbf{d}} \in \mathcal{R}^n$ is a vector of observed model outputs [e.g. *Tarantola*, 2005]. $P(\cdot)$ is the probability density function, a quantification of possible parameter values. Bayes' theorem provides a means to formulate the posterior distribution $P(\mathbf{m}|\hat{\mathbf{d}})$ by considering new data, and it is from this distribution we may draw conclusions.

Construction of the posterior requires two components. First, the likelihood $P(\hat{\mathbf{d}}|\mathbf{m})$ characterizes the probability of observing a realization of $\hat{\mathbf{d}}$ given $\mathbf{m}$. Evaluation of this term requires a solution to the forward model. Second, the prior model $P(\mathbf{m})$ is the supposed distribution of model parameters prior to consideration of data, including assumptions about the mean, covariance, and bounds.

All inverse problems incorporate prior information in one form or another (smoothness for example), which is necessary because of the ill-posedness of such problems. One particular advantage of Bayesian methods is that these prior assumptions, which are often vacuously defined in other inverse methodologies, are defined precisely here and are subject to scrutiny. Another is that changes in assumptions and the addition of new information are easily incorporated by changing the definitions of likelihood and prior without any structural changes in the inference procedure.

3.2.1 Observation Process

We define data to be observed quantities that have not been considered when forming the prior distribution. We denote data with a hat. Data may include observations of surface speed $\hat{\mathbf{U}}_s$, specific surface mass balance $\hat{\dot{\mathbf{b}}}$, thickness $\hat{\mathbf{H}}$, surface elevation rate of change $\widehat{\mathbf{\frac{\Delta S}{\Delta t}}}$, and (map plane) flow direction $\hat{\mathbf{N}}$. These components form the data vector

$$\hat{\mathbf{d}} = \begin{bmatrix} \hat{\mathbf{U}}_s \\ \hat{\dot{\mathbf{b}}} \\ \hat{\mathbf{H}} \\ \hat{\mathbf{N}} \\ \widehat{\mathbf{\frac{\Delta S}{\Delta t}}} \end{bmatrix}. \tag{3.2}$$

Bold indicates a vector: each observation may be available at any number of locations any number of times. We assume that each entry in $\hat{\mathbf{d}}$ is a random variable drawn from a distribution about a true mean value. The realization of such a variable is called the observation process. The observation process takes the form

$$\hat{\mathbf{d}} \sim \mathcal{F}(\mathbf{m}, \Sigma_{\mathbf{d}}), \tag{3.3}$$

where $\mathbf{m}$ is the vector of model parameters upon which the observation process $\mathcal{F}$ operates and $\Sigma_{\mathbf{d}}$ is a parameterization of observational uncertainty induced by the observation process. It includes but is not limited to measurement error. It may be the case that some of the subvectors in $\hat{\mathbf{d}}$ are empty, and in this case the distribution of that function is determined by the prior and its relationship to other parameters through the forward model. This is not to say that no observations have been involved; sometimes they have already been used in a different model to formulate a better prior.

3.2.2 Forward Model

While we assume that the uncertainties $\Sigma_{\mathbf{d}}$ induced by the observation process are approximately independent (i.e. observational errors at neighboring locations are uncorrelated), the model parameters are not. Thickness $H(\mathbf{x})$, mass balance $\dot{b}(\mathbf{x})$, and flow direction $\mathbf{N}(\mathbf{x})$ specify the depth-averaged speed $\bar{U}(\mathbf{x})$ through the continuity equation

$$\partial_t H(\mathbf{x}) \approx \partial_t S(\mathbf{x}) = -\nabla \cdot \left[\bar{U}(\mathbf{x}) H(\mathbf{x}) \mathbf{N}(\mathbf{x})\right] + \dot{b}(\mathbf{x}), \tag{3.4}$$

where $\nabla \cdot$ refers to the map plane divergence. We assume that basal melt is negligible. Also, we have used the approximation that thickness rate of change $\partial_t H$ is well approximated by the surface rate of change $\partial_t S$, which makes the assumption that bed elevations are stationary. This is often a good assumption, though not in regions experiencing rapid subglacial erosion [*Motyka et al.*, 2006].

Eq. (3.4) is valid only at an instant in time. On the contrary, data are observed over finite time intervals. The length of these intervals may vary between observables, and are often not aligned with one another (e.g. velocity may have been observed over the entire winter of 2008, while thickness observations have been taken almost instantaneously during the summer of 2005). Observations may thus be far from the cotemporal and instantaneous values required to make Eq. (3.4) valid. This incongruity in and between the characteristic time scales of the observation and model processes is responsible for an additional source of error that must be accounted for.

One way of constructing a valid mass conservation relationship over the entire observation period is to time integrate the forward model over the range $t \in [t_0, t_1]$, where t_0 and t_1 are the starting and ending times over which an estimate of time-averaged quantities are available. Integrating and dividing Eq. (3.4) by the interval $\Delta t = t_1 - t_0$ yields

$$\frac{\Delta S}{\Delta t} + \frac{1}{\Delta t} \int_{t_0}^{t_1} \left(\nabla \cdot \left[\bar{U} \mathbf{N} H \right] - \dot{b} \right) \mathrm{d}t = 0, \tag{3.5}$$

the time average over the observation period. Replacing each term with its average yields an equation similar to Eq. (3.4), but with a reinterpretation of observational uncertainty: not only imprecision in direct measurement and processing, but also the departure of the observation from the time average. We write this as

$$\Sigma_{\mathbf{d}} = \Sigma_{obs} + \Sigma_t, \tag{3.6}$$

where Σ_{obs} is uncertainty due to measurement, and Σ_t is the uncertainty due to measurements not being those of direct relevance to the forward model. s Velocity observations are only available at the surface, while the forward model Eq. (3.4) requires depth-averaged quantities. We assume that surface flow directions $\mathbf{N}$ are a good approximation of flow direction at depth. However, this is not always true for magnitudes and we introduce a spatially variable multiplicative factor $s(\mathbf{x})$ that serves to transfer between depth-averaged speed and surface speed

$$U_s(\mathbf{x}) = s(\mathbf{x}) \bar{U}(\mathbf{x}). \tag{3.7}$$

3.2.3 Model Simplification

Monte Carlo methods such as the MH algorithm are computationally expensive because the likelihood (and thus the forward model) must be evaluated many times. Fortunately, as long as $\mathbf{N}(\mathbf{x})$ is constant with depth, the mass conservation model is purely advective and we can exploit the independence of flow bands [*McNabb et al.*, 2012]. If we select two streamlines from the velocity field, the domain contained between them is independent of any other non-overlapping domain. If they are close, we can approximate the parameter variability transverse to flow with the transverse average, and the problem reduces from two dimensions to one. Neglecting curvature effects, the continuity equation becomes

$$\partial_r \left[w(r) \, \bar{U}(r) \, H(r) \right] = w(r) \left(\dot{b}(r) - \frac{\Delta S(r)}{\Delta t} \right), \tag{3.8}$$

where r is the along-flow coordinate, and $w(r)$ is the width of the streamline taken normal to the centerline. Parameters should again be viewed as averages over the temporal footprint of the observations. Note that the flow direction $\mathbf{N}(\mathbf{x})$ no longer enters the equation, since it has been subsumed by the width, which has its own observational uncertainty and prior. The model parameter vector is

$$\mathbf{m} = \begin{bmatrix} U(r; H(r), \dot{b}(r), \frac{\Delta S(r)}{\Delta t}) \\ H(r) \\ \dot{b}(r) \\ \frac{\Delta S(r)}{\Delta t} \\ s(r) \\ w(r) \end{bmatrix}. \tag{3.9}$$

This one dimensional formulation is efficient because it can be solved by integration:

$$\bar{U}(r) = \frac{1}{w(r)H(r)} \int_0^r w(r') \left(\dot{b} - \frac{\Delta S(r)}{\Delta t}\right) \mathrm{d}r', \tag{3.10}$$

which we evaluate with trapezoidal quadrature [e.g. *Atkinson*, 1978].

Non-dimensionalization of Eq. (3.10) produces an interesting result: the model depends on only a single non-dimensional parameter,

$$\gamma = \frac{L\widetilde{b}}{\widetilde{U}\widetilde{H}}, \tag{3.11}$$

where $\widetilde{b}$, $\widetilde{U}$, and $\widetilde{H}$ respectively refer to characteristic mass balance, velocity, and thickness, and L the length of the glacier, and thus it should be understood that results here apply similarly to differently scaled glaciers so long as γ remains constant.

3.2.4 Priors

The specification of prior distributions on model parameters depends on the variable and the problem being considered. However a universal requirement is that priors be chosen before considering any of the observations contained in $\hat{\mathbf{d}}$.

We assume no prior knowledge of the ice thickness $H(r)$ besides non-negativity and that it possesses some smoothness, and its prior should reflect these properties. This implies some knowledge of the covariance structure of ice thickness. A useful and general representation of $H(r)$ is as a Gaussian process (GP) [*Rasmussen*, 2006] with arbitrary mean and large variance, but with a specified covariance structure

$$H(r) \sim \mathrm{GP}(\cdot, \nu_H(r, r')), \tag{3.12}$$

where $\nu_H(r, r')$ is a spatial covariance function. The placeholder $\cdot$ indicates that the mean value should be arbitrary; the prior is sufficiently vague that the value of the mean does not affect the posterior distribution. We modify the Gaussian process so that negative values of H have zero probability. We use either the following Gaussian covariance function

$$\nu_i^G(r,r') = \sigma_i^2 \exp\left(-\frac{(r-r')^2}{l_i^2}\right) \tag{3.13}$$

or exponential covariance function

$$\nu_i^E(r,r') = \sigma_i^2 \exp\left(-\frac{|r-r'|}{l_i}\right), \tag{3.14}$$

for the remainder of this paper, where σ_i^2 is the prior variance, $r - r'$ is the pointwise distance between any two points, and l_i is the correlation length scale. Note that for $|r - r'| > 3l_i$, values no longer exhibit significant correlation [*Rasmussen*, 2006].

It is often preferable to specify a mass balance distribution in which the data have already been reanalyzed with a separate climate model or interpolated with some other method since these are more advanced than one we could include. Here we consider the estimated posterior distribution of a climate model to be the prior on mass balance, since observations have already been included. We assume that the mass balance prior is also a Gaussian process

$$\dot{b}(r) \sim \mathrm{GP}(\dot{b}_{\mathrm{prior}}(r), \nu_{\dot{b}}^G(r,r')). \tag{3.15}$$

We assume that the width function is a Gaussian process with a mean value given by the computed width,

$$w(r) \sim \mathrm{GP}(w_{\mathrm{prior}}(r), \nu_w^G(r,r')). \tag{3.16}$$

The use of a Gaussian process as a prior is the same key assumption as used in Kriging [*Williams*, 1998], and this technique should be seen as Kriging with additional information introduced to the likelihood function by continuity. Where no observations exist, the algorithm reverts to ordinary Kriging.

Once again, surface velocities and depth-averaged velocities are only equal when sliding accounts for all glacier movement. The other end-member, pure deformation, bounds the value of surface velocity at $U_s(r) = \frac{n+2}{n+1}\bar{U}(r)$, where $n = 3$ is Glen's flow law exponent [*Glen*, 1955]. This usually holds in polythermal glaciers, where the more complex rheological structure concentrates strain at the glacier base, and makes the depth averaged velocity closer to the surface velocity. We typically have no other *a priori* information about the sliding proportion s, and thus model it as a uniform distribution with the bounds given by the above argument

$$s(r) \sim \mathrm{Unif}\left(1, \frac{n+2}{n+1}\right). \tag{3.17}$$

Figure 3.1. Pseudocode describing the Metropolis-Hastings algorithm. Note that $q(\cdot|\cdot)$ is the proposal distribution, and $\bar{P}(\cdot|\mathbf{d})$ is the product of the likelihood and the prior evaluated at a point. Evaluation of the likelihood requires solution of the forward model.

```
Choose initial parameter values $\mathbf{m}_0$
for $i = 0, \ldots, N-1$ do
    Draw proposal sample $\mathbf{m}' \sim q(\mathbf{m}'|\mathbf{m}_i)$
    $\alpha_i \leftarrow \min\left[1, \frac{\bar{P}(\mathbf{m}'|\hat{\mathbf{d}})q(\mathbf{m}_i|\mathbf{m}')}{\bar{P}(\mathbf{m}_i|\hat{\mathbf{d}})q(\mathbf{m}'|\mathbf{m}_i)}\right]$
    Sample $u \sim \mathrm{Unif}(0,1)$
    if $u < \alpha_i$ then
        Set $\mathbf{m}_{i+1} = \mathbf{m}'$
    else
        Set $\mathbf{m}_{i+1} = \mathbf{m}_i$
    end if
end for
```

3.2.5 Metropolis-Hastings algorithm

The posterior distribution has no closed form and must be characterized by sampling. With samples in hand we can evaluate their statistical properties as a proxy for the posterior distribution. Our method of choice for this procedure is the Adaptive Step Metropolis-Hastings algorithm [*Hastings*, 1970]. Pseudocode for the algorithm is given in Algorithm 1.

The MH algorithm operates by travelling through parameter space according to steps drawn from a proposal distribution, in this simple case an independent multivariate normal centered around the current point. If the posterior probability is greater at the proposed point than at the current point, then the proposed step is accepted, and the algorithm continues from the new point. If the likelihood is lower at the proposed point, then a step is taken with probability equal to the ratio of the current and proposed points. In the adaptive step variant of the MH algorithm, the width of the proposal distribution is adjusted such that an optimal proportion of proposals are accepted. The algorithm is ergodic [*Hastings*, 1970], and after a sufficient number of steps, the samples converge to a set drawn from the posterior distribution. MH sampling is performed using the python package PyMC [*Patil et al.*, 2010].

Table 1. Table of relevant constants for the experiment outlined in Sections 3.3.1 and 3.3.2.

Parameter	Value	Units	Description
L	15	km	Domain length
z_{min}	0	m	Minimum elevation
z_{max}	1500	m	Maximum linear elevation
A_s	200	m	Sinusoidal topography amplitude
A_e	500	m	Headwall amplitude
L_e	700	m	Headwall decay length
σ_T	50	m	Random variability amplitude
l_T	250	m	Topographic correlation length
$\dot{b}_{min}$	-80	ma^{-1}	Minimum specific balance
$\dot{b}_{max}$	10	ma^{-1}	Maximum specific balance
c	3		Shape factor
l_H	250	m	Thickness correlation length
$l_{\dot{b}}$	1500	m	Mass balance correlation length
p_H	0.01		Relative thickness uncertainty factor
p_U	0.1		Relative speed uncertainty factor
$\sigma_{\dot{b}}$	2	ma^{-1}	Specific balance uncertainty
$\sigma_{u,min}$	10	ma^{-1}	Minimum velocity uncertainty
n_d	{11,6,3}		Number of data points
n_c	100		Number of grid cells

3.3 Synthetic glacier

3.3.1 Synthetic glacier generation

We used the finite element ice sheet model VarGlaS [*Brinkerhoff and Johnson*, 2013] to generate a synthetic steady state glacier in vertical profile. We neglect effects due to changing width and side drag. The glacier does not slide. The basal topography is given by

$$\begin{aligned} B(r) = {} & \frac{z_{min} - z_{max}}{L} r - A_s \sin \frac{2\pi r}{L} \\ & + A_e \exp - \frac{r + L}{L_e} + G(r), \end{aligned} \tag{3.18}$$

where z_{min} and z_{max} are minimum and maximum elevations, L is the domain length scale, A_s is the amplitude of a sinusoidal variability, A_e is the amplitude of an exponential term (simulating a steep headwall), L_e the decay rate of the exponential perturbation centered about the bergschrund, and $G(r)$ is a random topographic perturbation given by

$$G(r) \sim GP(0, \nu_T^G(r, r')). \tag{3.19}$$

Mass balance is an exponential function of surface elevation.

$$\begin{aligned} \dot{b}(S(r)) = \dot{b}_{min} + & \frac{\dot{b}_{max} - \dot{b}_{min}}{1 - \exp(-c)} \\ & \times \left[1 - \exp\left(-c\frac{S(r) - z_{min}}{z_{max} - z_{min}}\right)\right], \end{aligned} \tag{3.20}$$

where $\dot{b}_{min}$ and $\dot{b}_{max}$ are the mass balance minimum and maximum respectively, and c is a shape parameter. While the scale of the geometry will not be relevant to the results contained herein (q.v. Section 3.2.3), we performed the computations in dimensional form because it was easier to define reasonable parameter values that way. The parameters used are shown in Table 1, and correspond to a moderate-size maritime mountain glacier. The model was discretized over $n_c = 100$ equally spaced grid cells and time integrated until reaching a steady state. We assumed that surface elevation was stationary and known exactly.

3.3.2 Recovery of synthetic topography

As a first experiment, we used the MH algorithm to sample $\mathbf{m}$ under assumptions of uncertainty and data spacing that might correspond to a typical mountain glacier. First, we simulated observations of surface velocity by corrupting modeled velocity with uncorrelated Gaussian random noise with standard deviation $\sigma_{u,i} = \max(p_u U_{m,i}, \sigma_{u,\text{min}})$. Thickness measurements were assumed available at $\Delta d = L/n_d$ increments. Uncertainties in thickness measurements were assumed to be $\sigma_{H,i} = p_H H_i$. The above values were chosen not only to account for hypothetical instrument error, but also to simulate the deviation of measurements from the average required by Eq. 3.5. As discussed above, we assumed that mass balance observations had already been used to update the prior model.

We assumed that the error structure was known, and that the likelihood model was given by

$$\hat{\mathbf{d}} \sim \mathcal{N}(\mathbf{d}, \Sigma_{\mathbf{d}}), \tag{3.21}$$

where $\mathcal{N}(\mathbf{d}, \Sigma_{\mathbf{d}})$ is a multivariate normal distribution with mean $\mathbf{d}$ and covariance $\Sigma_{\mathbf{d}}$. The vector $\mathbf{d}$ contains model variables evaluated at observation points. We take $\Sigma_{\mathbf{d}}$ to be a diagonal matrix with entries given by the corresponding observational uncertainties stated above.

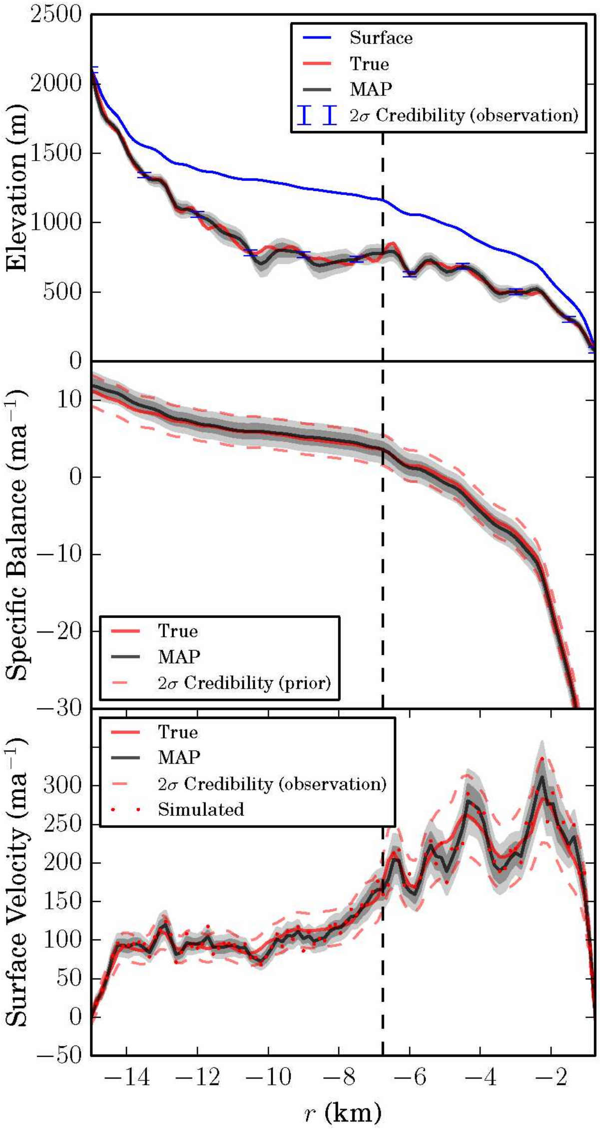

Figure 3.2. Recovered pointwise probability densities for a synthetic glacier. Dark and light gray shaded regions indicate the σ and 2σ posterior credibility interval respectively, and MAP refers to the solution corresponding to the maximum *a posteriori* probability. The dashed vertical line is the location at which the histograms in Figure 3.3 are computed.

We assumed no prior knowledge of the actual thickness values but that the thickness length scale was known, or that $l_H = l_T$, and that an informative prior model of mass balance with known covariance was available. Covariance amplitude was given by $\sigma_{\dot{b}} = r_{\dot{b}}\max(\dot{b})$ and $\nu^G_{\dot{b}}$ was Gaussian with correlation length $l_{\dot{b}}$.

Using the MH algorithm, we drew $n_i = 10^6$ iterations from the posterior distribution of $\mathbf{m}$, discarding the first 10^5 samples to eliminate transient behavior. We performed this procedure three times, using different (random) initial conditions for each sample. Evidence for the convergence of the samples to a stationary and correct posterior distribution will be demonstrated in Section 3.3.3.

Figure 3.2 shows the pointwise posterior distributions of the model parameters in $\mathbf{m}$. The most notable immediate result is that the algorithm produces correct credibility intervals: the 2σ credibility interval does indeed contain the true bed elevation at 95% of grid points, which provides confidence that the posterior distribution produced by the algorithm can recover meaningful information about the bed elevation.

Nonetheless, the maximum *a posteriori* prediction of bed elevations is not the true value. This is not a surprising result, since we assume a single measurement of velocity for a given location with which to constrain the mean of the velocity distribution there (though multiple observations at a point can be included naturally, see Section 3.4.1). The mean velocity is thus free to assume values in the space around the observation, but the most probable value of the velocity mean given one observation per grid point and in the absence of feedbacks from additional constraints on the velocity is the data point itself.

An interesting feature of the posterior distributions of both mass balance and surface velocity is that the posterior distribution is more specific than the prior. This implies that not only are these fields contributing information to the estimation of thickness, but data and smoothness constraints on the thickness also feed back.

3.3.3 Convergence

There are several mechanisms for assessing whether a distribution has become stationary, some heuristic and some quantitative. For a good approximation to the posterior distribution to be achieved, a) the sampler must traverse the support of the sampled function many times, b) the sampler must visit the entire support, and c) the region traversed by the sampler should be insensitive to initial conditions.

Examining traces (i.e. the history of parameter values at each sampler iteration) gives a heuris-

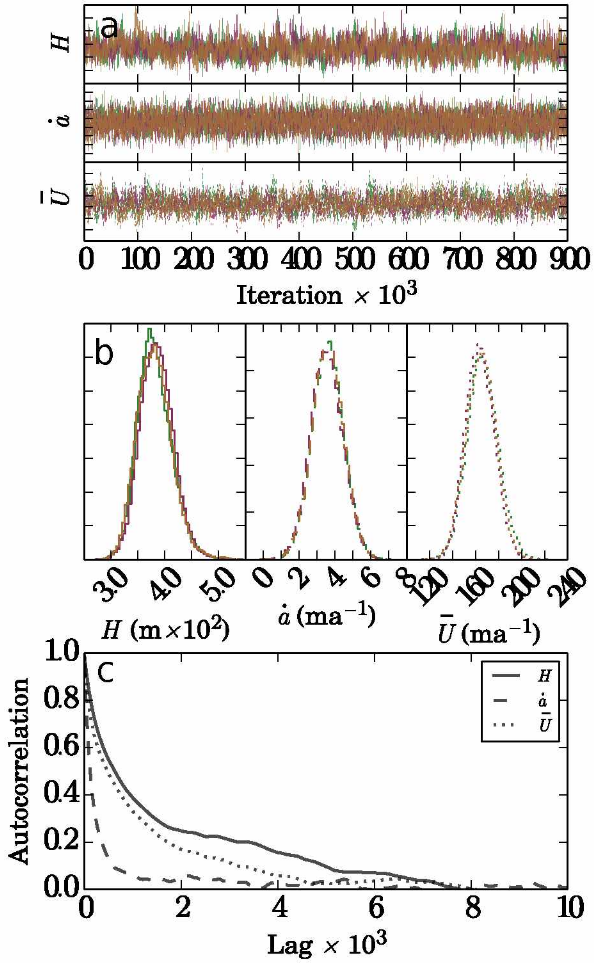

Figure 3.3. Convergence metrics for the grid point denoted by a vertical dashed line in Figure 3.2. (a) The trace for each sampler and for each variable. The traces indicate full exploration of the parameter space. (b) histograms of each variable for each sampler. (c) The auto-correlation for thickness H, mass balance $\dot{b}$, and speed U_s. For example, samples for thickness H become uncorrelated after approximately 8000 iterations.

tic means to assess convergence. Figure 3.3a shows the traces of the model vector $\mathbf{m}$ at the location indicated by the vertical dashed line in Figure 3.2. These plots are similar at any location. The 'fuzzy' pattern is an indication that the samples are well-mixed, which is to say that both criteria a) and c) are satisfied. While it is difficult to state definitively that there is not a distant probability maximum that is not being captured, this is unlikely due to physical intuition and the wide dispersion of initial conditions between samplers.

Figure 3.3b shows the posterior probability for several model variables. The densities associated with each sampler exhibit a high degree of similarity, which is further evidence that the samples produced by the MH algorithm have converged to the stationary posterior distribution.

The Gelman-Rubin statistic [*Gelman and Rubin*, 1992] provides a quantitative convergence statistic. This statistic compares the ratio between the interchain variance

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\mathbf{m}}_j - \bar{\bar{\mathbf{m}}})^2, \tag{3.22}$$

and the within-chain variance

$$W = \frac{1}{m} \sum_{j=1}^{m} \left[\frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{m}_{ij} - \bar{\mathbf{m}}_j)^2 \right], \tag{3.23}$$

where we have m chains each of length n, $\bar{\mathbf{m}}_j$ is the mean of chain j, and $\bar{\bar{\mathbf{m}}}$ is the mean of all chains. The marginal posterior estimate of the variance of $\mathbf{m}$ can be estimated by

$$\widehat{\mathrm{Var}}(\mathbf{m}|\hat{\mathbf{d}}) = \frac{n-1}{n} W + \frac{1}{n} B. \tag{3.24}$$

This quantity always overestimates the true value of the marginal variance. Simultaneously, W underestimates the within-chain variance for an underconverged chain. Thus, in the limit as $n \to \infty$, the ratio of these two quantities

$$R = \frac{\widehat{\mathrm{Var}}(\mathbf{m}|\hat{\mathbf{d}})}{W} \tag{3.25}$$

converges to one. Thus, for a fully converged distribution (i.e. one that exhibits behavior similar to the limiting case), $R \approx 1$. In practice, $R < 1.1$ is acceptable [*Gelman and Rubin*, 1992]. We find that for each component in $\mathbf{m}$, $R \ll 1.1$, providing evidence that the samples are drawn from the stationary distribution.

Figure 3.3c addresses a final numerical consideration. The MH algorithm produces samples that are autocorrelated, which can persist when parameters covary (as is usually the case with spatial processes). While this auto-correlation is not fatal to the algorithm's performance, it provides an important reminder to run each chain long enough to obtain unbiased sample statistics [*Christensen*

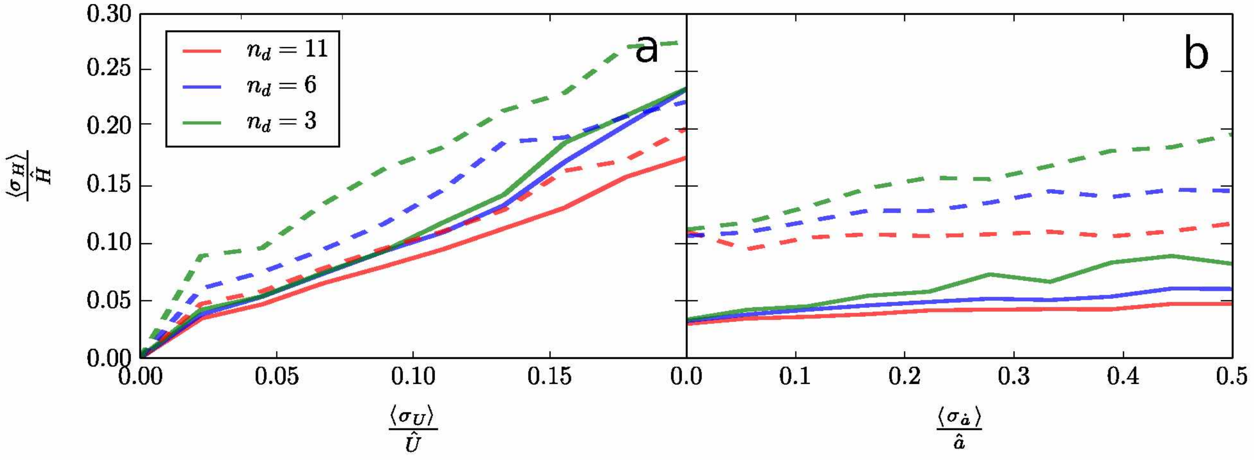

Figure 3.4. Relationship between average thickness uncertainty and velocity uncertainty (a) and mass balance uncertainty (b). Solid lines reflect the case where the other variable is being held constant at a low level of uncertainty (e.g. mass balance held fixed while velocity is varied), while dashed lines represent the case where the other variable is held constant at a high uncertainty. These curves are computed for three different data densities.

et al., 2011; *Link and Eaton*, 2012]. The Gelman-Rubin statistic suggests that the MH algorithm has been run long enough to overcome the difficulties due to autocorrelation.

3.3.4 Uncertainty Propagation

The relative uncertainty with which the ice thickness can be recovered using mass conservation methods is a function of the relative uncertainties and covariance structures of surface velocity measurements, mass balance measurements (or assimilated model output), ice thickness measurements, and ice thickness measurement density. It is reasonable to suspect that an improvement in any of these factors would lead to a commensurate improvement in posterior predictive capabilities with respect to thickness. Nonetheless, it is not clear which of these factors contributes the lion's share of posterior variance. This information is key in forming plans of additional data acquisition; we would like to know which data improvements (and at what densities) will most improve thickness estimates.

Here we examine the influence that surface velocity uncertainty, mass balance uncertainty, and measurement density have on the relative thickness uncertainty (quantified here as $\langle \sigma_H \rangle / \tilde{H}$). We neglect to consider sensitivity to uncertainty in thickness measurements. However, these uncertainties may well be important in many cases, particularly if the assumption of on-nadir bed returns is flawed or if these exhibit systematic errors. The assumption of a normally distributed error structure may also be inappropriate. Nonetheless, we proceed without a detailed quantification of these effects because a) the quantification of radar uncertainty is not the focus of this paper and b) we do not know the most appropriate way to proceed. However, it would be straightforward to include such a detailed uncertainty estimate within the framework presented here.

In order to assess the functional relationship between these uncertainties, we performed simulations over an array of 10 equally spaced relative velocity uncertainties between 0 and 20% of the true maximum surface velocity, 10 relative mass balance uncertainties between 0 and 50% of the true maximum mass balance, and the number of data points $n_d \in \{11, 6, 3\}$. Note that the last of these data spacings $n_d = 3$ corresponds to a thickness measurement at the midpoint of the glacier, along with the constraints of zero thickness at the glacier head and terminus. This corresponds to 300 simulations, each of which was run to convergence.

Figure 3.4a shows the mean standard deviation in thickness (normalized by max thickness) as a function of the mean velocity standard deviation (normalized by maximum velocity). Solid lines correspond to the case where the mass balance is known very precisely ($\sigma_{\dot{b}} = 0$), while the dashed lines represent an intermediate uncertainty in mass balance ($\langle\sigma_{\dot{b}}\rangle/\tilde{b} = 0.25$). Thickness and velocity uncertainties exhibit a linear relationship. The proportionality depends weakly on data density because of the lack of a smoothness constraint directly imposed on velocity; it is strictly a function of thickness and mass balance and only indirectly sees the locations at which thickness is known. Where mass balance is known precisely, the $n_d = 3$ and $n_d = 6$ solutions are similar. This is because the variance in velocities admissible with respect to forward model constraints when only a single measurement of thickness exists is less than the variance in velocity due to measurement uncertainty. When mass balance has large uncertainty, the forward model imposes a weaker constraint and the distribution of velocities becomes progressively wider as observational uncertainty increases.

Thickness precision as a function of mass balance uncertainty (Figure 3.4b) is also linear. However, this relationship depends more on data density due to the long correlation length imposed upon mass balance. While it can assume many values between thickness measurements, the local constraint imposed by mass conservation at thickness measurement locations forces it to a unique value there. If the measurement spacing is shorter than the mass balance correlation length, it cannot vary much between these two pinned points, regardless of its variance. This constraint becomes less active as measurement spacing increases, allowing the uncertainty in mass balance to more strongly influence uncertainty in the thickness.

3.4 Storglaciären

3.4.1 Characterization and data.

Storglaciären is a $\sim$3 km long polythermal glacier in northwestern Sweden. It possesses the longest spatially distributed surface mass balance record of any glacier [*Holmlund et al.*, 2005]. The relative simplicity of its geometry along with the density of mass balance data make it a useful test case. Also, the basal topography of Storglaciären is well known and thus this experiment provides a means to assess whether the algorithm correctly produces error bounds on a 'known' bed.

Thickness and surface observations were adapted from *Herzfeld et al.* [1993]. Surface velocities were derived from stake measurements [*Hooke et al.*, 1989; *Jansson*, 1997; *Kuriger*, 2002]. These observations are discrete rather than fields. However, this distinction from the previous synthetic example provides no particular difficulty from a technical standpoint.

We estimated the flowline width by using a first order glacier flow model [*Brinkerhoff and Johnson*, 2013] in the full map-plane domain to invert point measurements of surface speed for basal traction. We used the resulting velocity field to compute a flowband with centerline coordinates that passed approximately through the velocity observation points. We did not use this modeled velocity field as an observation. Instead, to reduce circularity, we used the original measurements.

A prior distribution on mass balance was generated from the measured annual net balance for Storglaciären between years 2000 and 2013 [*Jansson*, 1999]. The availability of multiple years of data allowed the computation of both the sample mean and the estimated covariance of the mean as a function of surface elevation. A spatial covariance function $\nu_{\dot{b}}^{G}(S(r), S(r'))$ was computed by fitting the sample mean covariance matrix with a Gaussian covariance function, similar to Equation (3.14). This covariance function was parameterized as a function of surface elevation.

Thickness measurement uncertainty was assumed to be $\sigma_H = 10$ m, and the thickness field was assigned a vague prior as in the synthetic case. We assumed a correlation length of $l_H = 250$ m, a velocity standard deviation of $\sigma_u = 5$ ma^{-1}, and a width uncertainty of $\sigma_w = 0.1$. The Storglaciären flowline was discretized with a horizontal resolution of approximately 35 m. Despite profile data being given as continuous, we artificially sampled the thickness every 350 m.

3.4.2 Recovery of known topography.

We ran the Metropolis-Hastings algorithm three times for 10^6 iterations each, discarding the first 10^5 samples. Each instance was given different initial values drawn from the prior distribution. We assessed the convergence of the samples using the same methods as discussed in Section 3.3.3. This

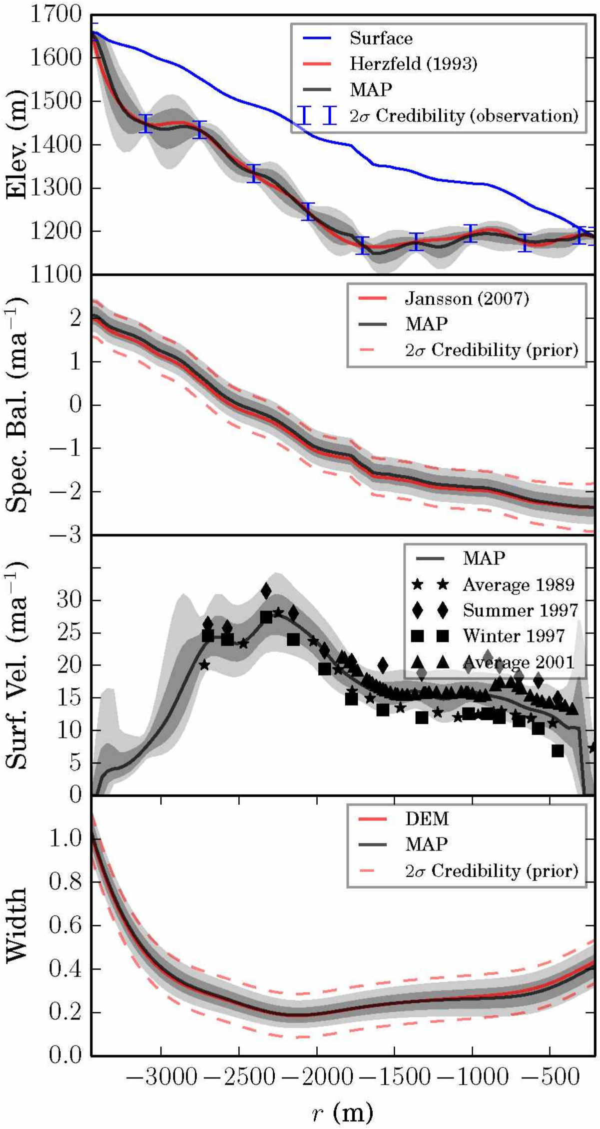

Figure 3.5. Recovered pointwise probability densities for Storglaciären. Dark and light gray shaded regions indicate the σ and 2σ posterior credibility interval, respectively.

analysis produced similar results to those of the synthetic case (not shown).

Figure 3.5 shows the computed posterior distributions and priors. The credibility intervals provide a correct if conservative estimate of the true value. However, it is not clear that in the presence of a large number of data points and a smooth bed (as for Storglaciären), that much is to be gained by using a mass conservation based interpolant. Rather, ordinary Kriging or some variant thereof would be equally useful in establishing a bed elevation model for dynamic modeling. This is less an indictment of mass conservation methods and more a reflection of the fact that Storgläciaren's simple topography, smooth bed, and high data density make for an ideal case for Kriging. Despite the fact that Storglaciären has a well constrained mass balance, we have already shown that the mass conservation method is insensitive to this quantity in cases where thickness measurements are closely spaced, and this dataset cannot be fully utilized. Simultaneously, surface velocities are discrete, and mass conservation cannot contribute much between observation locations. Nonetheless, the mass conservation technique produces a distribution that would be useful in forcing ensemble runs with dynamic models. Furthermore, it produces an estimate of unobserved surface velocities and a corresponding uncertainty estimate.

3.5 Jakobshavn Isbræ

3.5.1 Characterization and data.

Jakobshavn Isbræ is the largest basin (as ranked by discharge) on the Greenland ice sheet and is the fastest glacier on earth [*Rignot and Kanagaratnam*, 2006]. It drains 7% of the Greenland ice sheet by area. Due to these factors it is comparatively well studied. However, because of the large areal extent the absolute data density compared to Storglaciären (for example) is low.

Absolute mass balance rates over the ice sheet are small compared to a mountain glacier, with a maximum accumulation of around half a meter per year, an order of magnitude lower than a typical mountain glacier. Relative measurement errors are large given identical methods, and the distances over which measurements must be extrapolated are longer. Surface velocities range over four orders of magnitude and relative errors are high in the low velocity interior of the ice sheet and low in the fast flowing outlet regions. Soundings of ice thickness are made by airborne radar in the form of discrete radar flightlines, mostly running normal to the dominant ice flow direction. These can have high observational uncertainty in topographically complex regions [*Gogineni et al.*, 2014].

We consider the time domain $t \in [2003, 2014]$ since it overlaps observations of ice thickness

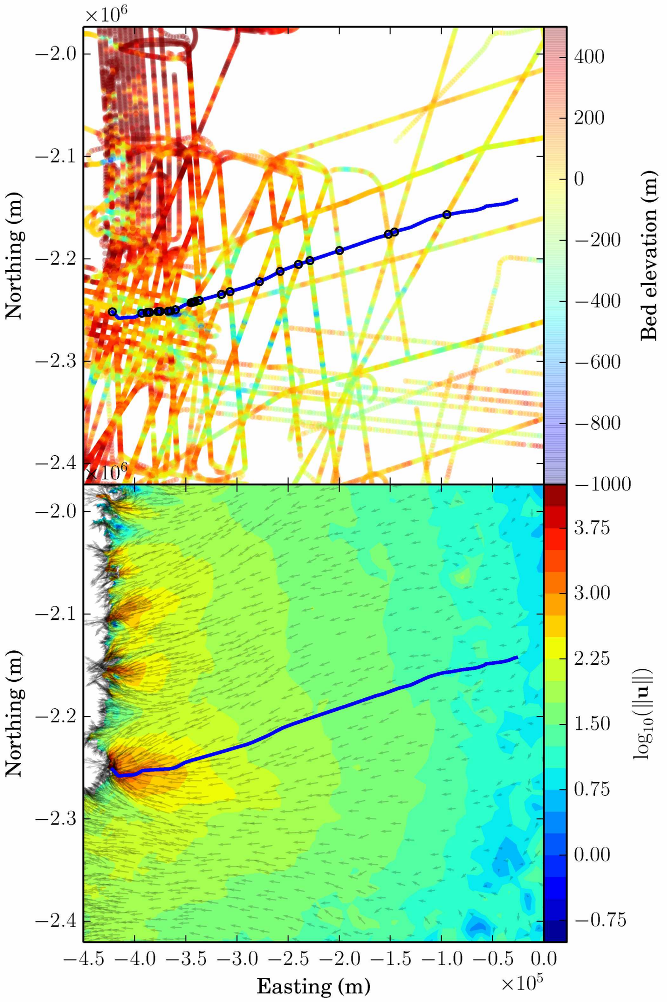

Figure 3.6. Centerline of the flowband considered for Jakobshavn Isbræ(Blue). Also shown are log-scaled velocity vectors from Rignot and Mouginot (2012), and bedrock elevations along Operation IceBridge flightlines derived from MCoRDS ice penetrating radar [*Allen et al.*, 2015].

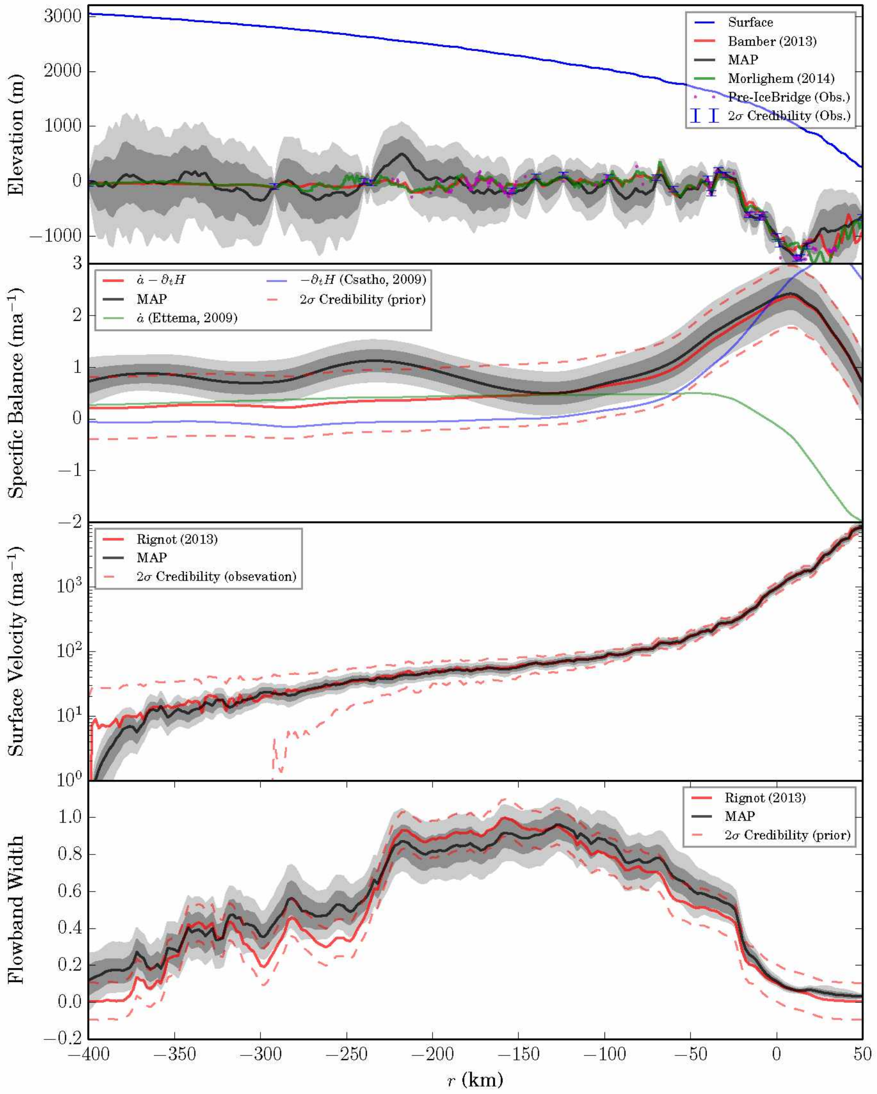

Figure 3.7. Recovered pointwise probability densities for a flowband over Jakobshavn Isbræ. Dark and light gray shaded regions indicate the σ and 2σ posterior credibility interval, respectively. Also included are the digital elevation models of *Bamber et al.* [2013] and *Morlighem et al.* [2014b], as well as pre-Icebridge radar observations that were not included in the solution procedure [*Allen et al.*, 2015].

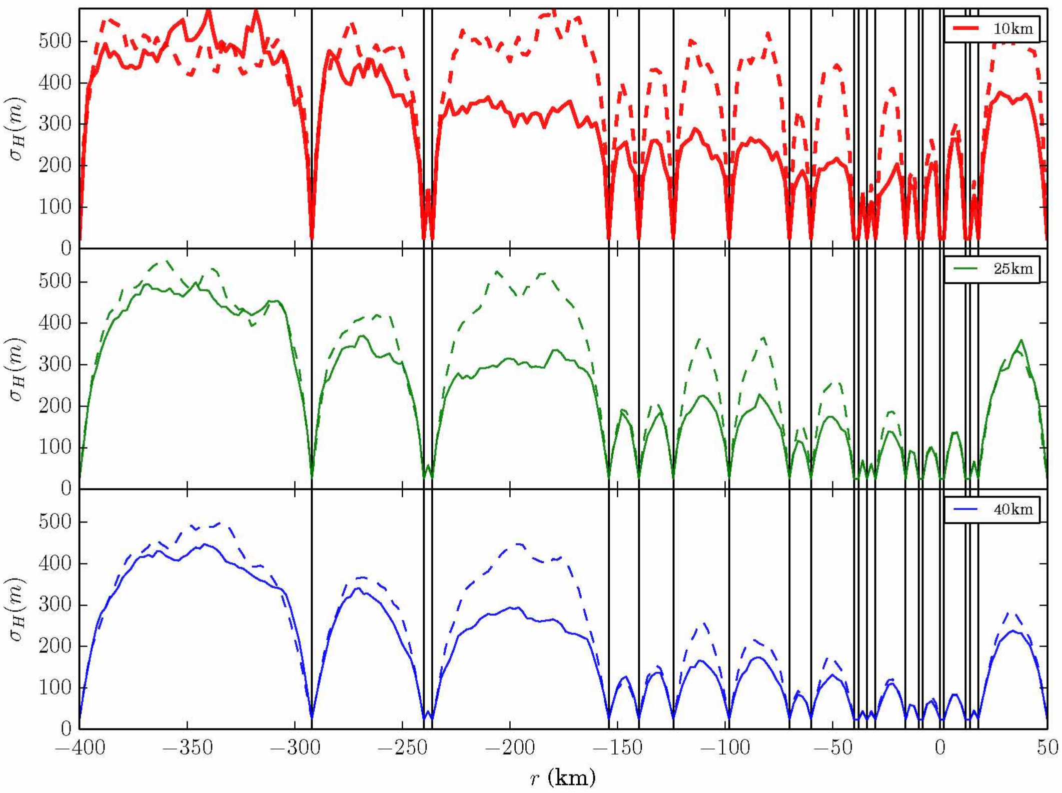

Figure 3.8. Posterior standard deviations of Jakobshavn Isbræ flowband under different assumptions about the correlation length. Flightline locations are denoted by a vertical black line.

[2011–2014, *Allen et al.*, 2015], surface velocity [2008, *Rignot and Mouginot*, 2012], surface rate of change [2003–2007, *Csatho et al.*, 2014], and mass balance [1958–2007, *Ettema et al.*, 2009]. We will describe each of these data sets and our estimate of their uncertainties in this application below.

We extracted a flowband from Jakobshavn Isbræ flowline by generating two streamlines from the InSAR derived velocity dataset of *Rignot and Mouginot* [2012]. The streamlines were spaced approximately 100 m apart at a location approximately 50 km upstream of the 2008 terminus location. The flowline used in this section is the centerline between these two flowlines, and all data are assumed to be width-averaged values contained therein. The width itself is the distance between the two flowlines along a line segment normal to the centerline. We assume a uniform 5% standard deviation in flowline width (i.e. $\sigma_w = 0.05$).

Velocity magnitudes were also derived from *Rignot and Mouginot* [2012]. Reported velocity uncertainties are approximately 10 ma^{-1} due to instrument and processing error. The velocity field spans the single winter of 2008, and as such can account for neither interannual nor intra-annual variability, yet there is a considerable amount of spatial and temporal variability in the basin's velocity field. For example, *Joughin et al.* [2012] showed that in the nearest 40 km to the Jakobshavn Isbræ terminus, velocities could vary by as much as 40% over the course of a year. Additionally, they show a long term trend in velocity over the period between 2004 and 2011. In each case, the degree of temporal variability was spatially heterogeneous, but roughly corresponded to the magnitude of the velocity. Observations did not always demonstrate a consistent seasonal signal and we cannot assume that the winter observation represents an annual velocity minimum.

A simple uncertainty estimate that can account for the factors given above assumes that observations have an uncertainty of $\sigma_U = \max(\hat{U}_s p_U, \sigma_{U,min})$. We take $\sigma_{U,min} = 10$ ma^{-1} as a reflection of the instrument uncertainty, a lower bound on the true uncertainty in annual average. We assume $p_U = 0.1$, which would correspond to a 2σ bound capturing the 40% intra-annual variability seen at some locations. While we believe that these uncertainties are roughly representative, we also acknowledge that there are more precise ways to parameterize this uncertainty, for example computing sample covariance between different velocity datasets and using a non-normal uncertainty distribution to reflect the fact that using solely a winter velocity skews the estimate of the mean velocity downwards.

Mass balance was drawn from the regional climate model RACMO2/GR [*Ettema et al.*, 2009], averaged over a period between 1958 and 2007, which is the longest temporal footprint of the data sets considered here. It lacks a covariance estimate, so we assume an uncertainty of $\sigma_{\dot{b}} = 0.2$ ma^{-1}

that reflects the fundamental uncertainty of the model output due to the long distances between controlling data points, systematic uncertainty in the reanalysis process, and processes such as basal melt and surface refreezing that might not be captured. It also accounts for uncertainty induced as a result of using this 49 year average as a proxy for the 11 year average we consider here. We assume a correlation length of $l_{\dot{b}} = 50$ km.

The Greenland ice sheet is not in steady state [*Csatho et al.*, 2014; *Joughin et al.*, 2012; *Motyka et al.*, 2010]. To account for this, we specify a thickness change field from *Csatho et al.* [2014]. This field is derived from repeat surface altimetry measurements taken between 2003 and 2009. Surface laser altimetry is precise and we thus assume measurement uncertainty in thickness change to be negligible compared to the uncertainty in mass balance estimates. Nonetheless, because the period of record is not contemporaneous with the specified averaging period, we assume that uncertainty in the average thickness rate of change contributes an additional $\sigma_{H_t} = 0.1$ ma^{-1} to uncertainty in apparent mass balance.

We utilize ice thickness data obtained by aerial radar using the Multichannel Coherent Radar Depth Sounder (MCoRDS) [*Allen et al.*, 2015], collected as part of Operation IceBridge, wherever it intersects the flowline. Note that we did not include pre-IceBridge observations with the same instrument, instead saving these for validation purposes. However, if we intended to produce a bed elevation field for further modeling use, we would include these observations when computing the posterior distribution. We assume a nominal standard deviation along the flightlines of 12.5 m, as specified in the data documentation. Furthermore, we keep only measurements that are rated as being of 'high quality.' This latter filter has the effect of eliminating most radar returns in the deep trough near Jakobshavn Isbræ's terminus, where off-nadir reflections make interpretation of radargrams uncertain. We neglect the small uncertainty resulting from thickness changes over the averaging period.

Variograms computed for IceBridge and pre-IceBridge data have placed the average range (the distance at which samples become uncorrelated) for Greenland between 58 km [*Morlighem et al.*, 2013] and 80 km [*Bamber et al.*, 2001] (erroneously referred to as the sill in both). Both of the above studies fit an exponential covariance model [*Rasmussen*, 2006]. We computed spatially explicit variograms over a moving 120 km footprint for all of Greenland. In the Jakobshavn Isbræ basin, we echo previous work in finding that the exponential model provides the best fit, with an average correlation length of approximately 10 km (the correlation length scale is 1/3 the range). In contrast, the average value over the ice sheet is approximately 25 km. These values are not precise because the assumption of independent samples in the computation of an empirical variogram is violated: nearby samples tend to be oriented parallel or perpendicular to flow. Therefore, we use these values as a guide for specifying the test length scales $l_H \in [10, 25, 40]$ km.

3.5.2 Recovery of basal topography for three correlation lengths.

We again ran the Metropolis-Hastings algorithm three times for 10^6 iterations each, discarding the first 10^5 samples, and each instance was given different initial values drawn from the prior distribution. We assessed the convergence of the samples using the same methods as discussed in Section 3.3.3, and consideration of traces, histograms drawn from different samplers, and the Gelman-Rubin statistic all indicated sampler convergence.

Figure 3.7 shows the pointwise posterior distributions produced by the MH algorithm for the flowline described above using $l_H = 10$ km. It is readily apparent that the bed elevation models of both *Bamber et al.* [2013] and *Morlighem et al.* [2014b] are admissible under the posterior distribution produced here, and neither is more probable than the other. The pre-IceBridge observations that were held back are also distributed according to the bounds produced by the posterior distribution, providing evidence for the predictive capabilities of the Bayesian approach.

The algorithm makes significant adjustments to the mass balance and width functions in order to accommodate velocity observations, particularly in the ice sheet interior. The requirement that the algorithm finds a posterior mass balance distribution that is improbable with respect to the prior distribution implies that at least one of the data sets considered herein may have a mean that is far from the true value. One hypothesis is that the surface mass balance model is underestimating aeolian snow redistribution.

Figure 3.8 shows the influence of correlation length on uncertainty estimates. Also included are the results of the method with the mass conservation relationship ignored, which is equivalent to ordinary Kriging. In regions of slow flow and relatively high uncertainty in velocity and mass balance rates, the uncertainty in derived thicknesses away from observations is large, and this is insensitive to correlation length. This is the case where mass continuity errors produce a similar amount of uncertainty to Kriging because it contains little useful information. Conversely, in regions of high data density, the covariance structure on thickness dominates. The smoothness constraint tends to dominate mass conservation, and considering the latter provides no advantage over Kriging in terms of precision. In the intermediate case, where thickness observations are sparse, but relative errors in the other constituents of the mass conservation model are low, the mass conserving interpolation scheme shines, producing uncertainties of less than half the Kriged case. The influence of correlation length on uncertainty estimates is also apparent: longer correlation lengths (i.e. smoother topography) produce lower uncertainties for a given data density.

3.6 Discussion

3.6.1 Application to the map-plane

The method presented herein was limited to the flowline case, and we argue that this context is useful for assessing the uncertainty that we expect from using mass conservation methods. However, many modeling applications require fields over the map plane. Because Monte Carlo methods require the solution of a forward model many times, the transition from solving the balance velocity equation on a flowline, which requires only quadrature, to the map plane, which requires the solution of a linear system of equations (not to mention an inherently greater number of degrees of freedom) is inherently expensive. However, this is not to say that the problem is intractable; because the Metropolis-Hastings algorithm is subject to the Ergodic Theorem, it is efficiently parallelizable in the sense that rather than running a single instance of the algorithm for many iterations, we can run many (independent) instances of the algorithm, and concatenate the resulting samples [*Murray*, 2010]. Aside from this numerical concern, the framework for utilizing the method in the map plane is given in Section 3.2.2 but without the simplifying assumptions of Section 3.2.3.

3.6.2 Selection of covariance models

The application of a geostatistical interpolation technique requires the selection of a model for the spatial covariance of the field in question, in this case ice thickness. In Bayesian methods this involves the *a priori* assumption of a correlation function with the specific form thereof derived from an empirical variogram or some other source [*Bamber et al.*, 2001; *Herzfeld et al.*, 1993]. Algorithms that cast the interpolation problem as one of PDE-constrained optimization impose an equivalent smoothness requirement through regularization [*Morlighem et al.*, 2011]. In either case, this choice introduces an influence on the pointwise posterior distribution of ice thickness of similar order to that of the mass conservation relationship.

These two processes (covariance function definition and regularization) are equivalent: regularizing on the square of the gradient is the same as using a Gaussian covariance function. The regularization parameter thus has physical relevance in that it is proportional to correlation length, and care must be taken in the selection of a smoothness parameter. It is not appropriate to regularize away the non-uniqueness of the problem with L-curve analysis because this selects the smoothest solution for which the data retains a good fit. However, this level of smoothness may not be physically mandated, and would tend to reduce the feasible range of solutions. Furthermore, the selection of a particular norm on the thickness gradient to minimize has the effect of choosing a covariance function. Experimental variograms suggest that the appropriate model for Greenland is usually exponential, yet the 2-norm regularization commonly used in basal topography inversions implies a Gaussian covariance. This may be desirable if a large scale volume estimate is the goal [*Bahr et al.*, 2014], but for local scale topographic estimation, it produces the wrong result. Instead, the degree of regularization or the covariance structure should be informed by independent analysis of bed covariance, and this must be incorporated into uncertainty estimates.

Nonetheless, deducing the appropriate covariance model and associated parameters is a difficult task, due to the sparsity of bedrock elevation measurements. This is compounded by the fact that glacier flow tends to produce landscapes with anisotropic topographic variability. This variability is spatially heterogeneous, and ice sheet interiors may have different geomorphic properties than the ice sheet margin or heavily glaciated mountain regions. The problem should be addressed by detailed radar soundings of ice thicknesses over regions deemed geomorphically representative of large scale conditions, as well as through analysis of recently deglaciated terrain. For ice sheets, this includes understanding the topographic variability in the heretofore less observed interior regions.

3.6.3 Flightline spacing

The selection of bed elevation measurement spacing is of great importance for future data acquisition campaigns, both in mountain and ice sheet environments. Considering Figure 3.8, and to a lesser extent Figure 3.4, there is an intermediate spacing regime over which mass conservation techniques can improve thickness uncertainty estimates. For long measurement spacing (i.e. greater than 20 ice thicknesses), mass conservation techniques may or may not outperform Kriging, depending on observational uncertainty. In either case, knowledge of the covariance structure does not improve uncertainty estimates, and errors may be large. For densely spaced flightlines, mass conservation does not offer an advantage over Kriging, because the smoothness imposed by the covariance model overrides any additional information due to continuity. Between these, mass conservation offers improvements in precision. We estimate this efficiency window to be one to two times the correlation length.

For Jakobshavn Isbræ, if we take the available ice-sheet wide empirical variograms computed from collected flightlines as guidance, then an isotropically oriented measurement spacing of approximately 10–20 km is appropriate. A more detailed look at existing measurements, particularly with an eye for discerning anisotropy in topographic covariance could improve estimates of uncertainty. This analysis could be improved further by collection of high density radar measurements over patches in order to better understand the short range topographic correlation structure that governs smoothness. For accurate assessment of the covariance structure, these patches would need to be at least 3 times the correlation length to a side. Regardless of topographic correlation structure, the admissible measurement spacing for a desired accuracy increases given more precise mass balance measurements (above a certain measurement spacing) and more precise velocities (to a point).

3.7 Conclusions

We have developed a Bayesian statistical model for inferring the posterior probability distribution of ice thickness given sparse observations thereof, coupled with observations of mass balance and surface velocity. Model variables are represented with Gaussian processes, which allow the specification of a covariance structure and prior information. These three parameters are linked through mass continuity. Our work advances upon previous methods in a few primary ways.

The continuity equation must be reformulated from acting instantaneously to over a finite time period. This point is subtle, because the resulting equation is similar to the original. However, the interpretation of the variables involved as time averages induces an additional step in the observation process, and we hence include an additional source of uncertainty (on top of measurement uncertainty) representing the deviation of observed quantities from their averages over the chosen averaging period of the continuity equation. Further constraining the form and magnitude of this uncertainty will be an important advance towards improving continuity-based interpolation schemes.

The Bayesian perspective allows the critical assumptions made in this model process to be elucidated. When assumptions of normality are used, they must be used explicitly. However, the method is general and assumptions of normality are not required. For example, this generality allows us to use a uniform distribution to model the relationship between surface and depth-averaged velocity.

Using Gaussian processes to represent fields such as thickness and mass balance allows for a straightforward and general way to impose smoothness constraints. We reiterate that there is a strong relationship between spatial covariance structure and regularization, and an accurate assessment of this value is critical in accurate modeling of basal topography as well as in determining the precision of those estimates.

For uncertainty estimates typical of velocities, mass balance, thickness, and data density in a mountain glacier, the algorithm reconstructed the basal topography, the distribution of which showed a relative uncertainty (as quantified by the normalized standard deviation) of around 10% at locations far from thickness measurement locations. Thickness uncertainty varies almost linearly with velocity uncertainty, due to the lack of an imposed covariance structure on velocity. On the contrary, the propagation of mass balance uncertainties is also influenced by the length scale of permissible variability in thickness and mass balance itself, as well as by data densities.

Application of the method to Storglaciären, a 3.1 km mountain glacier in the mountains of Sweden was successful, though the resulting uncertainties were relatively large. This is a result of the sparse nature of velocity measurements there, as well as the relatively large uncertainties in long-term average mass balance due to the high degree of interannual variability evident in the mass balance record.

At Jakobshavn Isbræ, the algorithm produced bed estimates with uncertainties varying greatly as a result of data density and relative uncertainty in different regions of the ice sheet. Consideration of mass conservation can improve thickness estimates in regions of low relative velocity and mass balance error, but tends to produce large uncertainties in regions of slow flow and small mass balance. We found that estimates of thickness uncertainty also depended strongly on correlation structure, which is equivalent to regularization in the PDE-constrained optimization context. Obtaining a better estimate of covariance structure would facilitate further application of mass conserving algorithms to ice sheets.

Based on our analysis, we suggest a flightline spacing of one to two times the topographic correlation length in order to best leverage mass conservation based interpolation techniques. For the Jakobshavn Isbræ region, we estimate this spacing to be between 10 and 20 km, but more work is necessary to determine the ideal value for the remainder of the ice sheet.

Acknowledgements

DB was supported by NSF Graduate Research Fellowship grant number DGE1242789. AA was supported by NASA grants NNX13AM16G and NNX13AK27G. MT was supported by NSF grant PLR 1107491. Thanks to Jesse Johnson, Ron Barry, Regine Hock, and Christina Carr for discussions

and review that led to great improvements to the manuscript.

Bibliography

Allen, C., C. Leuschen, P. Gogineni, F. Rodriguez-Morales, and J. Paden (2015), IceBridge MCoRDS L2 ice thickness for Greenland, National Snow and Ice Data Center Distributed Active Archive Center, doi:10.5067/GDQ0CUCVTE2Q.

Atkinson, K. E. (1978), *An Introduction to Numerical Analysis*, Wiley, New York.

Bahr, D. B., W. T. Pfeffer, and G. Kaser (2014), Glacier volume estimation as an ill-posed inversion, *Journal of Glaciology*, *60*(223), 922–934, doi:10.3189/2014JoG14J062.

Bamber, J. L., R. L. Layberry, and S. P. Gogineni (2001), A new ice thickness and bed data set for the Greenland Ice Sheet 1. measurement, data reduction, and errors, *J. Geophys. Res.*, *106*(D24), 33,773–33,780.

Bamber, J. L., J. A. Griggs, R. T. W. L. Hurkmans, J. A. Dowdeswell, S. P. Gogineni, I. Howat, J. Mouginot, J. Paden, S. Palmer, E. Rignot, and D. Steinhage (2013), A new bed elevation dataset for Greenland, *The Cryosphere*, *7*(2), 499–510, doi:10.5194/tc-7-499-2013.

Bindschadler, R. A., S. Nowicki, A. Abe-Ouchi, A. Aschwanden, H. Choi, J. Fastook, G. Granzow, R. Greve, G. Gutowski, U. Herzfeld, C. Jackson, J. Johnson, C. Khroulev, A. Levermann, W. H. Lipscomb, M. A. Martin, M. Morlighem, B. R. Parizek, D. Pollard, S. F. Price, D. Ren, F. Saito, T. Sato, H. Seddik, H. Seroussi, K. Takahashi, R. Walker, and W. L. Wang (2013), Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project), *Journal of Glaciology*, *59*(214), 195–224, doi:10.3189/2013JoG12J125.

Brinkerhoff, D., and J. Johnson (2015), A stabilized finite element method for calculating balance velocities in ice sheets, *Geoscientific Model Development*, *8*(5), 1275–1283, doi:10.5194/gmd-8-1275-2015.

Brinkerhoff, D. J., and J. V. Johnson (2013), Data assimilation and prognostic whole ice sheet modelling with the variationally derived, higher order, open source, and fully parallel ice sheet model VarGlaS, *The Cryosphere*, *7*(4), 1161–1184, doi:10.5194/tc-7-1161-2013.

Chandler, D. M., A. L. Hubbard, B. P. Hubbard, and P. W. Nienow (2006), A Monte Carlo error analysis for basal sliding velocity calculations, *J. Geophys. Res.*, *111*(F4), F04,005.

Christensen, R., W. Johnson, A. Branscum, and T. E. Hanson (2011), *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*, CRC Press.

Colgan, W., W. T. Pfeffer, H. Rajaram, W. Abdalati, and J. Balog (2012), Monte Carlo ice flow modeling projects a new stable configuration for Columbia Glacier, Alaska, c. 2020, *The Cryosphere*, *6*(6), 1395–1409, doi:10.5194/tc-6-1395-2012.

Csatho, B. M., A. F. Schenk, C. J. van der Veen, G. Babonis, K. Duncan, S. Rezvanbehbahani, M. R. van den Broeke, S. B. Simonsen, S. Nagarajan, and J. H. van Angelen (2014), Laser altimetry reveals complex pattern of Greenland Ice Sheet dynamics, *Proceedings of the National Academy of Sciences*, *111*(52), 18,478–18,483, doi:10.1073/pnas.1411680112.

Ettema, J., M. R. van den Broeke, E. van Meijgaard, W. J. van de Berg, J. L. Bamber, J. E. Box, and R. C. Bales (2009), Higher surface mass balance of the Greenland Ice Sheet revealed by high-resolution climate modeling, *Geophys. Res. Lett.*, *36*(12), L12,501.

Farinotti, D., M. Huss, A. Bauder, M. Funk, and M. Truffer (2009), A method to estimate the ice volume and ice-thickness distribution of alpine glaciers, *Journal of Glaciology*, *55*(191), 422–430, doi:10.3189/002214309788816759.

Gelman, A., and D. Rubin (1992), *A single series from the Gibbs sampler provides a false sense of security*, pp. 625–631, Oxford University Press.

Glen, J. W. (1955), The creep of polycrystalline ice, *Proceedings of the Royal Society, A*, *228*, 519–538.

Gogineni, S., J.-B. Yan, J. Paden, C. Leuschen, J. Li, F. Rodriguez-Morales, D. Braaten, K. Purdon, Z. Wang, W. Liu, and J. Gauch (2014), Bed topography of Jakobshavn Isbrae, Greenland, and Byrd Glacier, Antarctica, *Journal of Glaciology*, *60*(223), 813–833, doi:10.3189/2014JoG14J129.

Gudmundsson, G. H., and M. Raymond (2008), On the limit to resolution and information on basal properties obtainable from surface data on ice streams, *The Cryosphere*, *2*(2), 167–178, doi:10.5194/tc-2-167-2008.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika*, *57*(1), pp. 97–109.

Herzfeld, U. C., M. G. Eriksson, and P. Holmlund (1993), On the influence of kriging parameters on the cartographic output - a study in mapping subglacial topography, *Mathematical Geology*, *25*(7), 881–900.

Holmlund, P., P. Jansson, and R. Pettersson (2005), A re-analysis of the 58 year mass-balance record of Storglaciären, Sweden, *Annals of Glaciology*, *42*(1), 389–394, doi:10.3189/172756405781812547.

Hooke, R. L., P. Calla, P. Holmlund, M. Nilsson, and A. Stroeven (1989), A 3 year record of seasonal variations in surface velocity, Storglaciären, Sweden, *Journal of Glaciology*, *35*(120), 235–247.

Huss, M., and D. Farinotti (2012), Distributed ice thickness and volume of all glaciers around the globe, *Journal of Geophysical Research: Earth Surface*, *117*(F4), doi:10.1029/2012JF002523.

Jansson, P. (1997), Longitudinal coupling in ice flow across a subglacial ridge, *Annals of Glaciology*, *24*.

Jansson, P. (1999), Effects of uncertainties in measured variables on the calculated mass balance of Storglaciaren, *Geografiska Annaler*, *81A*(4), 633–642.

Joughin, I., B. E. Smith, I. M. Howat, D. Floricioiu, R. B. Alley, M. Truffer, and M. Fahnestock (2012), Seasonal to decadal scale variations in the surface velocity of Jakobshavn Isbræ, Greenland: Observation and model-based analysis, *Journal of Geophysical Research: Earth Surface*, *117*(F2), doi:10.1029/2011JF002110, f02030.

Kuriger, E. (2002), Analysis of structural features in the context of measured strain rates, Storglaciären, Sweden, Master's thesis, ETHZ, Zürich, Switzerland.

Li, H., Z. Li, M. Zhang, and W. Li (2011), An improved method based on shallow ice approximation to calculate ice thickness along flow-line and volume of mountain glaciers, *Journal of Earth Science*, *22*(4), 441–448, doi:10.1007/s12583-011-0198-1.

Link, W. A., and M. J. Eaton (2012), On thinning of chains in MCMC, *Methods in Ecology and Evolution*, *3*(1), 112–115, doi:10.1111/j.2041-210X.2011.00131.x.

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

McNabb, R. W., R. Hock, S. O'Neel, L. A. Rasmussen, Y. Ahn, M. Braun, H. Conway, S. Herreid, I. Joughin, W. T. Pfeffer, B. E. Smith, and M. Truffer (2012), Using surface velocities to calculate ice thickness and bed topography: a case study at Columbia Glacier, Alaska, USA, *Journal of Glaciology*, *58*, 1151–1164, doi:10.3189/2012JoG11J249.

Morlighem, M., E. Rignot, H. Seroussi, E. Larour, H. Ben Dhia, and D. Aubry (2010), Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica, *Geophysical Research Letters*, *37*(14), doi:10.1029/2010GL043853.

Morlighem, M., E. Rignot, H. Seroussi, E. Larour, H. Ben Dhia, and D. Aubry (2011), A mass conservation approach for mapping glacier ice thickness, *Geophysical Research Letters*, *38*(19), doi:10.1029/2011GL048659.

Morlighem, M., E. Rignot, J. Mouginot, X. Wu, H. Seroussi, E. Larour, and J. Paden (2013), High-resolution bed topography mapping of Russell Glacier, Greenland, inferred from Operation IceBridge data, *Journal of Glaciology*, *59*(218), 1015–1023, doi:10.3189/2013JoG12J235.

Morlighem, M., E. Rignot, J. Mouginot, H. Seroussi, and E. Larour (2014a), High-resolution ice-thickness mapping in South Greenland, *Annals of Glaciology*, *55*(67).

Morlighem, M., E. Rignot, J. Mouginot, H. Seroussi, and E. Larour (2014b), Deeply incised submarine glacial valleys beneath the Greenland Ice Sheet, *Nature Geoscience*, *7*, 418–422, doi:10.1038/ngeo2167.

Motyka, R. J., M. Truffer, E. M. Kuriger, and A. K. Bucki (2006), Rapid erosion of soft sediments by tidewater glacier advance: Taku Glacier, Alaska, USA, *Geophysical Research Letters*, *33*(24), doi:10.1029/2006GL028467, l24504.

Motyka, R. J., M. Fahnestock, and M. Truffer (2010), Volume change of Jakobshavn Isbræ, West Greenland: 1985–1997–2007, *Journal of Glaciology*, *56*(198), 635–646, doi:10.3189/002214310793146304.

Murray, L. (2010), Distributed Markov Chain Monte Carlo, *Proceedings of Neural Systems Workshop on Learning on Cores, Clusters, and Clouds*, *11*.

Patil, A., D. Huard, and C. J. Fonnesbeck (2010), Pymc: Bayesian stochastic modelling in python, *J. Stat. Softw*, pp. 1–81.

Perego, M., S. Price, and G. Stadler (2014), Optimal initial conditions for coupling ice sheet models to earth system models, *Journal of Geophysical Research: Earth Surface*, *119*(9), 1894–1917, doi: 10.1002/2014JF003181.

Petra, N., J. Martin, G. Stadler, and O. Ghattas (2013), A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet flow inverse problems, *ArXiv e-prints*.

Rasmussen, C. E. (2006), *Gaussian processes for machine learning*, MIT Press.

Rignot, E., and P. Kanagaratnam (2006), Changes in the velocity structure of the Greenland Ice Sheet, *Science*, *311*(5763), 986–990, doi:10.1126/science.1121381.

Rignot, E., and J. Mouginot (2012), Ice flow in Greenland for the International Polar Year 2008–2009, *Geophysical Research Letters*, *39*(11), doi:10.1029/2012GL051634, l11501.

Sergienko, O. V., T. T. Creyts, and R. C. A. Hindmarsh (2014), Similarity of organized patterns in driving and basal stresses of Antarctic and Greenland ice sheets beneath extensive areas of basal sliding, *Geophysical Research Letters*, *41*(11), 3925–3932, doi:10.1002/2014GL059976.

Seroussi, H., M. Morlighem, E. Rignot, E. Larour, D. Aubry, H. Ben Dhia, and S. S. Kristensen (2011), Ice flux divergence anomalies on 79 North Glacier, Greenland, *Geophysical Research Letters*, *38*(9), doi:10.1029/2011GL047338, l09501.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, doi:10.1137/1.9780898717921.

Williams, C. (1998), Prediction with Gaussian Processes: From linear regression to linear prediction and beyond, in *Learning in Graphical Models*, *NATO ASI Series*, vol. 89, edited by M. Jordan, pp. 599–621, Springer Netherlands, doi:10.1007/978-94-011-5014-9_23.

Chapter 4

Velocity variations at Columbia Glacier captured by particle filtering of oblique time-lapse images[1]

Abstract

We develop a probabilistic method for tracking glacier surface motion based on time-lapse imagery, which works by sequentially resampling a stochastic state-space model according to a likelihood determined through correlation between reference and test images. The method is robust due to its natural handling of periodic occlusion and its capacity to follow multiple hypothesis displacements between images, and can improve estimates of velocity magnitude and direction through the inclusion of observations from an arbitrary number of cameras. We apply the method to an annual record of images from two cameras near the terminus of Columbia Glacier. While the method produces velocities at daily resolution, we verify our results by comparing eleven-day means to TerraSar-X. We find that Columbia Glacier transitions between a winter state characterized by moderate velocities and little temporal variability, to an early summer speed-up in which velocities are sensitive to increases in melt- and rainwater, to a fall slowdown, where velocities drop to below their winter mean and become insensitive to external forcing, a pattern consistent with the development and collapse of efficient and inefficient subglacial hydrologic networks throughout the year.

4.1 Introduction

Motion is what defines a glacier, and measuring this motion is a principal concern for understanding changing ice dynamics. Observed over multiple years, ice velocity and acceleration inform the dynamic component of glacial contributions to sea level rise. [e.g. *Burgess et al.*, 2013; *Joughin et al.*, 2010]. In the very short term (e.g. hourly measurement intervals), velocity measurements can elucidate the physics of glacial response to diurnal or tidal forcing [*Dietrich et al.*, 2007; *Meier et al.*, 1994]. At intermediate scales, velocity changes provide information on the sensitivity of glacier flow to changes in liquid input (i.e. from storms or hot days) and on the configuration of the subglacial drainage system that determines this sensitivity [*Harper et al.*, 2007]. Analysis of image sequences captured by ground-based cameras provides a compelling mechanism for evaluating glacier velocity variations at short and intermediate time scales because of the low cost and ubiquity of the necessary equipment. While many photogrammetric techniques are applicable to both orthogonal and oblique

---

[1] Prepared for submission in *Journal of Glaciology* with S. O'Neel.

imagery, we confine our review to the latter. At the most basic level, methods of capturing the velocity consist of two elements: first, a means of tracking persistent features in image coordinates. Second, a relation between image and spatial coordinates.

The first quantitative application of time-lapse imagery to glaciers in was by *Flotron* [1973], who resolved both the seasonal signal of horizontal velocity, but also a small signal of vertical motion. In Alaska, *Harrison et al.* [1986] used automatic 35mm cameras to document daily speedups on Variegated Glacier during the three summers prior to its surge in 1982. *Krimmel and Rasmussen* [1986] used a similar system at the terminus to Columbia Glacier during the early stages of its retreat, manually tracking 30 points through a sequence of 3 frames per day of the summer of 1983. *Harrison et al.* [1992] describe an approach to tackling one of the more significant difficulties associated with tracking slow moving fluid: the camera also has a tendency to move and that control points are not always available. Despite the lack of surveyed ground control points or a stable camera position, they were able to document the surge of West Fork Susitna Glacier.

Since the advent of digital cameras, a handful of methods have emerged by which to perform automatic measurements with oblique cameras, in contrast to the manual methods used previously. *Evans* [2000] used a probabilistic metric to track multiple potential flow pathways. However, he stopped short of converting solutions from image coordinates to spatial coordinates, and so the utility of this tool for practical glaciology is limited. *Dietrich et al.* [2007] used automated tracking of features on the surface of Jakobshavn Isbræ, in conjunction with a photogrammetrically-derived DEM, resolving both horizontal and vertical motion of the Jakobshavn terminus under an *a priori* assumption of flow direction from a remotely-sensed velocity field. Cameras were calibrated and corrected with surveyed fiduciary points, and were oriented such that they were orthogonal to the primary flow direction. Their work represents best practices in camera installation, but these were usually not considered for cameras where feature tracking was not the initial intent. *Rosenau et al.* [2013] expanded this experiment over several more years, localizing the grounding line position at Jakobshavn based on presence or absence of tidal motion. At Helheim Glacier, feature tracking was similarly used to quantify tidal flexure near the glacier terminus [*Murray et al.*, 2015; *James et al.*, 2014]. *Ahn and Box* [2010] demonstrated a robust method for tracking flow fields utilizing multiple image pre-processing techniques simultaneously to reduce the incidence of incorrect matches between image pairs from several cameras in Greenland. The resulting velocities agree well with those derived from remote sensing. Finally, *Messerli and Grinsted* [2014] and *James et al.* [2016] provided software libraries that distill many of the essential methods required for oblique time-lapse velocity measurements into accessible and open-source packages.

In this paper, we address the problem of determining glacier velocities from an unstable camera under changing lighting conditions without precise ground control. In contrast to previous work, we apply a probabilistic approach that allows us to address some of the key difficulties of other methods while producing robust estimates of uncertainty under unfavorable conditions that are still ubiquitous 25 years after *Harrison et al.* [1992]. We apply a method called particle filtering, which sequentially updates the probability distribution of the state of a moving glacier surface, namely its position and velocity by considering a likelihood derived from matching characteristic features between images. Besides robust error accounting, this method immediately generalizes to multiple cameras, is robust to partial occlusion due to the use of predictions by an underlying physical model, and does not require the ill-posed step of projecting values in image coordinates onto the landscape. We apply the method to images collected by a pair of (non-stereo) cameras at Columbia Glacier, AK, that are subject to all of the potential uncertainty sources typical of time-lapse image analysis. We process images over nearly a full year, for the first time providing speeds over all seasons at daily temporal resolution, providing new insight into the evolution of basal processes near the terminus of a large tidewater glacier.

4.2 Columbia Glacier

Columbia Glacier (Fig. 4.1a) is a 52 km long tidewater glacier that drains the high central Chugach Mountains in southcentral Alaska. Its climate is strongly maritime with annual precipitation rates of near 3m at the terminus. The terminus itself is characterized by high speeds and vigorous calving [*Meier and Post*, 1987]. Winter temperatures are moderate, while early summers see strong melt. large rainstorms are common in the late summer and fall [*Bieniek et al.*, 2012].

Columbia Glacier is currently in the retreat phase of the tidewater glacier cycle [*Meier and Post*, 1987], and has retreated more than 25 km from its maximum in 1980. The specific mechanism for initiating the retreat is debatable [*Sikonia and Post*, 1980; *Carlson et al.*, 2017], as the glacier geometry at Columbia Glacier's maximum length was highly unstable due to the significant overdeepening in what is now Columbia's fjord. In any case, the current retreat is driven by glacial dynamics and bed topography, with climate assuming an ancillary role [*Pfeffer*, 2007].

The velocity of Columbia Glacier, particularly near the terminus, has varied greatly since the beginning of the retreat and these variations have been documented over a variety of temporal and spatial scales since the 1980s. *Vaughn et al.* [1985] used an automated laser rangefinder and a small

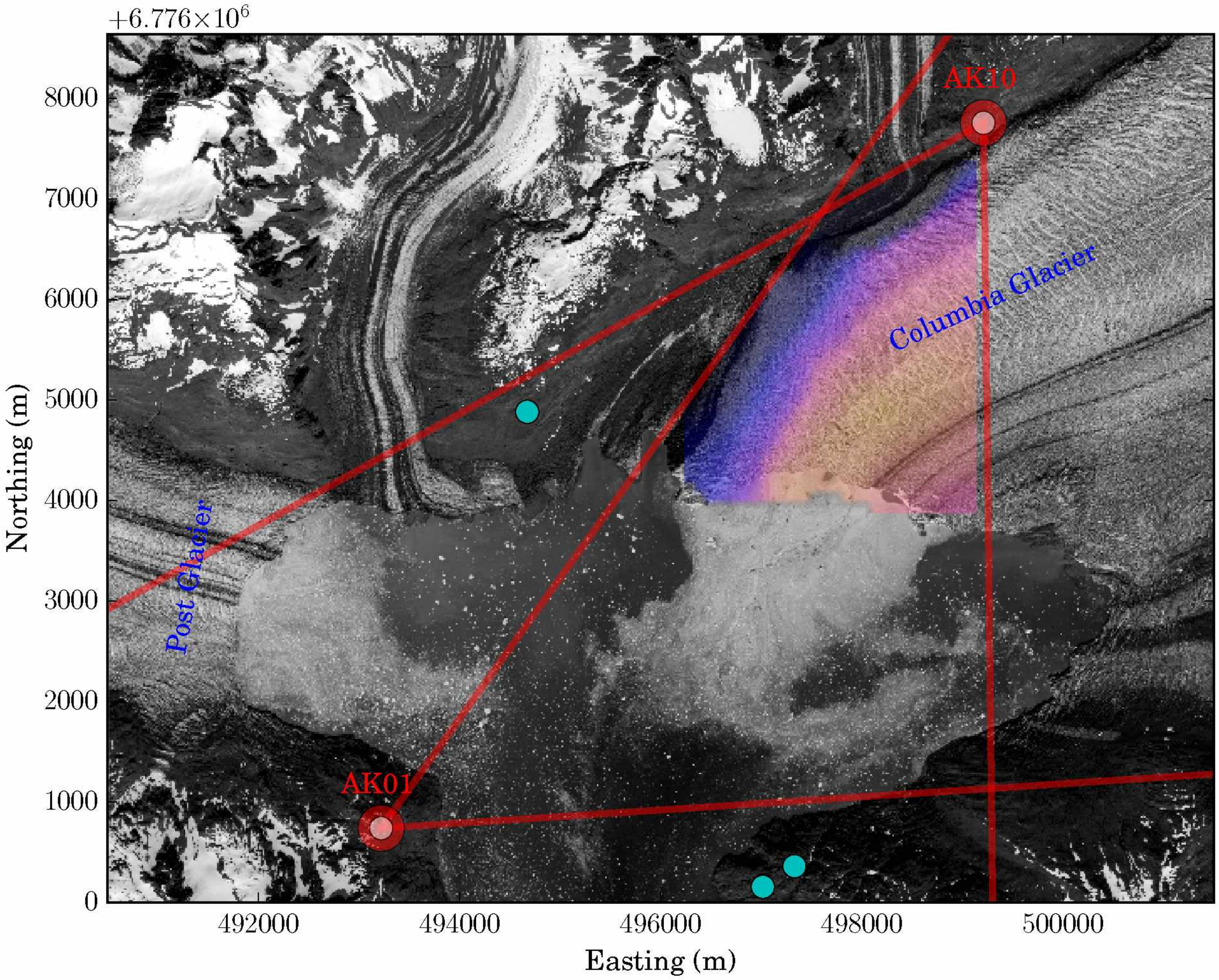

Figure 4.1. Columbia Bay and the Columbia Glacier terminus c. July 2013. Cameras used in this study are denoted by red bullseyes, and their field of view by red lines. Additional time-lapse cameras that were not included are given by cyan circles. We compute velocity fields in the glacierized region that falls within the field of view of both cameras. The colored surface is the glacier speed on 2013.07.15 (see Fig. 4.2 for colorbar). The temperature data shown in Fig. 4.5 is recorded at a station $\tilde{2}$ km out of frame to the northeast. Base image courtesy Polar Geospatial Center.

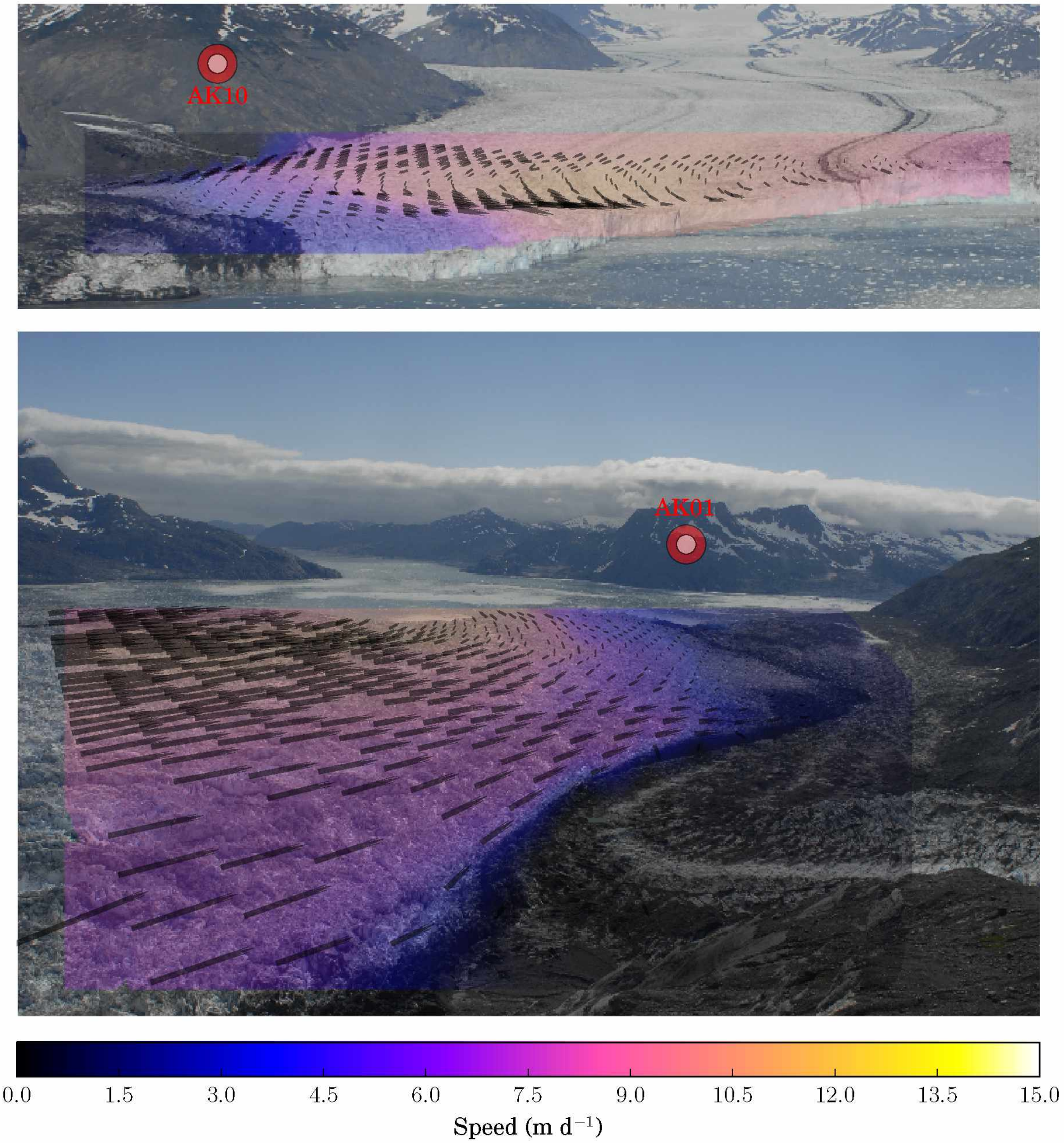

Figure 4.2. Representative images for AK01 (top) and AK10 (bottom), the two cameras used in this study. The location of the other camera is denoted by a bullseye. The colors represent glacier speed on 2013.07.15, with vectors indicating flow speed and direction in image coordinates. Note the presence of a strong shear margin which is well captured by the algorithm, as well as the very low image aspect ratios.

set of reflectors installed in the ice near the terminus to measure pointwise velocities at 15 minute intervals. Despite the excellent temporal resolution, this method was only applied to a handful of points, and the record only lasts for approximately 30 days during the summers of 1984-1986. *Meier et al.* [1994] performed similar measurements in 1987. *Krimmel and Rasmussen* [1986] performed one of the first examples of using oblique time-lapse photography to measure glacier velocities at daily resolution over both winter and summer. The camera, situated over 5 km from the study site was able to produce daily offsets with a nominal precision of 1m d^{-1}.

These temporally detailed results are supplemented with distributed velocity fields determined from repeat georectified aerial photographs with intervals of around two months [*Krimmel*, 2001], in which characteristic surface features such as crevasses were manually identified in image pairs. Since that time, major improvements in the availability of satellite imagery have allowed the automated generation of velocity fields with full coverage and enhanced temporal resolution. *Fahnestock et al.* [2016], used optical imagery from Landsat8 to produce velocity fields for each 15-day offset image pair in the satellite's brief record, excluding periods when the landscape was obscured by clouds. Circumventing the occlusion issues of optical imagery, *Burgess et al.* [2013] used synthetic aperture radar (SAR) to generate velocity fields at Columbia Glacier (and elsewhere) with a temporal resolution of 46 days at irregular intervals between 2007 and 2009. *Joughin et al.* [2010] and *Vijay and Braun* [2017] independently used SAR observations with a higher return frequency to produce 11-day average velocities between 2010 and 2016, with approximately monthly frequency. As a side note, *Vijay and Braun* [2017] produced commensurate digital elevation models that coincide with their velocity fields which are publicly available.

Taken in aggregate, a consistent story about the spatial and temporal variations in Columbia's flow has emerged, superimposed on tidewater retreat. At seasonal time scales, Columbia has a velocity minimum in October or November, before a slow ramping up to more consistent values through mid-winter into early spring. In late spring, the glacier rapidly accelerates, reaching a velocity maximum in May or June, before a decline back to the minimum state in the fall, a pattern that has persisted from when it was first enumerated [*Meier and Post*, 1987] through present [*Vijay and Braun*, 2017], albeit with a drift towards respective maxima and minima occurring later in the season. The magnitude of these variations have changed throughout the course of the retreat. In the early 1980s, maximum velocities were less than 10m d^{-1}, increased to 30 m d^{-1} during the periods of most vigorous retreat in the mid 1990s to 2000s [*O'Neel et al.*, 2005], and have been on the order of 15 m d^{-1} since 2010 [*Vijay and Braun*, 2017].

Since ice geometry changes relatively slowly, and these velocities are an order of magnitude higher than could be explained by deformation, seasonal evolution of subglacial pressure is commonly thought to drive these variations due to the ability of water pressure to partially offset the normal stress exerted by ice on the bed and thus reducing friction. Pressure variations are induced by changes in the availability of surface water (either melt or rainfall) or changes in the character of the subglacial drainage system [e.g. *Iken and Truffer*, 1997; *Werder et al.*, 2013]. Based on water pressure observations from terrestrial glaciers, it is hypothesized that during winter the lack of surface inputs lead to a weakly-connected drainage system with moderate pressure that produce moderate velocities [*Iken and Truffer*, 1997; *Truffer and Harrison*, 2006]. At the onset of the melt season in spring, the additional water overloads this system, leading to high pressures and velocities, but also causing an efficient drainage network to form. In the fall, as water input decreases, water pressure drops below the annual average before the drainage system can once again equilibrate to winter conditions.

The multiple field campaigns of the 1980s reported velocity fluctuations superimposed on these longer term signals during summer. Signals of acceleration were observed at both diurnal and tidal periods, presumably associated with short term changes in water pressure. Additionally, stochastic events such as foehn winds or large rainstorms were also observed to cause complex increases in velocity, often (but not always) followed by a decrease to below pre-event speeds. Interestingly, contemporaneous measurements of pressure and speed did not reveal a clear relationship between the two [*Kamb et al.*, 1994]. However, these velocity changes were well (but not perfectly) correlated with changes in water storage as inferred from proxies for influx and outflux [*Fahnestock*, 1991]. The observed mismatch in the direct pressure signal was ascribed to heterogeneity in the local subglacial drainage system relative to the area-averaged value relevant to glacier dynamics and aliased by total water storage.

Since the observations described above, few direct observations of Columbia Glacier's velocity have been collected. While remote sensing is extremely useful for describing synoptic features, it lacks the temporal resolution necessary for assessing glaciological response to short term changes in geometry (i.e. calving) or hydrology. Given the remarkable changes that have occurred over the last 30 years, it is not clear whether Columbia Glacier's dynamics resemble those observed c. 1987. Fortunately, Columbia Glacier has been the focus of an intense campaign of oblique time-lapse photography since 2007 as part of an Extreme Ice Survey and USGS monitoring program. Since 2007, cameras at seven sites have been active, generally taking images of the glacier terminus at sub-hourly intervals (Fig. 4.1b). Here we apply the method described below to these images to fill this data gap and to determine the degree to which short term velocity variability has changed since observations were last taken at such resolutions. Our method provides velocity fields at spatial resolutions of 100 m with daily estimates of the three-day average over 11 months.

4.3 A Bayesian method for tracking glacier surface features

The problem we seek to solve is as follows: given a set of sequential images of a glacier surface taken from an approximately identical vantage point, find the motion of the glacier and an associated estimate of uncertainty in a spatial (rather than image) frame of reference (See Fig. 4.3). While methods for tracking features between images are ubiquitous in computer vision, the particular problem of tracking glaciers presents a few notable challenges and requirements. First, since the images are oblique and we are interested in motion in a spatial frame of reference, the method must be amenable to the use of a projective transformation in some way. Projection from spatial coordinates to image coordinates is straightforward. However, the inverse operation of projecting an image onto a landscape that varies in the vertical coordinate is ill-posed because of discontinuities induced by partial occlusion of background elements by foreground elements, particularly as the angle between the camera direction and the surface becomes small. As such we, wish to avoid this inverse transformation. Second, glaciers occur outside and in bed weather, so the method must be robust to occasional occlusion and also to changes in lighting. Third, the method must be able to handle the considerable clutter present on a natural glacier surface. As a means to address all of these problems, we have developed a method that combines so-called template matching with particle tracking.

Template matching works by finding the mismatch between a reference sub-image and a larger test sub-image as a function of pixel displacement. For example, if a test sub-image were constructed by rigidly translating a reference sub-image by one pixel up and to the right, then the error surface computed by template matching would be minimal for a displacement of one pixel up and to the right. For natural images, this error surface likely contains multiple minima of various degrees, representing several feasible image offsets of varying certainty. To improve robustness to changing lighting conditions, we apply local image processing to both reference and test sub-images. Template matching utilizes large neighborhoods of data and so is robust to clutter and mismatching compared to similar methods that compare local features [e.g. *Lowe*, 1999], but it cannot handle large deformation and is somewhat computationally expensive. Fortunately, neither

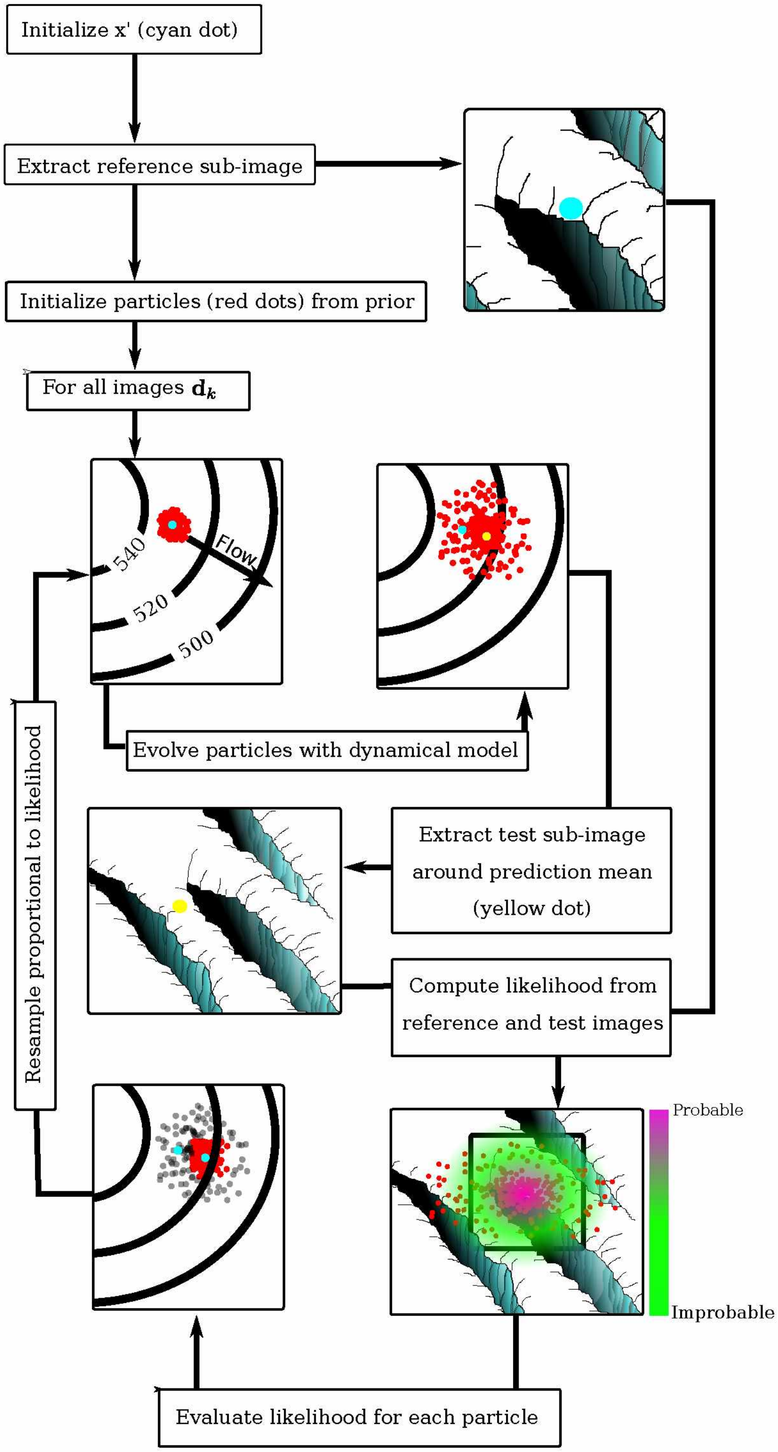

Figure 4.3. Graphical depiction of the steps of the hybrid template matching-particle filtering developed in this paper.

of these restrictions are too onerous in the context of glaciology, where motion is slow and real-time processing is not required.

Particle tracking is a method for sequentially updating the probability distribution of the state (primarily position and velocity) of a dynamical model based on a stream of observations (in this case images from a time-lapse camera). Relevant statistics (such as the mean or maximum likelihood estimator) can then be extracted from the distribution. The dynamical model that we consider essentially says that a point on the glacier surface moves with nearly constant velocity, subject to small random accelerations typical of glacial velocity variations. The probability distribution of the state is approximated by a large number of 'particles,' each representing a potential state, which are evolved through time based on the stochastic dynamical model. When observations are available, particles that are unlikely with respect to these observations are culled, while observations that are likely are replicated. The dynamical model exists entirely in the spatial reference frame. To determine which particles are likely, they are projected into image space according to a numerical model of the camera, and deemed likely or unlikely with respect to the error surface computed with template matching. Note that this allows the algorithm to explore multiple hypotheses, since template matching may produce more than one likely offset. While many particles are required to fully sample the space of possible velocities, the projection of image offsets back into spatial coordinates is never required. Additionally, in the absence of observations (or if observations are of low quality due to bad weather), the state evolves according to the dynamical model only.

We can repeat the above steps for any location desired on the glacier surface. By applying this method to a grid of points and applying interpolation to the resulting solutions, we thus produce velocity fields.

4.3.1 Glacier motion model

We use a Lagrangian state-space model (i.e. a set of input variables $\mathbf{m}_k$, a set of output variables $\mathbf{m}_{k+1}$ and a set of first order discrete transfer functions) to represent the motion of trackable features along the surface of a glacier. The model state variables are the map-plane coordinates $\mathbf{x}$, the map-plane velocities $\mathbf{v}$, the elevation z, and a systematic elevation offset from a datum δS, which form the state vector at the k-th time step $\mathbf{m}_k = [\mathbf{x}_k, \mathbf{v}_k, z_k, \delta S_k]$. We assume that a specified point of interest moves tangent to an imprecisely known glacier surface with its velocity subject to random accelerations, which yields the discrete difference equations

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \Delta t\, \mathbf{v}_k + \frac{\Delta t^2}{2}\, \mathbf{a}_k && (4.1) \\ \mathbf{v}_{k+1} &= \mathbf{v}_k + \Delta t\, \mathbf{a}_k && (4.2) \\ z_{k+1} &= S(\mathbf{x}_{k+1}) + \delta S_{k+1} && (4.3) \\ \delta S_{k+1} &= \delta S_k + \sigma_z \left\| \mathbf{v}_k \right\| \Delta t. && (4.4) \end{aligned}$$

$\mathbf{a}_k$ are normally distributed random accelerations in both horizontal directions

$$\mathbf{a}_k \sim \mathcal{N}(0, \Sigma_{v,k}), \tag{4.5}$$

where $\Sigma_{v,k}$ is a diagonal covariance matrix with entries given by an assumed characteristic variance in glacier velocities. $S(\mathbf{x}_{k+1})$ is a reference surface elevation field, for example an interpolant to a DEM. Since errors in this reference surface are likely to be systematic, we assume that uncertainty in the glacier surface evolves as a random walk that depends on how far the particle moves over the DEM, and a characteristic slope of small scale features σ_z. We neglect the evolution of the surface elevation due to melting, assuming that this is at least partially accounted for by emergence velocity. In regions where the ice is flowing quickly the error between the true surface elevation and the reference elevation can evolve quickly, while in regions that are not moving at all, the error should remain approximately constant.

The initial state $\mathbf{m}_0$ is specified as

$$\begin{aligned} \mathbf{x}_0 &\sim \mathcal{N}(\mathbf{x}', \Sigma_{\mathbf{x}}), && (4.6) \\ \mathbf{v}_0 &\sim \mathcal{N}(\mathbf{v}', \Sigma_{\mathbf{v}}), && (4.7) \\ z_0 &= S(\mathbf{x}_0) + \delta S_0, && (4.8) \\ \delta S_0 &\sim \mathcal{N}(0, \Sigma_S), && (4.9) \end{aligned}$$

where $\mathbf{x}'$ is the nominal location to be tracked, $\mathbf{v}'$ is an initial guess of the velocity, and the various Σ are covariance matrices associated with these initial distributions. A stochastic state-space model can also be written as a random vector drawn from a distribution conditioned on the previous state, or

$$P(\mathbf{m}_k | \mathbf{m}_{k-1}) = \mathcal{N}(\mathcal{F}(\mathbf{m}_{k-1}), \Sigma_{k-1}), \tag{4.10}$$

where $\mathcal{F}$ is the deterministic component of Eqs. 1–4, and Σ_k is the covariance matrix associated with the noise.

4.3.2 Applying Bayes' Theorem

The distribution of potential solutions produced by the model described above when used with reasonable estimates of initial distributions and process noise is large. We wish to determine which of these solutions are likely with respect to a sequence of error-prone observations, in this case the displacement of a characteristic pattern of surface features associated with the the nominal tracked location $\mathbf{x}'$ between oblique images. Stated more rigorously, we seek the probability distribution of a current state $\mathbf{m}_k$ as constrained by all images up to and including that at the current time $\mathcal{D}_k = \{\mathbf{d}_i : i \in 1, \ldots, k\}$, where $\mathbf{d}_i$ is an image at time i. We restrict our consideration to the case that two assumptions hold, both of which are true in this context. First, the transition between states must be a Markov process, which is to say that $\mathbf{m}_k$ is independent of all states except the previous one $\mathbf{m}_{k-1}$, or

$$P(\mathbf{m}_k|\mathbf{m}_{k-1}, \ldots, \mathbf{m}_0) = P(\mathbf{m}_k|\mathbf{m}_{k-1}). \tag{4.11}$$

This means that the transition between states has no memory. Second, observations must depend only on the contemporaneous state, and are independent of all other observations and non-contemporaneous states:

$$P(\mathcal{D}|\mathbf{m}_k, \ldots, \mathbf{m}_0) = \prod_{i=1}^{n} P(\mathbf{d}_i, \mathbf{m}_i). \tag{4.12}$$

Using these assumptions combined with Bayes' Rule [*Tarantola*, 2005] allows us to sequentially update our belief in the state distribution as additional measurements are added:

$$P(\mathbf{m}_k|\mathcal{D}_k) \quad \propto \quad P(\mathbf{d}_k|\mathbf{m}_k)\, P(\mathbf{m}_k|\mathcal{D}_{k-1}), \tag{4.13}$$

where $P(\mathbf{d}_k|\mathbf{m}_k)$ is the *likelihood*, which describes how likely it is to observe the current measurement $\mathbf{d}_k$ assuming a state $\mathbf{m}_k$, and $P(\mathbf{m}_k|\mathcal{D}_{k-1})$ is a *prior* distribution that describes how feasible a state is given all previous images but before considering the present image. Multiplying the likelihood and the prior together yields the *posterior* distribution (to a normalizing constant), which is the probability distribution of states after having considered all available observations. Note that the posterior distribution (the probability density function of position and velocity of a given point after considering a set of images) can be dominated by either the likelihood or the prior. In the former case, if an observation is equally likely given any state $\mathbf{m}$ (for example, in the case of complete occlusion of the image by, say, a cloud), then the likelihood is constant and the posterior distribution is only proportional to the prior: the observation has added no new information and reverts to the prior. Conversely, if observations are very certain and the prior relatively vague, then the posterior distribution will be governed by observations.

The prior distribution is constructed by propagating the posterior distribution at $k-1$ through the state model, which is to say that the best guess for the current state is the fully-constrained previous state updated with the model dynamics. This turns out to be true, which can be seen by factoring the expression for the prior as

$$P(\mathbf{m}_k|\mathcal{D}_{k-1}) = \int P(\mathbf{m}_k|\mathbf{m}_{k-1})P(\mathbf{m}_{k-1}|\mathcal{D}_{k-1})\mathrm{d}\mathbf{m}_{k-1}, \tag{4.14}$$

which is a form of the Chapman-Kolmogorov forward equation [*Doucet and Johansen*, 2009]. In this equation, the first term is the probability distribution of the new state given the old state (or the forward model), and the second term is the posterior distribution from the previous time step. This equation forms a new distribution by applying the system dynamics to the posterior at the previous time step. In the case that all distributions are Gaussian and the system dynamics are linear, Eq. 4.14 can be solved analytically (leading to the well-known Kalman Filter). However, in this case neither assumption is true due to the non-linear constraint of surface tangent motion, and also due to non-linearity in the process of making observations, described below.

4.3.3 Measurement

The likelihood can be interpreted as follows: given a known state, what is the probability of a camera recording a given (sub-)image. The process we adopt here can be summarized as a) project the mean state position into image coordinates, b) extract a sub-image in the neighborhood around the projected point and perform local image processing, and c) compute the sum of squared differences between the sub-image and all other sub-images in the neighborhood, which is interpreted as the scaled logarithm of the likelihood.

Specification of a camera model

Our method relies on possessing an accurate function for projecting coordinates in physical space to coordinates in image space. We adopt the model of *Claus and Fitzgibbon* [2005], which is specified by camera position, camera orientation, focal length, camera sensor size, and radial and tangential lens distortion. *A priori*, only the camera location is known, and that with limited precision in this case (namely the precision of a hand-held GPS). We solve for the remaining parameters by minimizing the misfit between a set of points that are uniquely identifiable in both a digital elevation model and in a reference image (mostly prominent features such as mountain peaks and shoreline outcrops) using Powell's algorithm [*Powell*, 1964].

Reference and search sub-images

We begin by projecting the nominal location of the point we wish to track into image coordinates. We then find the nearest integer pixel and extract a $m_r \times n_r$ sub-image T, which becomes the reference sub-image that we track throughout the period of interest. As a preprocessing step, we perform a whitened principal components transform to convert the image from RGB to a Z-normalized intensity [*Smith et al.*, 2002]. We then apply a highpass median filter to highlight edges and partially remove the effect of shadows.

Each time an image becomes available, we compute the weighted mean from $P(\mathbf{m}_k|\mathcal{D}_{k-1})$ (i.e. the prediction step) and project it into image space. We then extract a test sub-image I with size $m_t \times n_t, m_t > m_r, n_t > n_r$. The sub-image sizes are found through trial and error, with a tradeoff between feature uniqueness and clutter as sub-image size increases. We apply the same preprocessing steps with the addition of histogram matching step for each band of the test sub-image, such that the color profile matches that of the reference template. This helps to ameliorate some of the effects of changing illumination.

Computation of the likelihood

With preprocessed reference and test sub-images in hand, we compute the area-averaged sum of squared differences between the reference template and test template for all possible pixel offsets u, v for which the reference template falls entirely within the test template

$$\ell(u,v) = \frac{1}{m_r n_r} \sum_{u',v'} (T(u',v') - I(u+u', v+v'))^2. \tag{4.15}$$

We then define the likelihood as

$$\mathcal{L}(u,v) \propto \exp\left(-\frac{\ell(u,v)}{\sigma_\ell^2 + \sigma_m^2}\right), \tag{4.16}$$

where σ_ℓ is the measurement uncertainty due to illumination changes and deformation, and σ_m is the uncertainty due to camera motion in pixels [*Nakhmani and Tannenbaum*, 2008]. The deterministic procedure would be to find $\text{argmax}_{u,v}\ \mathcal{L}(u,v)$ (often with a sub-grid parameterization to increase precision) as the (single) measurement. However, because of the quasi-periodic nature of trackable glacier surface features (namely crevasses) as well as changes in illumination, there are often several peaks in the likelihood, often of comparable magnitude. Here, we do not try to find this peak, instead retaining the complete likelihood for the update of the posterior distribution. This allows particles to explore multiple hypotheses for each new image. Because spurious offsets tend to be inconsistent between images, while the "true" peak is persistent (even if it is not the most probable solution for a given scene), incorrect hypotheses tend to be ephemeral, while good solutions remain probable over multiple images.

Camera model correction

In practice, cameras are not perfectly stable due to changes in temperature, wind, and other unknown factors which conspire to produce small offsets between image pairs [e.g. *Harrison et al.*, 1992]. To determine this offset, we track a set of points $\{u_g, v_g\}$ that are on land (as opposed to ice or water), and thus assumed to be stationary. We then compute the maximum likelihood solution with subgrid precision for each control point, and fit a 3 parameter rigid rotation-translation model for each image using the RANSAC algorithm [*Fischler and Bolles*, 1981] to eliminate outliers induced by occlusions from clouds, errors in choosing the correct motion, etc. The residual of this fit divided by the number of control points (and penalized for excluding data points) is used as σ_m, the measurement uncertainty due to camera motion. When computing the likelihood for non-stationary points, we then add the offset predicted by the rotation-translation model to the projected image coordinates, which minimizes the influence of camera motion.

Multiple cameras

Assuming n_c cameras, the above steps can be performed for each, and due to independence, the resulting probabilities can be multiplied. Because an accounting of observational uncertainty is inherent in the system, the information derived from each camera is properly weighted: cameras far from a given point or with unfavorable geometry are weighted less because likelihood maxima are more diffuse.

Even cameras that do very little in terms of specifying flow speed on their own can dramatically increase the accuracy of other cameras. A schematic justification for this is given in Fig. 4.4. In the figure, Camera a successfully captures the component of the flow vector parallel to the image plane. Unfortunately, there are multiple vectors that can appear identically to the camera. Though in the real world, this is somewhat offset by the camera being elevated above the surface, this advantage is partially offset by uncertainty in the location of the surface (i.e. DEM errors). Camera b has a different problem: because the flow vector is normal to the image plane, the camera cannot determine anything about the magnitude of the flow. However, it does have the advantage of having full knowledge of the flow direction. Individually, neither of these measurements is very

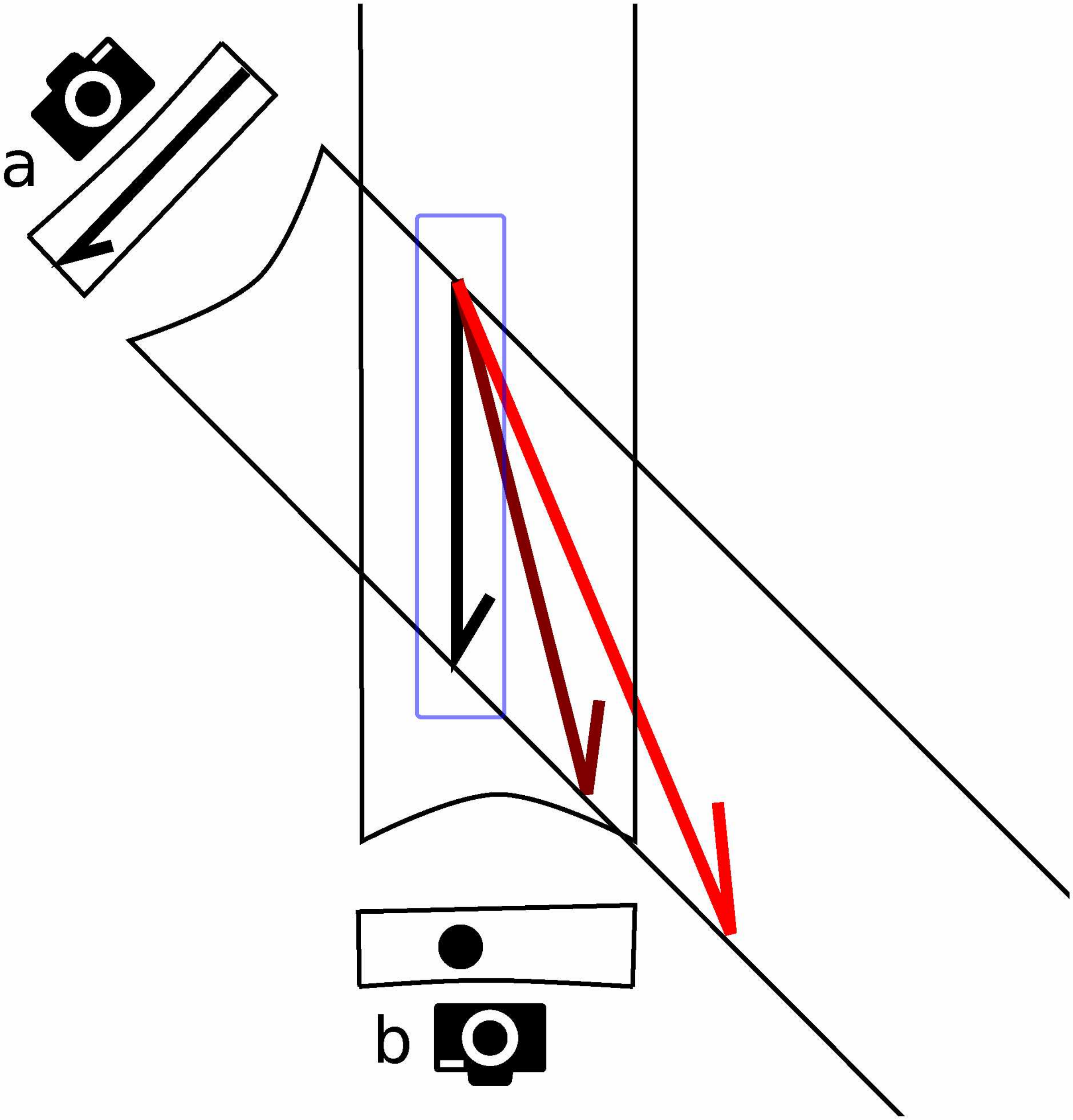

Figure 4.4. Schematic illustrating the utility of multiple cameras in oblique flow tracking. Even if both cameras are insufficient to fully resolve flow direction and magnitude on their own, two cameras operating in tandem typically provide strong constraints on one another.

satisfying. In tandem however, the information from camera b specifies which vector camera a has measured, producing the correct measurement of both magnitude and direction of the offset vector. As shown in Fig. 4.1, the situation at Columbia Glacier is similar to the hypothetical scenario of Fig. 4.4.

4.3.4 Particle Filtering

Given that our problem precludes analytical solution, we must instead find an approximate numerical solution. An effective method for dealing with problems of this type is known variously as sequential Monte Carlo [*Doucet and Johansen*, 2009], sequential importance (re-)sampling, bootstrap filtering [*Gordon et al.*, 1993], particle filtering, or the Condensation Algorithm[*Blake and Isard*, 1997]. We will refer to the method as particle filtering for the remainder of this work. As the name implies, we rely upon a random sample of feasible states ('particles') that are sequentially updated as new observations become available. The central assumption is that a probability distribution $P(\mathbf{m})$ can be represented as a weighted set $\{(\mathbf{m}^i, w^i) : i \in 1, \ldots, N\}$ of random samples, forming a new probability mass function

$$P(\mathbf{m}) \approx \sum_{i=1}^{N} w^i \delta(\mathbf{m} - \mathbf{m}^i), \tag{4.17}$$

where $\delta(\cdot)$ is the Dirac delta function, and N is the number of random samples. As N increases, the quality of the approximation increases. The random samples are drawn from a proposal density $q(\mathbf{m}_k^i|\mathbf{m}_{k-1}^i, \mathbf{d}_k)$, and the weights are computed as

$$w_k^i \propto w_{k-1}^i \frac{P(\mathbf{d}_k|\mathbf{m}_k^i)P(\mathbf{m}_k^i|\mathbf{m}_{k-1}^i)}{q(\mathbf{m}_k^i|\mathbf{m}_{k-1}^i, \mathbf{d_k})}, \tag{4.18}$$

which are subsequently normalized. The proposal distribution is arbitrary, but some choices are better than others with respect to capturing the posterior distribution with a minimum number of samples. The practical and intuitive choice is that the proposal distribution should be the prior distribution at time k

$$q(\mathbf{m}_k^i|\mathbf{m}_{k-1}^i, \mathbf{z}_k) = P(\mathbf{m}_k^i|\mathbf{m}_{k-1}). \tag{4.19}$$

While this method ostensibly captures the posterior distribution given enough particles, the diffusive nature of the state equations means that eventually very few particles will be probable with respect to observations. We overcome this by resampling at each time step from the samples $\mathbf{m}^i$ with probability given by w^i using a systematic resampling method [*Carpenter et al.*, 1999].

Resampling produces the same distribution with particles now distributed proportional to the weights, which are then reset to $1/N$. Weights are now simply proportional to the likelihood

$$w_k^i \propto P(\mathbf{d}_k|\mathbf{m}_k^i). \tag{4.20}$$

The resulting distribution converges to the true posterior probability

$$P(\mathbf{m}_k|\mathcal{D}_k) = \lim_{N\to\infty} \sum_{i=1}^{N} w_k^i \delta(\mathbf{m}_k - \mathbf{m}_k^i), \tag{4.21}$$

proof of which can be found in *Blake and Isard* [1997].

4.4 Application to Columbia Glacier

We apply the above algorithm to two cameras, dubbed AK01 and AK10 (Fig. 4.1) from 2013.06.10 through 2013.09.25 and from 2013.11.06 through 2014.4.30, during which both cameras operated continuously at 20 minute intervals. AK01, a Nikon D200, recorded 5.8 megapixel images in JPEG format at quality level 99. AK10, a newer D200, recorded 10 megapixel images at quality level 92. The break between camera epochs naturally leads to a ‘summer’ and ‘winter’ record, and we divide our analysis along those lines.

To specify the surface elevation $S(\mathbf{x}, t)$, we linearly interpolate (in time) between the nearest two members of a set of 10m-resolution digital elevation models derived from the TanDEM-X satellite [*Vijay and Braun*, 2017], to which we fit a 3rd order spline for sub-pixel interpolation. DEMs at this level of precision capture transient crevasse features. To account for this, we sequentially apply a maximum filter followed by a Gaussian smoother over the glacierized area, each with a 30m kernel, which has the practical effect of ‘filling’ crevasses. More complex approaches to surface processing are possible [*Messerli and Grinsted*, 2014], but experimentation has shown the results to be relatively insensitive to the smoothing method.

We specify initial locations of points to track as the vertices of a grid with 100m spacing in both map-plane coordinates, so long as those points are within the field of view of both cameras, and the elevation is more than 20 m above sea level (Vectors in Fig. 4.7 correspond to these points). We start the algorithm for each day in the record period at noon local time (20:00 UTC), and track for 3 days. The algorithm works equally well running backwards in time, so we also track the point backwards starting at the final image of the forward run. We take the mean of the two resulting velocities weighted by the inverse of the Frobenius norm of the sample covariance (i.e. cases in which the last or first image has bad weather has high covariance so contributes little to the mean).

We must make a few choices regarding process noise, and these are mostly informed by heuristics. We assume that random accelerations have a standard deviation of approximately 2 m d^{-2} in both directions. This is based on the characteristic velocity variations observed from high resolution surveys [*Meier et al.*, 1994; *Krimmel and Rasmussen*, 1986], and also so that the model has the potential to capture the abrupt slowdowns sometimes observed at Columbia Glacier in the fall (O'Neel, unpublished data). We specify the standard deviation of local slope as $\sigma_z = 0.1$.

We use a square reference image with size $m_r = n_r = 15$ pixels, and a test image size of $m_t = n_t = 25$ pixels, which assumes a maximum search distance of 10 pixels. Note that this is a form of prior information that we are explicitly introducing into the results: we assume that the probability that a particle moves more than 10 pixels in image space is zero. In practice, this turns out to be a reasonable assumption, and serves to limit spurious correlations and relieve computational effort. We assume $\sigma_t = 0.25$, which implies that a correlation peak is localized to within a 1 pixel range with 2σ credibility. Values of σ_m vary between images, but when matching two images in good weather under similar illumination, the value is close to zero, while unfavorable conditions where many points are occluded will produce uncertainties of greater than $\sigma_m = 5$ pixels, which effectively means that the likelihood is uniform across the test image.

We approximate the probability distribution at each point we wish to track with $N = 3000$ samples. This is probably overkill; however, since our application does not need to be run in real time, the improved convergence associated with using many particles is worth the increased computational costs. Since each sample point is independent, the code is naively parallelizable. The lion's share of memory is taken up by storing many high-resolution images. We utilize a shared read-only memory structure, so that computation for each point in the grid can draw upon the same location in memory. Otherwise each grid point is independent. On a laptop with 8 cores, processing each scene requires around 150 s, depending on the time of year (with maximum and minimum computational times falling on the solstice due to the abundance or dearth of usable images). The processing time would decrease linearly with more cores.

The result is a three-day running average for each initial point at daily resolution, encoded in the form of a collection of sample from which we can draw statistics. While the converged velocity distributions are not normal, they are sufficiently close that they can be described by a mean and covariance matrix. To eliminate outliers and to smooth the resulting fields, we replace each velocity and covariance component at each point with the medians of neighboring points within 150m. Henceforth, when we refer to velocities (and uncertainties), unless otherwise specified, we refer to the distribution mean (and covariance) smoothed in this way.

4.5 Results

4.5.1 Pointwise velocity evolution

The velocity of the near-terminus shows strong temporal variability over seasonal and sub-seasonal time scales (Fig. 4.5). A point near the terminus attains a velocity maximum of nearly 15 m d^{-1} on 2013.06.12, near the start of the record, before decreasing to a minimum of 3 m d^{-1} in late September. During the summer, the glacier also shows a variety of distinct speed-ups. These speed-ups are correlated with warm periods during June, but this correlation is only weakly evident by later in the season. Large rainstorms in August and September produce speedups as well, each lasting around three days. Due to the tendency for the glacier to be occluded during these weather events, they are also the most uncertain.

Following the fall velocity minimum, Columbia Glacier begins to accelerate in mid-November. The speed then remains relatively consistent throughout the winter, with a maximum velocity of around 12 m d^{-1} in February before falling to 10 m d^{-1} throughout March and April. During the winter, the glacier is non-responsive to variations in precipitation in temperature, presumably because temperatures are too cold for liquid water to be present at the glacier surface in large volumes (though plotted temperatures are recorded at 1000 m and are usually several degrees lower than those at the elevation of the glacier terminus). Note that during the winter months, the general level of uncertainty in speed is higher. This is because fewer usable images exist at this time due to less daylight: less data means the flow is less precisely resolved.

4.5.2 Spatial patterns of velocity change

In the nearest 2.5 km from the terminus along the centerline, Columbia Glacier shows relatively small spatial variability in speed (Fig. 4.6a). Nonetheless, because the baseline speed at Columbia Glacier is so fast, these small variations are still sufficient to produce large strains and the evident extensive crevassing. A slight acceleration in the lowest 500 m seems to be transient, and we could detect no specific reason for why this occurs. Changes in the stress regime due to individual calving events could be a factor, but previous work indicating velocity changes over this scale have been for a floating terminus [*Murray et al.*, 2015; *Ahn and Box*, 2010], whereas Columbia Glacier's terminus is grounded over the period in question.

Looking at the glacier in cross-section, a migrating shear margin is evident. Ice near the edge

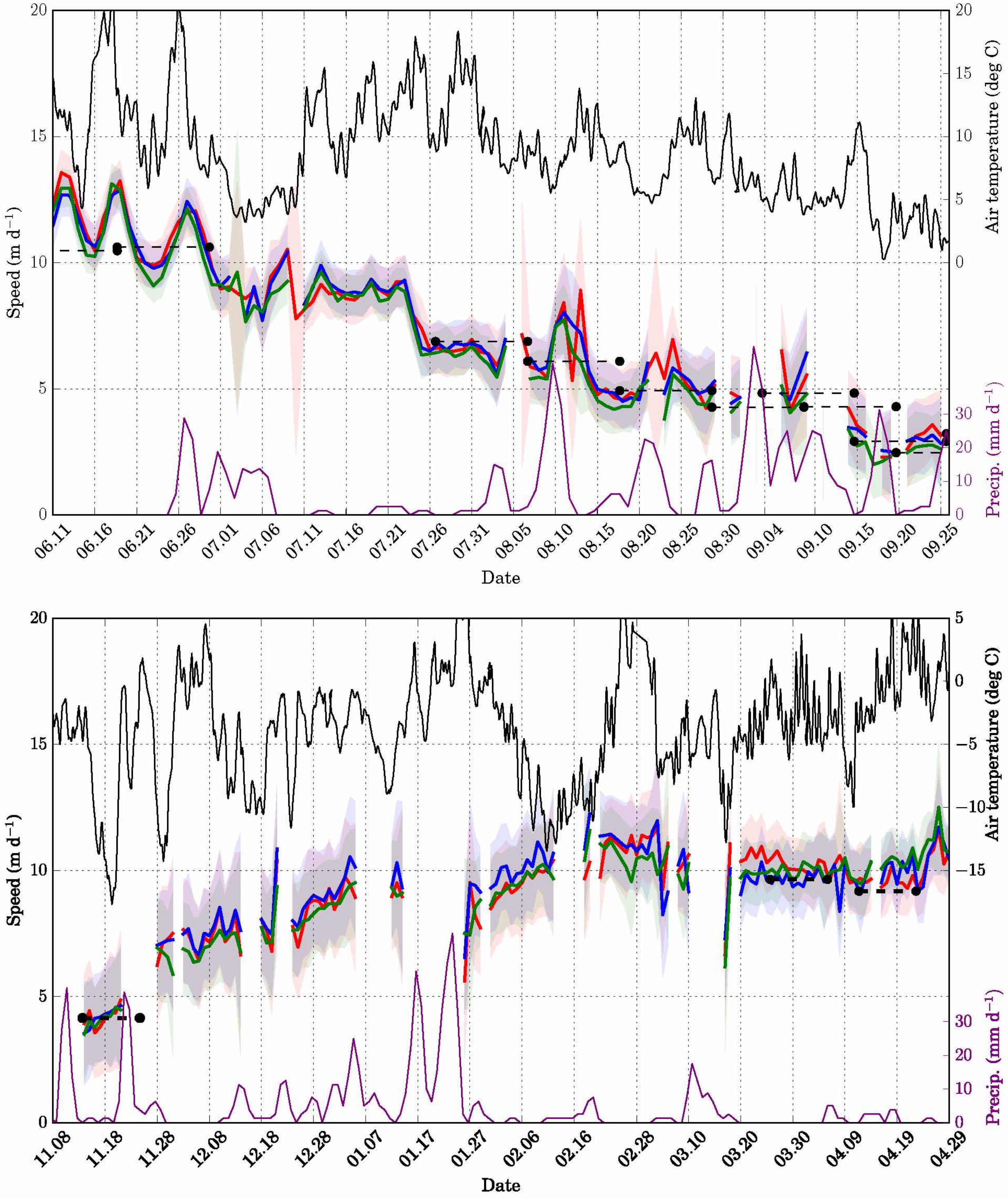

Figure 4.5. Time series of ice speed at points corresponding to line color in Fig. 4.7 in summer 2013 (top) and winter 2013-2014 (bottom). Black barbells are TSX speeds over the time periods indicated by the line endpoints. The black line in the upper portion of the plot is air temperature at a weather station approximately 5 km upstream from the Columbia Glacier terminus and at an elevation of $\tilde{1}000$ m, while the purple line is precipitation rate recorded in the village of Tatitlek, near the mouth of Columbia Bay.

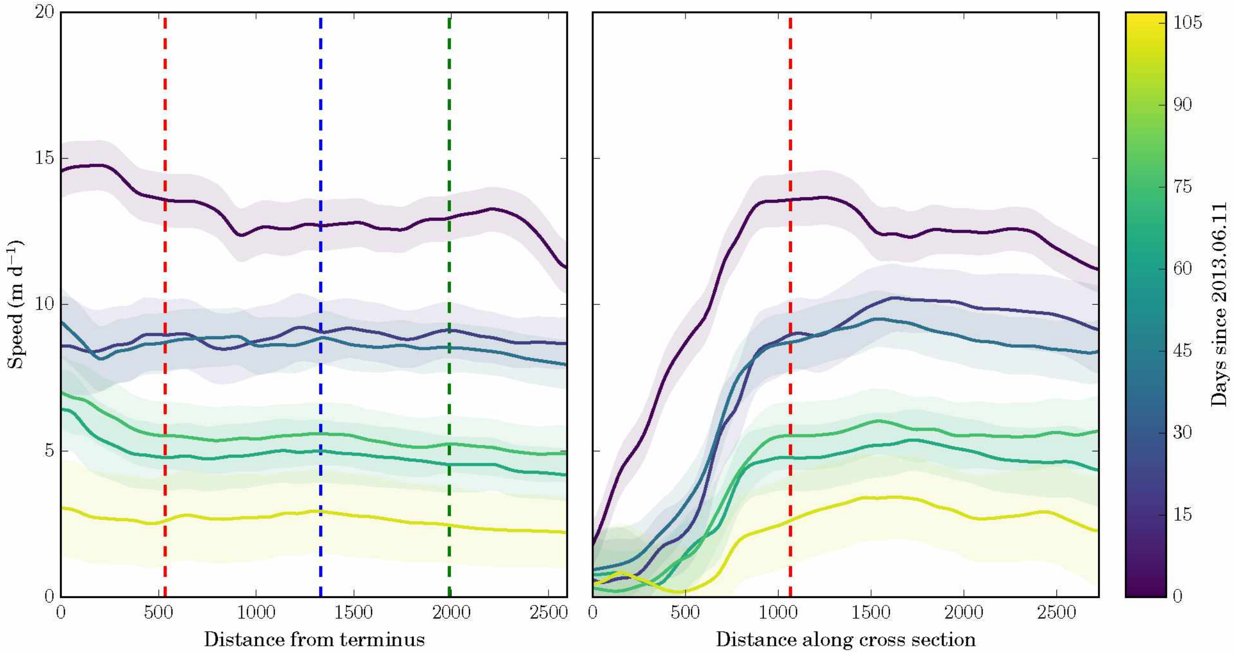

Figure 4.6. Along-flow (left) and across-flow (right) speed profiles during summer 2013, with colored lines representing the location at which the time-series in Fig. 4.5 were extracted. Note the relative homogeneity of speeds in the along-flow direction, which are mostly constant except for a ephemeral acceleration within 500 m of the terminus. In the cross-sections, note the activation and deactivation of marginal ice between 0 and 500 m.

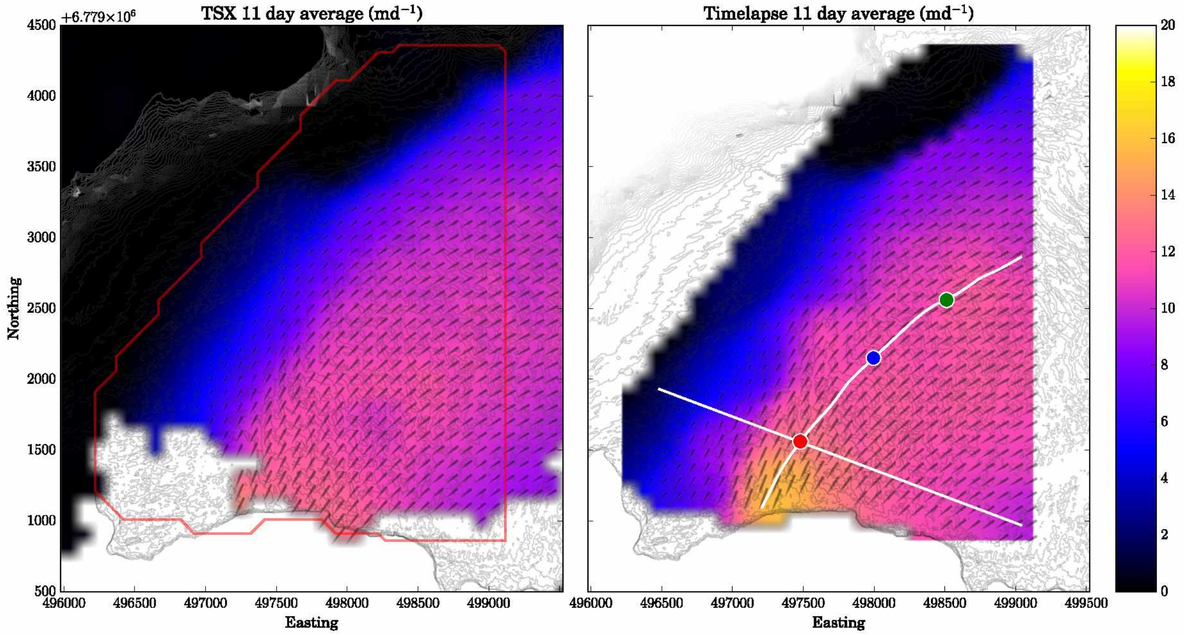

Figure 4.7. Comparison of a TSX-derived velocity field covering 2013.06.10 through 2013.06.21 (left) to ones produced using this method averaged over the same period (right). The colored dots are the locations at which the lines in Fig. 4.5 are extracted, and the white lines correspond to the long profile and cross-section of Fig. 4.6. The distinct shear margin in the northern part of the field corresponds to the one visible in Fig. 4.2.

of the glacier is activated or deactivated based on fast flow in the center. Extrapolating from the edge of the data, at its fastest rate in mid-June, the entire glacier width is in motion. Later in the year, when centerline speeds have dropped to 20% of maximum, a nearly 1000 m wide strip of marginal ice become stagnant.

4.5.3 Validation

It is necessary to confirm that the method is producing results at least consistent with other contemporaneous observations, if not with reality. To this end, we compare the time average of the daily velocity computed with the particle filtering method over the same epoch as a velocity field derived from the radar interferometer TerraSAR-X (TSX) [*Joughin et al.*, 2010]. Fig. 4.7 shows a TSX velocity field taken for 2013.06.10 through 2013.06.21 and the stacked velocity field derived for time-lapse for the same period. The qualitative agreement between the two fields is excellent, with the shear margin in the northeast corner captured in both, as well as the locations of fast flow.

The correspondence between speeds produced by both methods is shown in Fig. 4.8. A robust linear regression gives a slope of nearly unity, $r^2 = 0.97$, and a bias of 0.7 m d^{-1}, with our particle

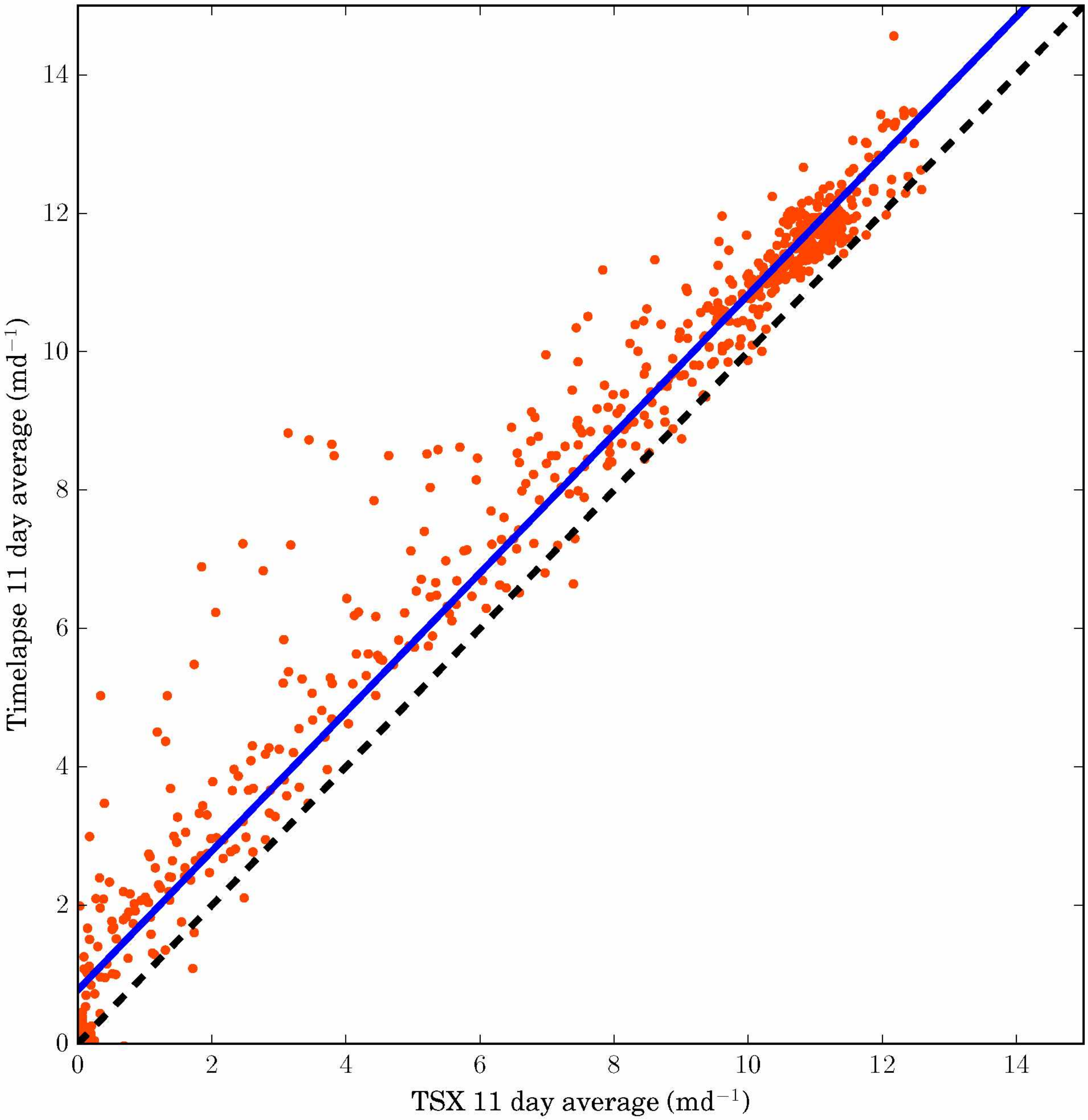

Figure 4.8. The correspondence between the speeds of Fig. 4.7. The black dashed line is 1:1. The blue line is the best fit line from a robust regression, with slope close to 1 and a bias of 0.7 m d^{-1}. Other epochs show a similar level of agreement.

filtering method producing the faster speeds. The average standard deviation in the time-lapse speed over the averaging period is $\sigma_V = 1.7$m d^{-1}, and computed standard deviations encompass the TSX velocities in nearly all cases. Reported uncertainties for TSX range between 1–10m d^{-1}, and similarly encompass the time-lapse speeds. Results are similar for all epochs for which optical conditions allow for good velocity solutions. The good fidelity between datasets lends confidence to the method for subsequent interpretation.

It is also interesting to examine the differences between the two velocity fields. The particle filtering method shows a patch of fast ice in the ~200 m upstream from the terminus. Given the high velocities in this region, it would make sense that this speedup is not captured by SAR: trackable surface features would be calved off in the eleven-day interval between satellite images, while the higher temporal resolution offered by time-lapse can track features much nearer the calving front.

As an additional validation measure, TSX velocities, when they exist, are overlain in Fig. 4.5. In general, there is excellent agreement, but with our method showing considerably more variance than can be captured in the 11-day averages of TSX. This variance is due to the additional resolution granted by analyzing time-lapse imagery, which allows us to capture speed changes associated with short term changes in water input or perhaps individual calving events [*Ahn and Box*, 2010].

4.6 Discussion

4.6.1 Implications for the subglacial hydrologic system

Our results strongly support the conclusions of previous studies that suggest that annual and sub-annual velocity variations are driven by the evolution of the subglacial hydrologic system in response to changes in external input in the form of either melt or rain [*Kamb et al.*, 1994; *Fahnestock*, 1991; *Vijay and Braun*, 2017]. The seasonality evident in our results is consistent with observations throughout Columbia's retreat history, and we support the following explanation for it.

During mid-winter, external water input to the glacier base is likely small, particularly given the availability of copious surface snow to absorb moisture, even in the few cases where winter temperatures climb above freezing. The hydrologic system at this time is in a dormant state where basal water pressures are governed only by basal melt, and the ocean, although remnant water from the ongoing drainage of firn may also contribute. The tidewater influence is important: baseline hydraulic head is still a significant fraction of flotation, and this allows the glacier to maintain a minimum flow speed an order of magnitude greater than the maximum speeds of most land-terminating glaciers. Nonetheless, because there is little water flux, efficient drainage elements such as Rothlisberger channels [*Röthlisberger*, 1972] likely cannot persist because there is insufficient turbulent heat dissipation to maintain them. Thus we echo *Kamb et al.* [1994] and hypothesize that the winter state is characterized by an inefficient drainage system that lead to moderately high water pressure and flow speed. We would anticipate that velocities during winter would be sensitive to the odd winter water pulse (e.g. a so-called pineapple express where sub-tropical air and moisture are transported to northerly latitudes). However, such events in 2013–2014 were invariably accompanied by severe camera occlusion, and we could not discern velocities during this time period with any certainty. Nonetheless, there is some evidence of short lived slow-downs following large precipitation events in January and March.

At the onset of melting in the spring, the subglacial drainage system is overwhelmed by the availability of water. Since the drainage system is still inefficient, additional input must be accommodated by an increase in water pressure, which leads to fast flow. This sensitivity is shown clearly during June of 2013 (See Fig. 4.5a), during which the velocity and air temperature (a proxy for melt) are highly correlated. This strong correlation is short-lived, however. We propose that as the subglacial drainage system develops, more water is required to maintain high velocities. The excess water from subsequent warm spells loses its impact come July, with speeds decreasing throughout the summer. Interestingly, this slowdown does not appear to occur gradually; rather it decreases in discrete steps, the most distinct of which is seen around 2013.07.24. This event is particularly interesting because there seems to be no discernible external forcing, although it is worth noting that changes in melt rates may be independent of temperature [*Sicart et al.*, 2008]. Temperatures had been high but not anomalous, and precipitation was minimal. It seems possible that this could be triggered by a sudden change in the transmissivity of the subglacial hydrologic system, such as the confluence of two large Rothlisberger channels, but this is purely speculative.

Throughout late summer, major rainfall events drive short-lived (~ 3 day) speed peaks, though it is difficult to be sure of the magnitude of these peaks due to occlusion of the camera by the same precipitation that produced the speed-up. However, two events in August were observed with reasonable precision. A simple calculation of melt using a temperature index model [*Hock*, 2003] implies that meltwater and rainwater flux during these large rain events were both around 40 mm of water equivalent per day, though this likely underestimates melt due to latent heat transfer from precipitation. This combined influx produced speed-ups that were slightly lower in magnitude to the purely melt-induced speed-ups of the early season. The modest speedups relative to enhanced influx indicates that the subglacial drainage system has evolved to efficiently transmit the extra water. In the long term, rain events may enhance the process of seasonal slowdown. After the rain event of 2013.08.08-2013.08.12, velocities dropped by 30% compared to the pre-storm average, though as discussed above, such a drop is not necessarily associated with external forcing. In any case, these slowdowns are consistent between tidewater glacier systems, but not typically evident in terrestrially terminating glaciers. This may be caused by the lower limit on water pressure imposed by the marine margin, which keeps conduit pressure high enough to resist fast closure [*Podrasky et al.*, 2012].

In early fall, the situation is opposite to that of mid-winter. The drainage system is well-developed, but as temperatures drop and rain transitions to snow, influx becomes small. What little water is available is drained through an efficient hydrologic system kept open by the background water pressure imposed by sea level. Eventually, as creep closure destroys this efficient drainage system and slow flow keeps subglacial storage capacity to a minimum [*Bartholomaus et al.*, 2011], water pressure and speeds recovers to winter levels, which seems to be complete by approximately January.

4.6.2 Improvements and extensions

As with all numerical methods, there are a number of ways in which the algorithm's performance can be improved. First, the problem of image registration and co-registration yields a source of systematic uncertainty that is difficult to quantify. Cameras tend to be placed such that glaciers take up most of the field of view. Unfortunately, glaciers are useless for the purpose of establishing ground control for the calibration of camera models. Instead, ground control is established by attempting to match features in error-prone DEMs to recognizable features in images that are typically very far (e.g. mountain peaks) or very near. Variable atmospheric distortion compounds these errors. Additionally, manual digitization of ground control points is labor-intensive and subject to picking errors.

A potential method for circumventing these difficulties is to utilize automated detection of horizons and other strong image edges, and minimize the difference between these and horizons computed from a digital elevation model. Such a method could be used simultaneously to find static camera parameters such as focal length and lens distortion and also time-varying camera orientation [*Baboud et al.*, 2011], without any human intervention.

Second, an adaptive procedure for selecting reference images could be included. As it stands, we track a sub-image extracted from the first image in a sequence, irrespective of the fact that it could be occluded, which obviously leads to failure (fortunately this failure is reflected in uncertainty estimates). Instead, we could compare the statistical properties of each image in a sequence in order to find one that is of good quality for tracking. The method would still fail if too many images in a sequence were occluded, but it may help fill some of the gaps evident in Fig. 4.5. However, we would also not know the location of the selected sub-image at the beginning of the sequence, and a more carefully selected prior would be needed to ensure that the image correlation procedure does not become lost. One potential choice is to use the posterior from the last epoch in which the velocity fell below a certain error threshold, or to initialize the algorithm with SAR-derived velocities.

Finally, while the results presented here rely on the use of two cameras in order to properly specify flow directions, those directions deviated little throughout the observation period despite large changes in speed. This suggests that we could extend the observational record to include epochs in which only one camera was active by specifying mean flow direction as a strong prior on the computed velocities. While flow directions are likely to change substantially in the long term, this could be extremely useful for filling gaps, such as the one between 2013.09.25 and 2013.11.06 that exists in the data presented here, wherein AK10 failed while the other camera continued to take images uninterrupted.

4.6.3 Conclusions

We developed a probabilistic method for tracking glacier surface motion based on time-lapse imagery. The method operates by evolving a set of particles according to a stochastic dynamical model, while culling particles that are improbable and reproducing probable ones, with the likelihood determined by computing the sum of squared differences between a reference image and test image. The resulting set of solutions converges to the true posterior distribution of glacier velocity at a given point. The method is robust to occlusion and false matching, provides rigorous uncertainty, and easily accommodates the refinement of velocity measurements with the use of multiple cameras.

We apply the developed method to just under a year's worth of images collected by two cameras near the terminus of Columbia Glacier between 2013.06 and 2014.05. Based on image geometry and distances, the method was able to extract three day running-average velocities over all time periods during which the glacier surface was visible. To ensure that the resulting velocity fields were valid, we compared TerraSar-X derived velocities to quantities computed with this method and averaged over the same temporal footprint, finding excellent agreement over the entire record.

At seasonal time scales, our findings mirror those of previous workers in showing that Columbia Glacier transitions between a winter state characterized by moderate velocities, to an early summer speed-up, to a fall slowdown, in which velocities drop to well below their winter state before eventually recovering in early winter. Our method resolves velocity correlations with melt and rainfall events, though the glacier's sensitivity to these events appears seasonal: during the spring, before an efficient drainage network has developed the glacier sees strong melt- and rainwater induced variability, while in the fall the system responds very little to these forcings. The velocity fields produced here may help to constrain future simulations of tidewater glacier hydrology and the resulting changes in ice dynamics.

Bibliography

Ahn, Y., and J. E. Box (2010), Glacier velocities from time-lapse photos: technique development and first results from the Extreme Ice Survey (EIS) in Greenland, *Journal of Glaciology*, *56*(198), 723–734.

Baboud, L., M. Čadík, E. Eisemann, and H.-P. Seidel (2011), Automatic photo-to-terrain alignment for the annotation of mountain pictures, in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 41–48, IEEE.

Bartholomaus, T. C., R. S. Anderson, and S. P. Anderson (2011), Growth and collapse of the distributed subglacial hydrologic system of Kennicott Glacier, Alaska, USA, and its effects on basal motion, *Journal of Glaciology*, *57*(206), 985–1002.

Bieniek, P. A., U. S. Bhatt, R. L. Thoman, H. Angeloff, J. Partain, J. Papineau, F. Fritsch, E. Holloway, J. E. Walsh, C. Daly, et al. (2012), Climate divisions for Alaska based on objective methods, *Journal of Applied Meteorology and Climatology*, *51*(7), 1276–1289.

Blake, A., and M. Isard (1997), The CONDENSATION algorithm-conditional density propagation and applications to visual tracking, in *Advances in Neural Information Processing Systems*, pp. 361–367.

Burgess, E. W., R. R. Forster, and C. F. Larsen (2013), Flow velocities of Alaskan glaciers, *4*.

Carlson, A. E., Z. Kilmer, L. B. Ziegler, J. S. Stoner, G. C. Wiles, K. Starr, M. H. Walczak,

W. Colgan, A. V. Reyes, D. J. Leydet, et al. (2017), Recent retreat of Columbia Glacier, Alaska: Millennial context, *Geology*, *45*(6), 547–550.

Carpenter, J., P. Clifford, and P. Fearnhead (1999), Improved particle filter for nonlinear problems, *IEE Proceedings - Radar, Sonar and Navigation*, *146*(1), 2–7, doi:10.1049/ip-rsn:19990255.

Claus, D., and A. W. Fitzgibbon (2005), A rational function lens distortion model for general cameras, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 213–219, IEEE.

Dietrich, R., H.-G. Maas, M. Baessler, A. Rülke, A. Richter, E. Schwalbe, and P. Westfeld (2007), Jakobshavn Isbræ, West Greenland: Flow velocities and tidal interaction of the front area from 2004 field observations, *Journal of Geophysical Research: Earth Surface*, *112*(F3).

Doucet, A., and A. M. Johansen (2009), A tutorial on particle filtering and smoothing: Fifteen years later, *Handbook of nonlinear filtering*, *12*(656-704), 3.

Evans, A. N. (2000), Glacier surface motion computation from digital image sequences, *IEEE Transactions on Geoscience and Remote Sensing*, *38*(2), 1064–1072.

Fahnestock, M., T. Scambos, T. Moon, A. Gardner, T. Haran, and M. Klinger (2016), Rapid large-area mapping of ice flow using Landsat 8, *Remote Sensing of Environment*, *185*, 84–94.

Fahnestock, M. A. (1991), Hydrologic control of sliding velocity in two Alaskan glaciers: observation and theory, Ph.D. thesis, California Institute of Technology.

Fischler, M. A., and R. C. Bolles (1981), Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, *24*(6), 381–395.

Flotron, A. (1973), Photogrammetrische Messung von Glescherbewgungen mit autmatischer Kamera, *Schweizerisch Zeitschrift f ur Vermessung, Kulturtchnik, und Photogrammetrie*, *1*(73).

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993), Novel approach to nonlinear/non-Gaussian Bayesian state estimation, in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113, IET.

Harper, J. T., N. F. Humphrey, W. T. Pfeffer, and B. Lazar (2007), Two modes of accelerated glacier sliding related to water, *Geophysical Research Letters*, *34*(12), n/a–n/a, doi:10.1029/2007GL030233, l12503.

Harrison, W., C. Raymond, and P. MacKeith (1986), Short period motion events on Variegated Glacier as observed by automatic photography and seismic methods, *Annals of Glaciology*, *8*(1), 82–89.

Harrison, W., K. Echelmeyer, D. Cosgrove, and C. Raymond (1992), The determination of glacier speed by time-lapse photography under unfavorable conditions, *Journal of Glaciology*, *38*(129), 257–265.

Hock, R. (2003), Temperature index melt modelling in mountain areas, *Journal of hydrology*, *282*(1), 104–115.

Iken, A., and M. Truffer (1997), The relationship between subglacial water pressure and velocity of Findelengletscher, Switzerland, during its advance and retreat, *Journal of Glaciology*, *43*(144), 328–338.

James, M. R., P. How, and P. M. Wynn (2016), Pointcatcher software: analysis of glacial time-lapse photography and integration with multitemporal digital elevation models, *Journal of Glaciology*, *62*(231), 159–169.

James, T. D., T. Murray, N. Selmes, K. Scharrer, and M. O'leary (2014), Buoyant flexure and basal crevassing in dynamic mass loss at Helheim Glacier, *Nature Geoscience*, *7*(8), 593.

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

Kamb, B., H. Engelhardt, M. A. Fahnestock, N. Humphrey, M. Meier, and D. Stone (1994), Mechanical and hydrologic basis for the rapid motion of a large tidewater glacier: 2. Interpretation, *Journal of Geophysical Research: Solid Earth*, *99*(B8), 15,231–15,244.

Krimmel, R., and L. Rasmussen (1986), Using sequential photography to estimate ice velocity at the terminus of Columbia Glacier, Alaska, *Annals of Glaciology*, *8*(1), 117–123.

Krimmel, R. M. (2001), Photogrammetric data set, 1957-2000, and bathymetric measurements for Columbia Glacier, Alaska, *Tech. rep.*, US Geological Survey.

Lowe, D. G. (1999), Object recognition from local scale-invariant features, in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, IEEE.

Meier, M., and A. Post (1987), Fast tidewater glaciers, *Journal of Geophysical Research: Solid Earth*, *92*(B9), 9051–9058.

Meier, M., S. Lundstrom, D. Stone, B. Kamb, H. Engelhardt, N. Humphrey, W. W. Dunlap, M. Fahnestock, R. M. Krimmel, and R. Walters (1994), Mechanical and hydrologic basis for the rapid motion of a large tidewater glacier: 1. Observations, *Journal of Geophysical Research: Solid Earth*, *99*(B8), 15,219–15,229.

Messerli, A., and A. Grinsted (2014), Image georectification and feature tracking toolbox: ImGRAFT, *Geoscientific Instrumentation, Methods and Data Systems Discussions*, *4*, 491–513.

Murray, T., N. Selmes, T. D. James, S. Edwards, I. Martin, T. O'Farrell, R. Aspey, I. Rutt, M. Nettles, and T. Baugé (2015), Dynamics of glacier calving at the ungrounded margin of Helheim Glacier, southeast Greenland, *Journal of Geophysical Research: Earth Surface*, *120*(6), 964–982.

Nakhmani, A., and A. Tannenbaum (2008), Particle filtering using multiple cross-correlations for tracking occluded objects in cluttered scenes, in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pp. 652–657, IEEE.

O'Neel, S., W. T. Pfeffer, R. Krimmel, and M. Meier (2005), Evolving force balance at Columbia Glacier, Alaska, during its rapid retreat, *Journal of Geophysical Research: Earth Surface*, *110*(F3).

Pfeffer, W. (2007), A simple mechanism for irreversible tidewater glacier retreat, *Journal of Geophysical Research: Earth Surface*, *112*(F3).

Podrasky, D., M. Truffer, M. Fahnestock, J. M. Amundson, R. Cassotto, and I. Joughin (2012), Outlet glacier response to forcing over hourly to interannual timescales, jakobshavn isbræ, greenland, *Journal of Glaciology*, *58*(212), 1212–1226.

Powell, M. J. (1964), An efficient method for finding the minimum of a function of several variables without calculating derivatives, *The Computer Journal*, *7*(2), 155–162.

Rosenau, R., E. Schwalbe, H.-G. Maas, M. Baessler, and R. Dietrich (2013), Grounding line migration and high-resolution calving dynamics of Jakobshavn Isbræ, West Greenland, *Journal of Geophysical Research: Earth Surface*, *118*(2), 382–395.

Röthlisberger, H. (1972), Water pressure in intra-and subglacial channels, *Journal of Glaciology*, *11*(62), 177–203.

Sicart, J. E., R. Hock, and D. Six (2008), Glacier melt, air temperature, and energy balance in different climates: The bolivian tropics, the french alps, and northern Sweden, *Journal of Geophysical Research: Atmospheres*, *113*(D24).

Sikonia, W. G., and A. Post (1980), Columbia Glacier, Alaska recent ice loss and its relationship to seasonal terminal embayments, thinning and glacial flow, *Tech. rep.*, report.

Smith, L. I., et al. (2002), A tutorial on principal components analysis, *Cornell University, USA*, *51*(52), 65.

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, SIAM.

Truffer, M., and W. D. Harrison (2006), In situ measurements of till deformation and water pressure, *Journal of Glaciology*, *52*(177), 175–182.

Vaughn, B., C. Raymond, L. A. Rasmussen, D. Miller, C. Michaelson, M. Meier, R. Krimmel, A. Fountain, W. Dunlap, and C. Brown (1985), Short-term velocity measurements at Columbia Glacier, Alaska; August-September 1984, *Tech. rep.*, US Geological Survey,.

Vijay, S., and M. Braun (2017), Seasonal and interannual variability of Columbia Glacier, Alaska (2011–2016): Ice velocity, mass flux, surface elevation and front position, *Remote Sensing*, *9*(6), 635.

Werder, M. A., I. J. Hewitt, C. G. Schoof, and G. E. Flowers (2013), Modeling channelized and distributed subglacial drainage in two dimensions, *Journal of Geophysical Research: Earth Surface*, *118*(4), 2140–2158.

Chapter 5

Conclusions

Bayesian sampling offers a powerful technique for inferring the unknown parameters that control glacier dynamics from surface observations. Such techniques offer an advantage over other inverse methods by producing the complete probability distribution of model configurations. This information can be used to estimate uncertainty, estimate covariance between parameters, and to ensure uniqueness. The Bayesian methodology also allows for the specification of complex and intuitive error structures that have no meaningful analog in deterministic inversion, as well as for a wide variety of prior constraints, including smoothness, positivity, and dynamical constraints based on physical principles. The cost for these advantages is increased computational complexity as we try to recover a high-dimensional function from a finite number of samples.

In Chapter 2, we applied the Metropolis-Hastings algorithm to the problem of inferring properties of the glacial hydrologic system from observations of surface speed and terminal water flux. We found that the mechanical opening of linked cavities due to glacier sliding over a rough bed is similar in magnitude to the thermal opening due to the turbulent dissipation of heat by flowing water. Contrary to previous work, we found it to be highly likely that significant amounts of water are stored englacially, buffering changes in pressure due to changes in water storage. The majority of uncertainty in model results was attributed to model sensitivity to the parameters that control the rate of turbulent dissipation. Observations of the parameters controlling this process like the hydraulic gradient and subglacial cavity size would go a long way towards constraining glacier hydrology models.

In Chapter 3, we again used the Metropolis-Hastings algorithm, this time sampling from the probability distribution of ice thickness for several test cases. For the first time, we presented a consistent framework for understanding the problem of so-called mass conserving interpolation, in particular addressing the heretofore issue of implicit time averaging. This averaging produces additional uncertainty that must be included in any error budget. We found that the relative uncertainty in ice thickness depends almost linearly on the relative uncertainty in both surface velocity and surface mass balance observations. The presence of local observations of the ice thickness reduces the strength of this dependency, particular when observation spacing becomes less than the characteristic length of topographic variability. With this information in mind, we were able to suggest an optimal measurement spacing for future radar campaigns such as NASA's Operation IceBridge.

In Chapter 4, we used particle filtering to find estimates of glacier surface velocity from oblique time-lapse photography. The method worked by proposing a large number of potential solutions for the motion and position of a certain location on the glacier surface, and then filtering solutions that were unlikely. The likelihood was specified by extracting a sub-image around the initial position of the solution ensemble projected into image coordinates, and searching for that sub-image in subsequent images. By design, our method was robust to temporary occlusion of the glacier surface by bad weather and to bad matches between images. We applied the method to Columbia Glacier over a year's worth of images collected at half-hourly resolution, and were able to extract velocities with daily resolution. We found that Columbia Glacier exhibits strong seasonal variability, upon which is superimposed shorter-term responses to individual melt and rainfall events. These velocity variations were consistent with the seasonal growth and collapse of an efficient subglacial drainage system.

Outlook

The methods presented in this thesis offer a blueprint for how to apply Bayesian methods to glaciological models. The easiest way that this work might be extended is to apply identical methods and models to new glaciers and observational datasets. This more operational context would help us to understand the degree to which the results demonstrated in this thesis, particularly Chapters 2 and 3, might be widely applicable. This would also help to identify shortcomings in choices of physical and statistical models: we can learn about important processes by recognizing that their absence makes matching a model to observations impossible. The next possibility for applying Bayesian methods to glaciology is to extend the methods here to more complicated instances of the same model problem. For example, we have already extended the methods of Chapter 3 to the two-dimensional case. It seems similarly plausible to apply the methods of Chapter 2 to more complex glacier hydrology models that are spatially distributed and operate over varying spatial scales. The methods of Chapter 4 could also be used to constrain more complicated models of surface velocity. For example, rather than the simple pointwise constant-velocity model featured in this thesis, we could use time-lapse photography to constrain a distributed ice flow model.

The three problems presented here are by no means the only application of Bayesian methods in glaciology or in geophysics at large. Indeed, they have utility in any case where we need to make an inference about an observable quantity given imperfect observations, and in particular where the uniqueness and uncertainty in the inference might have qualitative importance. As data about the natural world becomes available at an ever faster rate, efforts to use these data to inform unobserved (and perhaps unobservable) processes will also increase. Modern satellite observations are already being used to automatically generate solutions to variants of two of the problems addressed in this thesis. Glaciological inversion has gone beyond being commonplace; it now operational. Such extensive capabilities have the potential to be transformative. However, we must be careful not to let abundance overwhelm caution. These inverse problems are still ill-posed and error-prone, and likely come with a high degree of uncertainty, and we need to be ensure that this is correctly quantified. Bayes' theorem offers a consistent methodology for doing this.