Spatio-temporal statistical models for glaciology

Giridhar (Giri) R. Gopalan

Faculty of Physical Sciences
University of Iceland
2019

Spatio-temporal statistical models for glaciology

Giridhar (Giri) R. Gopalan

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Statistics

Advisor
Dr. Birgir Hrafnkelsson

PhD Committee
Dr. Guðfinna Aðalgeirsdóttir
Dr. Birgir Hrafnkelsson
Dr. Christopher K. Wikle

Opponents
Dr. Oksana Chkrebtii
Dr. Hilmar Gudmundsson

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, August 2019

Spatio-temporal statistical models for glaciology
Glacial statistics
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Statistics

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Dunhagi 5
107, Reykjavik
Iceland

Telephone: +354-525-4700

Bibliographic information:
Giridhar (Giri) R. Gopalan, 2019, *Spatio-temporal statistical models for glaciology*, PhD dissertation, Faculty of Physical Sciences, University of Iceland, 120 pp.

ISBN 978-9935-9452-2-8

Printing: Háskólaprent
Reykjavik, Iceland, August 2019

Abstract

The purpose of this thesis is to develop spatio-temporal statistical models for glaciology, using the Bayesian hierarchical framework. Specifically, the process level is modeled as a time series of computer simulator outputs (i.e., from a numerical partial differential equation solver or an emulator) added to an error-correcting statistical process, closely related to the concept of model discrepancy. This error-correcting process accounts for spatial variability in simulator inaccuracies, as well as the accumulation of simulator inaccuracies forward in time. For computational efficiency, linear algebra for bandwidth-limited matrices is used for evaluating the likelihood of the model, and first-order emulator inference allows for the fast approximation of numerical solvers. Additionally, a computationally efficient approximation for the likelihood is derived. Analytical solutions to the shallow ice approximation (SIA) of the full Stokes equation system for stress balance of ice are used to examine the speed and accuracy of the computational methods used, in addition to the validity of modeling assumptions. Moreover, the modeling and methodology within this thesis are tested on data sets collected by the University of Iceland Institute of Earth Science (UI-IES) glaciology team, including bi-yearly mass balance measurements at 25 fixed sites at Langjökull (a glacier) over 19 years, in addition to 100 meter resolution digital elevation maps. As a byproduct of the construction of the Bayesian hierarchical model, a novel finite difference method is derived for solving the SIA partial differential equation (PDE). Although the application domain of this work is glaciology, the model and methods developed in this thesis can be applied to other geophysical domains.

The thesis is structured around three papers. The first of these papers reviews dynamical modeling of glacial flow, introduces a second-order finite difference method for solving the SIA PDE, presents a Bayesian hierarchical model involving this numerical solver, and validates the model with analytical solutions to the SIA PDE. The second of these papers generalizes the statistical model of the first paper, probes higher-

order random walks for representing model discrepancy, incorporates first-order emulators, and analyzes methods for efficient log-likelihood evaluation. The third of these papers applies the model framework of the first two papers to mass balance and surface elevation data at Langjökull.

The major contributions of the thesis are the derivation of a new numerical method for solving the SIA PDE in two spatial dimensions and time, the use of a random walk to represent model discrepancy (i.e., an *error-correcting process*), efficient methods for log-likelihood evaluation, and the application of spatio-temporal statistical modeling to Langjökull, one of Iceland's main glaciers.

Útdráttur

Markmið þessarar ritgerðar er að þróa tíma- og rýmisháð tölfræðileg líkön fyrir jökla með því að nota stigskipt Bayesísk líkön. Hér er sá hluti stigskipta Bayesíska líkansins sem snýr að undirliggjandi líkani fyrir ferlið sem er verið að skoða, útfærður þannig að tímaraðirnar sem koma frá tölulegum hermi (þ.e. frá tölulegri lausn á hlutafleiðujöfnu eða nálgun á slíkri tölulegri lausn) eru lagðar saman við líkindafræðilegt ferli sem hefur það hlutverk að leiðrétta fyrir mismuninn á milli tölulega hermisins og raunverulega ferlisins, og er nátengd hugmyndinni um misræmi líkana. Þetta líkindafræðilega ferli leiðréttir fyrir rýmisháð frávik og tekur tillit til að frávikin safnist upp yfir tíma. Til að flýta fyrir útreikningum þá er línuleg algebra fyrir rýr fylki notuð til að reikna sennileikafall líkansins og fyrstu gráðu hermar notaðir til að flýta fyrir útreikningum á tölulegum lausnum hlutafleiðujafna eða öðrum kerfum. Að auki er ný reiknisparandi nálgun fundin fyrir sennileikafallið. Fræðilegar lausnir á þunnjökla nálguninni sem byggir á jöfnum Stokes fyrir spennur í jöklum, eru notaðar til að meta reiknihraða og nákvæmni tölulegra nálgana, og hversu vel líkanið fellur að gögnunum. Að auki er líkönunum og aðferðafræðinni í þessari ritgerð beitt á raunveruleg gagnasöfn sem Jarðvísindastofnun Háskóla Íslands hefur sett saman, þar með taldar afkomumælingar á 22-25 föstum stöðum sem teknar eru tvisvar á ári á Langjökli yfir 19 ára tímabil, auk hæðarkorts sem hefur 100 metra upplausn. Aukaafurð sem kom til við smíði á stigskipta Bayesíska líkaninu, er ný mismunaaðferð til að leysa tölulega hlutafleiðujöfnuna fyrir þunnjökla nálgunina. Þó svo að aðferðirnar sem hér eru settar fram séu fyrir jöklafræði þá má útfæra þær fyrir önnur jarðeðlisfræðileg gögn og tilsvarandi líkön.

Ritgerðin byggir á þremur vísindagreinum. Í fyrstu greininni er farið yfir þau líkön sem hafa verið þróuð til að lýsa hreyfingu jökla, annarar gráðu mismunaaðferð til að leysa tölulegu hlutafleiðujöfnuna fyrir þunnjökla nálgunina er kynnt sem og stigskipt Bayesískt líkan sem notar tölulegu lausnina, og mat byggt á stigskipta Bayesíska líkaninu er borið saman við fræðilega lausn þunnjökla nálgunarinnar. Í grein

tvö er fjallað um nánari útfærslu á tölfræðinni og útreikningunum fyrir stigskipta Bayesíska líkanið, slembigangur af stigi hærra en einn er skoðaður sem líkindafræðilegt líkan fyrir mismuninn á milli tölulega hermisins og raunveruleikans, sýnd notkun á fyrstu gráðu hermum, og ný reiknisparandi nálgun fyrir sennileikafallið er kynnt. Í þriðju greininni er aðferðafræði fyrstu tveggja greinanna beitt á afkomugögn og hæðargögn frá Langjökli.

Framlag ritgerðarinnar felst í: (i) nýrri annarrar gráðu mismunaaðferð til að leysa tölulegu hlutafleiðujöfnuna fyrir þunnjökla nálgunina í tveimur rúmvíddum og tíma, (ii) notkun slembigangs til að lýsa mismuninum á milli tölulega hermisins og raunverulega ferlisins, (iii) nýrri reiknisparandi nálgun fyrir sennileikafallið, (iv) að beita nýju tíma- og rýmisháðu tölfræðilegu líkani fyrir jökla við greiningu gagna frá Langjökli, einum af stærstu jöklum Íslands.

Dedication

This thesis is dedicated to all who have supported my pursuit of scholarship.

Table of Contents

List of Figures

List of Tables

List of Original Papers

This thesis is structured around three papers. These papers are as follows:

1. **Gopalan, G.**, Hrafnkelsson, B., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F.: A Bayesian hierarchical model for glacial dynamics based on the shallow ice approximation and its evaluation using analytical solutions, The Cryosphere, 12, 2229-2248, https://doi.org/10.5194/tc-12-2229-2018, 2018.

2. **Gopalan, G.**, Hrafnkelsson, B., Wikle, C.K., Rue, H., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F.: A Hierarchical Spatio-Temporal Statistical Model Motivated by Glaciology. Published by the Journal of Agricultural, Biological, and Environmental Statistics, 2019. *Winner of the 2019 American Statistical Association (ASA) Section on Bayesian Statistical Science (SBSS) Laplace Award.*

3. **Gopalan, G.**, Hrafnkelsson, B., Aðalgeirsdóttir, G., and Pálsson, F.: Spatio-temporal statistical modeling of Langjökull, to be submitted to the Annals of Applied Statistics, 2019.

Abbreviations

- **SIA**: shallow ice approximation
- **ODE**: ordinary differential equation
- **PDE**: partial differential equation
- **SPDE**: stochastic partial differential equation
- **GP**: Gaussian process
- **BHM**: Bayesian hierarchical model
- **MVN**: multivariate normal
- **IID**: independent and identically distributed
- **RV**: random variable
- **RW**: random walk
- **PC**: penalized complexity
- **AIC**: Akaike information criterion
- **ABC**: approximate Bayesian computation
- **INLA**: integrated nested Laplace approximation
- **DLM**: dynamic linear model
- **MCMC**: Markov chain Monte Carlo
- **RMSE**: root mean square error
- **GMRF**: Gaussian Markov random field
- **FEM**: finite element method

- **MB**: mass balance
- **ASL**: above sea level
- **GPS**: global positioning system
- **InSAR**: interferometric synthetic aperture radar
- **LIDAR**: light detection and ranging
- **UI-IES**: University of Iceland Institute of Earth Sciences
- **GPUs**: graphics processing units
- **CPUs**: central processing units

Acknowledgments

Never in my wildest dreams did I expect to be a member of the University of Iceland! This has been an amazing experience that I will always look upon fondly. In some ways, I have ended a long educational journey, but in other ways this is a new and exciting beginning. Doubtless, I would not be at this career juncture if it were not for the guidance of many gracious individuals.

First and foremost, I must thank my advisor, Professor Birgir Hrafnkelsson. Birgir has been incredibly supportive during the entirety of the Ph.D., both intellectually and personally. Birgir had architected an extremely well-thought-out Ph.D. project in a fascinating statistical area of research. Moreover, his command of statistics and intelligence have been apparent during our time working together, as he has never been short of good ideas and conversations regarding next steps to take the research. But beyond having statistical prowess, Birgir has also been a reliable and empathetic friend. I will always be grateful that he has encouraged me to maintain both a professional and personal life. Birgir's kindness has left an indelible impression, and I only hope to be able to impart the same excellence in supervision to any future students.

Furthermore, my glaciologist collaborators, Guðfinna Aðalgeirsdóttir, Alex Jarosch, and Finnur Pálsson, have been a joy to work with. Together, they possess an encyclopedic knowledge of glaciology and the Icelandic glaciers, and are always filled with sharp ideas. Additionally, traveling to Vatnajökull with the glaciology team was, both literally and figuratively, a once-in-a-lifetime experience. I will never forget gazing upon the caldera at Grímsvötn, imagining the pools of water that were hidden beneath the ice.

The third committee member, Chris Wikle, has been an excellent mentor throughout the project. He graciously hosted me at the Department of Statistics at the University of Missouri, and within a week had been able to suggest a seemingly endless number of directions to take the research. I am thoroughly impressed by how someone so academically

accomplished can be such a fun person to learn from and interact with.

Beyond this, I have enjoyed thought-provoking conversations and interactions with many other experts around the globe, including Håvard Rue, Haakon Bakka, and Stefan Siegert.

I cannot be thankful enough to the Icelandic Research Fund (RANNIS) for supporting the Statistical Models for Glaciology project. Due to their funding, I have been able to forge an academic career, meet brilliant people all over the globe, and work on tantalizing scientific problems. Simply put, there is nothing more a scholar can ask for from a funding agency.

Besides the individuals above, I must thank profusely all of the teachers, mentors, researchers, and educational institutions that have helped me along the journey.

1 Introduction

1.1 Motivation

The main purpose of this doctoral thesis is to amalgamate Bayesian hierarchical modeling approaches with physical modeling from glaciology. While Paper 1 reviews Bayesian hierarchical modeling for physical-statistical problems in greater detail, the motivation for this sort of modeling is most aptly summarized in the abstract of Berliner (2003):

"Two powerful formulas have been available to scientists for more than two centuries: Newton's second law, providing a foundation for classical physics, and Bayes's theorem, prescribing probabilistic learning about unknown quantities based on observations. For the most part the use of these formulas has been separated, with Newton being the more dominant in geophysics. This separation is arguably surprising since numerous sources of uncertainty arise in the application of classical physics in complex situations."

The current state of science, statistics, applied mathematics, and computation is a particularly exciting one to be pursuing the Bayesian hierarchical approach in geophysics, for a number of reasons. One reason is the trend in the sciences (i.e., geosciences, astronomy, and bioinformatics) to store and retrieve prodigious amounts of data. A second reason is a rapid growth in hardware to compute efficiently, for instance, with the aid of supercomputing clusters, graphics processing units (GPUs), parallel computation, and faster central processing units (CPUs). A third main reason is the commensurate development of efficient computational methodologies and algorithms to deal with analytically intractable Bayesian models. These include traditional Markov chain Monte Carlo (MCMC) methodologies such as Metropolis-Hastings and Gibbs sampling (Tierney, 1994), more recent MCMC methodologies such as Hamiltonian Monte Carlo (Neal, 2011) and slice sampling (Neal, 2003), and approximate inferential techniques such as variational inference (Blei et al., 2017),

integrated nested Laplace approximations (INLA) (Rue et al., 2009), and approximate Bayesian computation (ABC) (Marin et al., 2012). Additionally, recent work in the uncertainty quantification and probabilistic numerics branches of applied mathematics has focused on numerical methods from a Bayesian perspective, such as gamblets and Bayesian homogenization (Owhadi and Scovel, 2017), which come with proven computational complexity results.

This work develops physical-statistical models and uses data from Langjökull to validate the models and associated methods. Langjökull is Iceland's second largest glacier, with an area of 900 km^2, a volume of 190 km^3 and mean thickness of 210 m (Björnsson and Pálsson, 2008). The small relative thickness of the glacier (in comparison to horizontal dimensions) allows for valid application of the SIA (Fowler and Larson, 1978; Hutter, 1982, 1983; Flowers et al., 2005). While models applying the SIA have been used for predicting the evolution of other Icelandic glaciers, such as those from Aðalgeirsdóttir (2003) and Flowers et al. (2005), an advantage of the Bayesian hierarchical approach is that it gives full probability distributions for future glacial thickness values at various spatial locations (instead of just point predictions). Furthermore, along with a posterior predictive distribution for glacier thickness values at various spatio-temporal coordinates, the Bayesian hierarchical modeling approach yields full probability distributions over the ice viscosity and basal sliding parameters, providing plausible ranges for these parameters in light of the observed data. It is important to note that, even if a Bayesian method is not used, these parameters are very difficult to directly measure, and so some type of estimation is necessary.

In summary, this thesis develops spatio-temporal Bayesian hierarchical models for glacial thickness, validated with simulations and data from Langjökull. The same approaches can be applied to glaciers and ice sheets in other locations of the globe, such as Greenland. Besides contributing new modeling approaches and methods to the statistics literature, the ideas developed within this thesis can be used to predict, probabilistically, how glaciers will respond to a changing climate. Moreover, the physical-statistical modeling and methodology developed for glaciology can be applied to other scientific and engineering domains.

The next few subsections of the introduction give a concise overview of the key statistical elements employed in the three papers of this thesis: Bayesian modeling, Bayesian hierarchical modeling, spatio-temporal

statistics, computer simulators, model discrepancy, and emulators. These subsections are followed by a brief overview of glacial dynamics and the application of Bayesian modeling to glaciology.

1.2 Bayesian modeling

Because this thesis relies upon Bayesian modeling, it is important to give an overview of what Bayesian modeling entails. In an empirical study, one's objective is typically to make inferential statements about unknown quantities (e.g., scientifically relevant parameters or future, unobserved quantities) given observed data (e.g., the results of one or many experiments). Such inferential statements may then be used to make decisions of practical importance or arrive at scientifically relevant deductions or conclusions. For the sake of this exposition, it may be useful to refer to the unknown objects of inferential interest as θ, a vector in $\mathbb{R}^n$, but in a more general setting θ may belong to an infinite-dimensional set. The Bayesian approach models the joint distribution of all unknowns and potentially observable data with a probability distribution. Typically, this joint distribution is specified as a product of two components called a likelihood function, a probability density for the observed data as a function of θ, and a prior distribution, a probability density for θ. Finally, inference regarding the unknowns is accomplished by deriving the distribution of the unknowns conditional on the observed data, canonically referred to as the posterior distribution.

An important issue in Bayesian inference involves the choice, interpretation, and subjectivity of the prior distribution. One interpretation of the prior distribution is that it encodes an individual's subjective prior beliefs about inferential objects of interest before an experiment has been conducted. In some instances, priors are chosen for analytical convenience, such as in the case of conjugate priors, where the posterior is analytically tractable and also happens to have some important theoretical properties, as shown by Diaconis and Ylvisaker (1979) in the exponential family setting. Other schools of thought attempt to define default priors based on invariance principles, such as the Jeffreys priors (Jeffreys, 1946), or information-theoretic principles such as reference priors (Berger et al., 2009) and the penalized-complexity priors due to Simpson et al. (2017). Yet another interpretation of the prior distribution is that it is a weighting

function that regularizes estimates, and for instance both the widely-used Lasso and ridge regression estimators have an interpretation in this vein. That is, the Lasso estimate (Tibshirani, 1996) may be seen as the posterior mode of a typical linear regression setup with independent, zero-mean Laplacian priors on the coefficients, and the ridge regression estimate (Hoerl and Kennard, 1970) can be seen as the posterior mode of a typical linear regression setup with independent, zero-mean normal priors on the coefficients. In addition to many variations for prior choice, it is important to realize that the conditional distribution of the observed data given the parameters of interest can be arbitrary, and it is also often times used in a frequentist analysis.

Complementary to the Bayesian approach, the frequentist approach provides a way of assessing the goodness of an inferential procedure. One frequentist interpretation treats the inferential objects of interest as fixed and unknown quantities, where any probability distribution ascribed to them is ultimately an inferential construct. Variation in observable data is thus modeled by the probability distribution underpinning the likelihood function. In other words, the appropriate distribution to reference in a frequentist validation is the conditional distribution of the data given the unknowns, as opposed to the joint distribution of the data and unknowns. A frequentist analysis answers the question: how does this estimator or confidence interval perform under repeated sampling or observations according to a particular data distribution, treating unknowns as fixed quantities? So for example, a 95 percent confidence interval should, in theory, contain the unknown of inferential interest in 95 percent of samples over the long run. However, a given interval either does or does not contain the unknown.

There are a number of mathematically rigorous ways to reconcile Bayesian inference with frequentist criteria, of which two are highlighted. The first justification is asymptotic; under weak regularity conditions and assuming data are generated independently, identically, and conditioned on a fixed, unknown parameter, the posterior concentrates around the fixed parameter as more data are collected, e.g., as in the Bernstein von Mises theorem (van der Vaart, 2000; Shen and Wasserman, 2001). There are also similar nonparametric results, for instance in Castillo and Nickl (2014). Nonetheless, the reader is pointed to the simulation studies of Paper 1, which show an interesting sort of asymptotic behavior for the posterior inference of ice viscosity when more data are collected in

a fixed time window. The second justification comes from statistical decision theory, in which the complete class theorem loosely states that the admissible estimators are the generalized Bayes rules (Lehmann and Casella, 2003); in particular, assuming continuity of risk functions and a strictly positive prior distribution, Bayes rules (i.e., estimators that minimize a loss function over the joint distribution of data and unknowns) are admissible (see (Lehmann and Casella, 2003) for the precise mathematical definition). The continuity of risk functions holds in common settings, such as when the data distribution comes from the exponential family (Lehmann and Casella, 2003; Robert, 2007).

1.3 Bayesian hierarchical modeling

As the terminology suggests, Bayesian hierarchical modeling is an extension of Bayesian modeling (Cressie and Wikle, 2011). A hierarchical model specifies the data distribution through conditional distributions and latent variables. By taking the product of these distributions, and then marginalizing (i.e., integrating) out latent variables, one arrives at the marginal distribution of the data. Let us introduce a latent variable, S, and assume the data variable is Y. Their respective probability distributions are $p(S)$ and $p(Y)$. Rather than specifying $p(Y)$ directly, it may be conceptually simpler to specify $p(S)$ and $p(Y|S)$, the *conditional* distribution of Y given S. In the context of modeling in Cressie and Wikle (2011), S typically represents a latent physical process (e.g., humidity) and Y represents observable data than can be easily linked to the physical process (e.g., inches of rainfall in some region). The marginal distribution of Y, $p(Y)$, can then be obtained as $\int p(Y|S)p(S)ds$.

Furthermore, if there are physical parameters, represented as θ, we can denote the distribution of the process level as $p(S|\theta)$, emphasizing the dependence of the physical process on some physical parameters. Additionally, a prior distribution, $p(\theta)$, can be ascribed to θ. Assuming that Y is conditionally independent of θ given S, then the distribution of the data conditional on the physical parameter, $p(Y|\theta)$, is $\int p(Y|S)p(S|\theta)ds$. Sometimes this integral can be computed analytically. This holds in this thesis because the latent distribution ($p(S|\theta)$) and the data level distribution ($p(Y|S)$) are multivariate normal. In such a situation, the marginal distribution, $p(Y|\theta)$, also has a multivariate normal form. How-

ever, in cases when this does not hold, one can use other computational procedures (e.g., MCMC (Tierney, 1994), INLA (Rue et al., 2009), etc.).

1.4 Spatial and spatio-temporal statistics

The most basic statistical models assume independent and identically distributed (IID) (hence uncorrelated) measurements. In many real-world problems, however, there will be correlation between data points. This is true, for instance, in the case of time series; measurements that are taken close together in time will tend to exhibit more correlation than those that are taken far apart in time. Likewise, if data are spatially referenced, it is natural to assume that nearby data points (e.g., measurements of temperature) will tend to be more correlated, and this correlation drops off as the distance between points increases.

Let $s_1, s_2, ..., s_m$ be elements of $\mathbb{R}^d$. A common way of representing spatial covariance is to assume that some stochastic process, Y, which can be indexed spatially as $Y(s_1), Y(s_2), ..., Y(s_m)$, is distributed according to a Gaussian process (GP), with some mean function $\mu(s)$ and a covariance kernel that is often a function of the distance between points. For instance, a common covariance kernel is the squared exponential kernel, in which the covariance between points decays as the exponential of the squared distance between points:

$$C(s_a, s_b) = \sigma^2 \exp \frac{-||s_a - s_b||^2}{2\phi^2}.$$

Here, σ is the marginal standard deviation and ϕ is the length-scale parameter.

Another covariance function that is often used in spatial statistics is the Matérn kernel (Bakka et al., 2018), which is described more in Paper 3, but whose formula is repeated here for sake of comparison:

$$C(s_a, s_b) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{8\nu}||s_a - s_b||/\rho) K_\nu (\sqrt{8\nu}||s_a - s_b||/\rho).$$

Here, σ is the marginal standard deviation, ρ is the spatial range parameter, ν is the smoothness parameter, K_ν is the modified Bessel function of second kind with order ν, and Γ is the gamma function.

The squared exponential kernel is actually a version of the Matérn kernel as the smoothness parameter tends to infinity (Rasmussen and Williams, 2006). Both of these covariance functions are depicted below. Besides these two kernels, there are other choices at the disposal of the modeler, which can be found, for instance, in Rasmussen and Williams (2006) and Banerjee et al. (2003).

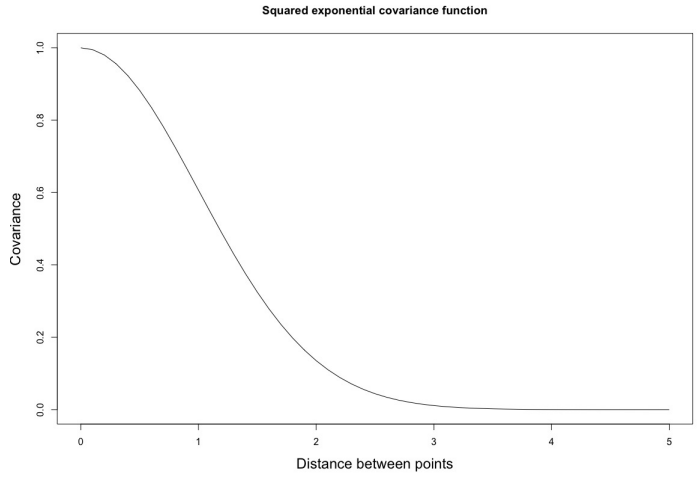

Figure 1.1. Squared exponential covariance function as a function of distance. Both σ and ϕ are 1.

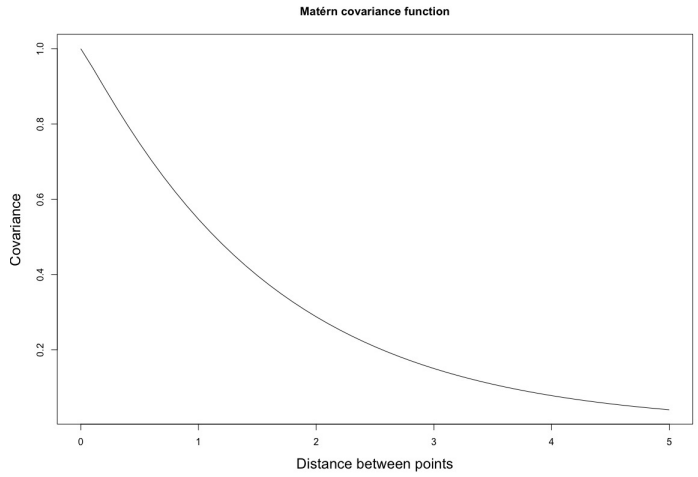

Figure 1.2. Matérn covariance function as a function of distance. Both σ and ρ are 1, and the smoothness parameter is .55.

If data are spatially referenced to a lattice (i.e., a finite number of spatial locations), and their joint distribution is a multivariate normal

distribution, then we can refer to these data as a Gaussian Markov random field (GMRF) (Rue and Held, 2005). A useful property of a multivariate normal precision matrix (i.e., inverse of the covariance matrix) is that 0 entries imply conditional independence. Specifically, let Y have a multivariate normal distribution with mean μ, precision matrix Q, and the components of Y are referred to by subscript. Then Y_i is independent of Y_j conditioning on all other components of Y if and only if $Q_{ij} = 0$. Furthermore, it is possible to link geostatistical fields (discussed previously) with GMRFs through stochastic partial differential equations (SPDEs), an approach that is described more in Lindgren et al. (2011). We use this approach, implemented in the R-INLA software, to make mass balance predictions in Paper 3.

If data have both spatial and temporal coordinates, it is possible to implement correlation structures that are a function of both space *and* time. This thesis develops spatio-temporal correlation with a multivariate random walk where the error term of the random walk is derived from a squared exponential kernel. The full covariance under this model specification is described in greater detail in Paper 2; in particular, there is correlation across different space *and* time coordinates, consequently inducing spatio-temporal correlation structure.

1.5 Computer simulators, model discrepancy, and emulators

Consider the most basic linear model of the form $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$; Y_i is the observed data, β_0 is an intercept term, X_i is a covariate, and β_1 is an associated parameter. If ε_i is a Gaussian random variable (RV) with mean 0, then Y_i is a Gaussian RV with mean $\beta_0 + \beta_1 X_i$, which is a linear function of a covariate. However, in a physical situation, it is typical that some nonlinear function, $f(.)$, is used instead of a linear function. Such a function is often times the output of a computer program or simulator, such as a numerical PDE solver. This sort of a model takes the form $Y_i = f(\theta, \phi) + \varepsilon_i$, where θ is a physically important parameter and ϕ represents additional parameters needed for the computer program to run.

Kennedy and O'Hagan (2001) consider such models,which are further developed by Brynjarsdóttir and O'Hagan (2014). In particular, often

times the output of a computer simulator will not perfectly capture the physical process being modeled (e.g., due to numerical inaccuracy or approximate dynamics upon which the simulator is based). For this reason, Kennedy and O'Hagan (2001) suggest including an additional term called model discrepancy, a function denoted by $\delta(\phi)$. In other words, $Y_i = f(\theta, \phi) + \delta(\phi) + \varepsilon_i$. The introduction of Paper 2 elaborates more on model discrepancy and its relationship to the model within this thesis.

One complication in such models is that a computer simulator ($f(\theta, \phi)$) can take a lot of computational resources to evaluate, i.e., it takes a long time to run. Most inference methods in such models will require repeated evaluations of $f(\theta, \phi)$, which can take an unfeasible amount of time if evaluating $f(\theta, \phi)$ is slow. Emulators (also referred to as surrogate models) can mitigate this computational burden by replacing $f(\theta, \phi)$ with a function $g(\theta, \phi)$ that mimics the output of $f(\theta, \phi)$, but is less computationally expensive to evaluate. The emulator is constructed by running $f(\theta, \phi)$ for some pre-defined values of θ, ϕ, and using the output to train g. Paper 2 uses first-order emulation (Hooten et al., 2011) to emulate the numerical PDE solver from Paper 1. In addition to first-order emulators, there are other types of emulators in the literature, notably polynomial-chaos expansions (Sargsyan, 2016) and Gaussian process based emulators (Gu et al., 2018).

1.6 Glacial dynamics

The emphasis of the introduction has been on statistical concepts that are used in this thesis, though a preliminary overview of glacial dynamics is important to include. The time evolution of glacial thickness is attributed to the flow of ice and the rate of change of ice mass, which is caused mostly by melting of ice and snow accumulation (and compaction). Ice flow is essentially caused by two sources: 1) deformation due to gravity and 2) interaction of the bed of the glacier with bedrock, which can either act in the direction of flow (sliding) or impede the flow (friction). The direction of flow is in the negative of the surface gradient, confirmed by observations from Minchew et al. (2015). In the SIA used for this thesis, the exact formulation for deformation due to gravity comes from Glen's flow law (Glen, 1955), and the basal sliding component comes

from Weertman's sliding relation (Weertman, 1964); more is described in Paper 1.

Additionally, besides deformation due to gravity and basal sliding, there are other factors not included in the SIA that can have an effect on glacial dynamics. One such factor is a glacial surge, when a glacier suddenly flows at a much faster rate than usual. It is estimated that 40 km^3 of ice had been transported from accumulation areas to melting areas at Vatnajökull during the 1990s (Björnsson and Pálsson, 2008). Another main event is a jökulhlaup, a flood that occurs when there is an outburst of meltwater stored in a glacier. Such a flood is caused by geothermal fields that continuously melt ice, the build up of water at ice dams, or a volcanic eruption that melts ice that subsequently drains at the glacier margin (Björnsson and Pálsson, 2008).

While the application of Bayesian modeling is relatively new to glaciology, there have been some notable instances of such work in the literature. Some important examples include Berliner et al. (2008), Pralong and Gudmundsson (2011), Brinkerhoff et al. (2016), Isaac et al. (2015), Minchew et al. (2015), and Guan et al. (2016). More details for these papers are given in Paper 1. Our work differs from these treatments because it includes both 2 spatial dimensions and time, as well as model discrepancy, through a multivariate random walk.

2 Materials and methods

All of the programming in this thesis was done in the R programming language (R Core Team, 2016). Code was written for simulating glacial dynamics based on the analytical solutions to the SIA model from Bueler et al. (2005), implementation of the finite difference solver in Paper 1, posterior inference within the main Bayesian hierarchical model, emulation, and mass balance predictions. The major pieces of code are included at the end of the thesis. The main R packages used in the thesis are:

1. mvtnorm_1.0-8 (for multivariate normal routines) (Genz et al., 2019),

2. FastGP_1.2 (for the elliptical slice sampling method) (Gopalan and Bornn, 2015),

3. randomForest_4.6-12 (for training emulators) (Liaw and Wiener, 2002)

4. raster_2.8-4 (for making spatial maps) (Hijmans, 2018), and

5. INLA_18.07.12 (for making mass balance predictions) (Rue et al., 2009).

Additionally, R Studio Server was run on an Amazon Web Services (AWS) instance (m4.2xlarge) for the analysis of Paper 3, which processes Langjökull data. R version 3.3.1 was used, and with the exception of the AWS server, R was run on a 2015 MacBook Pro, 2.7 GHz Intel Core i5 processor with 8 GB 1867 MHz DDR3.

The main statistical approach used for constructing the models in this thesis is the Bayesian hierarchical modeling framework. We used the version of Bayesian hierarchical modeling as delineated in Cressie and Wikle (2011). This is the parameter, process, data level hierarchical modeling framework, described in more depth in the introduction of Paper 1.

The data used was provided by Finnur Pálsson, member of the glaciology group of the University of Iceland Institute of Earth Sciences (UI-IES). The first piece of data consists of a 100 m resolution surface digital elevation map of Langjökull, based on measurements taken in late April to early May of 1997. The second piece of data consists of a 100 m resolution surface topographical map of the Langjökull bedrock, also collected during late April to early May of 1997. The final piece of data consists of the locations (i.e., latitude, longitude), surface elevation, and summer and winter mass balance measurements at 22-25 measurement sites across Langjökull, between 1997-2015 inclusive. More information about the data and how they were collected is given in Section 2 of Paper 3.

3 Summary of papers, contributions, and conclusions

3.1 Summary of papers

The papers of this project were designed to develop models and validation in a natural sequence. As such, the first paper introduces a prototypical Bayesian hierarchical model for glacial flow that allows for the prediction of glacier surface elevation and the inference of ice viscosity. This model is validated with simulation studies that use analytical solutions to the SIA PDE, introduced by Bueler et al. (2005). The process level of the Bayesian hierarchical model embeds a second-order finite difference solver for the SIA PDE, which is based on the Lax-Wendroff method (Hudson, 1998). This paper is novel for constructing a Bayesian hierarchical model that operates in 2 spatial dimensions and time – previous work either does not involve time or involves only one spatial dimension with time (Berliner et al., 2008; Brinkerhoff et al., 2016; Guan et al., 2016). Paper 1 has been published in The Cryosphere, a glaciology journal (Gopalan et al., 2018).

Paper 2 develops the general applicability of the model from Paper 1 and shows ways to improve computation. In particular, as an extension of Paper 1, Paper 2:

1. Demonstrates the use of a first-order emulator instead of a numerical solver (the resultant speed-up in computation is demonstrated),
2. uses random walks (RWs) of order greater than 1 (the first paper only considered RW(1)), and performs associated residual analyses,
3. derives sparsity results for log-likelihood evaluation,
4. derives the computational complexity of log-likelihood evaluation,
5. demonstrates empirical run-time results, and

6. provides a method to fit an error-correcting process when little prior information is available.

Additionally, connections are made to the literature on model discrepancy (Kennedy and O'Hagan, 2001; Brynjarsdóttir and O'Hagan, 2014). This paper has been selected as the winner of the 2019 American Statistical Association (ASA) Section on Bayesian Statistical Science (SBSS) Laplace Award. Also, it has been published in the Journal of Agricultural, Biological, and Environmental Statistics (Gopalan et al., 2019).

Paper 3 applies the methodology and modeling from the first two papers to data sets from Langjökull. The data sets include the 1997 surface elevation of Langjökull, the bedrock topography as measured with radio-echo sounding by the University of Iceland glaciology team, and winter and summer measurements of mass balance from 1997-2015. This paper includes the application of R-INLA to make spatial predictions of mass balance at Langjökull, a new contribution to glaciology. It also derives a posterior estimate for ice viscosity and predictions for surface elevation at Langjökull. In contrast to the simulation studies of Paper 1, the posterior for ice viscosity concentrates sharply at a single value, which is in the domain of expected ice viscosity values to be expected at Vatnajökull (Aðalgeirsdóttir et al., 2000). Additionally, the surface elevation predictions for 2015 at the measurement sites, trained on data from 1997-2014, are within a few meters of the observed surface elevations. The final paper is in preparation for submission to the Annals of Applied Statistics.

3.2 Contributions to the literature

It is important to underscore the contributions that these papers have made to both the glaciology and statistics literature. The Bayesian hierarchical spatio-temporal model contributed in this thesis is novel in that it incorporates two spatial dimensions and time. Many of the previous applications of Bayesian modeling in glaciology have either not included time, or have included a dynamical equation that operates only along the flowline in one spatial dimension (Berliner et al., 2008; Brinkerhoff et al., 2016; Guan et al., 2016). The contribution of a Bayesian hierarchical model in time and two spatial dimensions is important because it allows for probabilistic forecasts of surface elevation across the entire area of

the glacier, as opposed to a point estimate of how glaciers will develop in the future (Aðalgeirsdóttir et al., 2006; Flowers et al., 2005). Since there are numerous sources of uncertainty in glaciological modeling (e.g., physical model used, numerical inaccuracies, uncertainty in fundamental constants), it is important that predictions for glacier evolution are probabilistic. An additional contribution to the glaciology literature is a new finite difference method, based on the Lax-Wendroff method, which is a second-order method. Most of the finite difference methods in glaciology have been first-order, such as in Bueler et al. (2005); Jarosch et al. (2013); Aðalgeirsdóttir (2003). Nonetheless, the numerical errors of this method are qualitatively similar to the errors exhibited by first-order finite difference methods, as demonstrated in Bueler et al. (2005). This knowledge was used in constructing the error-correcting process of the aforementioned Bayesian hierarchical model. A third main contribution to the glaciology literature is using a formal spatio-temporal statistical model to make mass balance predictions (i.e., interpolations) at Langjökull. Previously, this has been done manually (Pálsson et al., 2012).

This thesis contributes to the statistics literature as well. One main contribution is the use of a random walk model to account for model discrepancy (i.e., the deviation between a numerical PDE solver and the real spatio-temporal physical process). Model discrepancy has mostly been inferred with a GP prior over a space of functions (Kennedy and O'Hagan, 2001; Brynjarsdóttir and O'Hagan, 2014). It has been shown that, under the random walk model, the log-likelihood can be computed exactly in an efficient manner, due to the use of bandwidth-limited matrix algebra. Moreover, an easily-parallelizable approximation to the log-likelihood has been developed as well. The computational complexity for the evaluation of both the exact and approximate likelihood has been derived. Additionally, simulating from the posterior predictive distribution can be done simply in such a model. Another major statistical contribution is the use of first-order emulation for emulating a numerical PDE solver; in Paper 2, a computational improvement is afforded without any appreciable difference in the posterior as compared to a posterior derived from a numerical PDE solver.

3.3 Conclusions

One conclusion of this work is that prior knowledge for the error-correcting process (i.e., model discrepancy), embedded in the covariance matrix of the random walk, can reduce the bias of physical parameter estimates. This is consistent with a conclusion of Brynjarsdóttir and O'Hagan (2014). In Papers 1 and 2, this phenomenon is demonstrated by using prior knowledge of the discrepancy between finite difference solutions and exact analytical solutions to the SIA, from the work of Bueler et al. (2005). Another main conclusion is that the Bayesian hierarchical model can generate good predictions of glacial surface elevation, which is demonstrated both in the simulation studies of Paper 1 and the Langjökull data of Paper 3. Unexpectedly, however, the posterior for ice viscosity has much higher precision in the real-data test case in comparison to the simulation studies of Papers 1 and 2. Another main conclusion is that computational improvements are possible for performing Bayesian inference within the proposed model, by using bandwidth-limited matrix linear algebra, a parallelizable approximation to the log-likelihood, and emulating a numerical PDE solver with first-order emulators.

The three papers that this thesis consists of suggest that the Bayesian hierarchical model developed, along with associated methodologies, provide a useful, computationally efficient way to understand glaciological systems and make probabilistic forecasts of surface elevation of temperate glaciers.

4 Paper 1

A Bayesian hierarchical model for glacial dynamics based on the shallow ice approximation and its evaluation using analytical solutions

Gopalan, G., Hrafnkelsson, B., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F.: A Bayesian hierarchical model for glacial dynamics based on the shallow ice approximation and its evaluation using analytical solutions, The Cryosphere, 12, 2229-2248, https://doi.org/10.5194/tc-12-2229-2018, 2018.

Abstract: Bayesian hierarchical modeling can assist the study of glacial dynamics and ice flow properties. This approach will allow glaciologists to make fully probabilistic predictions for the thickness of a glacier at unobserved spatio-temporal coordinates, and it will also allow for the derivation of posterior probability distributions for key physical parameters such as ice viscosity and basal sliding. The goal of this paper is to develop a proof of concept for a Bayesian hierarchical model constructed, which uses exact analytical solutions for the shallow ice approximation (SIA) introduced by Bueler et al. (2005). A suite of test simulations utilizing these exact solutions suggests that this approach is able to adequately model numerical errors and produce useful physical parameter posterior distributions and predictions. A byproduct of the development of the Bayesian hierarchical model is the derivation of a novel finite difference method for solving the SIA partial differential equation (PDE). An additional novelty of this work is the correction of numerical errors induced through a numerical solution using a statistical model. This error-correcting process models numerical errors that accumulate forward in time and spatial variation of numerical errors between the dome, interior, and margin of a glacier.

4.1 Introduction

The shallow ice approximation (SIA) is a nonlinear partial differential equation (PDE) that describes ice flow when glacier thickness is relatively small compared to the horizontal dimensions. Derived from the principle of mass conservation, the SIA PDE depends on two key physical parameters: ice viscosity and basal sliding (sometimes described as basal friction or drag). The primary objective of this paper is to develop a Bayesian hierarchical model (BHM) for glacier flow utilizing the framework espoused by Wikle (2016) and Cressie and Wikle (2011), which allows one to: 1) infer ice viscosity and basal sliding parameters and 2) make probabilistic predictions for glacial thickness at unobserved spatio-temporal coordinates. This BHM relies upon a finite difference scheme for solving the SIA that is inspired by the Lax-Wendroff method (Hudson, 1998). To validate this BHM, we utilize exact analytical solutions from Bueler et al. (2005). Hence, in addition to the development of a BHM for shallow glaciers, this paper serves as a case study for the strategy of using exact analytical solutions to validate or tune BHMs governed by physical dynamics. Moreover, the BHM developed can be applied to the general “physical-statistical" problem (Berliner, 2003). This BHM is verified and diagnosed through a combination of assessments of posterior probability intervals, checks of predictive accuracy for glacial thickness prediction, and a comparison between observed and expected errors due to the numerical solution of the SIA.

4.1.1 An overview of Bayesian modeling and BHMs

Before describing how BHMs are used in physical-statistical models, particularly for geophysical problems, a very terse overview of Bayesian modeling and Bayesian hierarchical modeling is given for the uninitiated reader. A main component of Bayesian statistics is the use of probability distributions to model parameters thought to be fixed quantities (i.e., scientific constants); this assumption allows one to use rules of conditional probability (i.e., Bayes’ theorem) to derive probability distributions for scientific quantities of interest, such as physical constants or predictions of future quantities of a system being studied. Typically, the major assumptions required as input to the analysis are prior distributions for parameters as well as a probabilistic model for the data. The output is a

probability distribution for parameters or predictions conditional on data that has been collected or observed; canonically, this is referred to as the posterior distribution.

A BHM is a Bayesian model in which the probabilistic model for data is specified in a hierarchy. Working with such a hierarchy has a number of advantages – it is usually easier to conceptualize the probabilistic model for the data, and it is also easier to model various parts of a system of interest modularly instead of all at once. Such an approach is conducive to the construction of a probabilistic model that tightly corresponds to a scientific system of interest, which is naturally thought of in separate components or modules. In a BHM, the rules of conditional probability can be used to specify the relevant distributions. For example, let us consider a mock system that has parameter vector θ, an intermediate unobserved vector S, and observations Y. θ might be statistical or physical parameters, S could be a quantity of scientific interest, and Y could be noisy observations of S. A schematic for such a model is given in Figure 4.3, and the joint probability distribution is

$$p(\theta, S, Y) \quad = \quad p(\theta)p(S|\theta)p(Y|S,\theta).$$

The distribution $p(\theta)$ represents prior beliefs about parameters before data are collected, while $p(S|\theta)$ represents prior knowledge or assumptions for how S is generated given parameters. For instance, this prior knowledge could entail clustering or some dependence between the elements of S. The process that models Y conditional on S and θ is $p(Y|S,\theta)$. The posterior distribution of scientific quantities of interest, $p(\theta, S|Y)$, is proportional to $p(\theta, S, Y)$ by Bayes' theorem. Estimates and assessments of uncertainty of scientific parameters and quantities can be extracted from the posterior distribution.

4.1.2 Physical-statistical modeling with BHMs

The case for applying Bayesian hierarchical modeling and methodology in geophysics is strongly made by Berliner (2003), which he describes as "physical-statistical modeling". Particularly, employing the Bayesian hierarchical approach has the primary advantage of incorporating all relevant sources of uncertainty and randomness into one coherent probabilistic framework. The sources typically modeled together are: 1) measurement errors in the data collection process, 2) lack of full knowledge of the

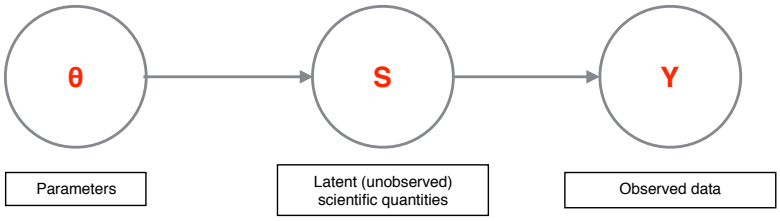

Figure 4.3. Schematic of a simple Bayesian hierarchical model; here, θ represents physical parameters, S represents unobserved scientific quantities of interest, and Y represents the observed data.

precise functional form of the underlying physical equations describing the physical phenomenon being modeled, or else simplification of the physical system description 3) numerical errors induced while approximating the solution to a system of partial differential equation PDEs, and 4) lack of precise knowledge of fundamental parameters (constants) in the underlying PDEs describing said phenomenon. In the Bayesian hierarchical framework (Berliner, 1996; Wikle, 2016; Cressie and Wikle, 2011) each of these sources of uncertainty is modeled by conditioning on the appropriate quantities, and inference is performed by sampling from or approximating the posterior distribution (the distribution of the unknown quantities of interest conditional on the observed data).

At the highest level of a BHM, prior probability distributions are laid out for the physical parameters of interest. At the intermediary level, a probability distribution for the physical process of interest is laid out conditional on the parameters, which is typically motivated by a numerical scheme for solving PDEs. In particular, this level may be modeled as the sum of the output from a numerical solver and an error-correcting process. Finally, at the observed level, a probability distribution is set forth for the observed data conditional on the latent physical process and other relevant measurement parameters, which include variances of measuring procedures or devices. The product of these probability distributions specifies the joint distribution of all relevant quantities, which is proportional to the posterior distribution by the definition of conditional probability. While a traditional analysis may handle each of these disparate sources of uncertainty in an ad-hoc and disjointed fashion, the Bayesian hierarchical approach leverages probability measures to cohesively model major sources of uncertainty

and undertake inference in a principled manner. Figure 4.4 diagrams what a prototypical physical-statistical Bayesian hierarchical model might look like.

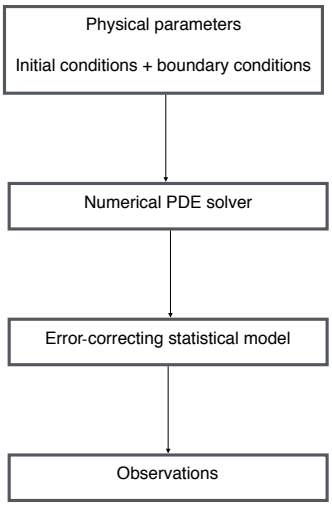

Figure 4.4. Schematic of a prototypical physical-statistical Bayesian hierarchical model. At the top layer, physical parameters, initial conditions, and boundary conditions are fed into a numerical solver, and the output of this is corrected with an error-correcting process; finally the actual observations are dependent on the actual physical process values.

While the BHM approach to physical-statistical problems offers many advantages, it is not an infallible approach. In particular, while constructing a BHM may be straightforward, actually fitting a BHM to data can be computationally difficult. In the analysis that follows, there are only one to two physical parameters and the likelihood function is tractable, so posterior computation is not difficult. In more complex scenarios with many physical parameters (e.g., a basal sliding field with a parameter for each grid point), it becomes more difficult to compute the posterior or draw samples from it. There are now many new tools, however, for Bayesian inference of complicated and high dimensional posterior distributions, such as Stan (Stan Development Team, 2018) and INLA (Rue et al., 2017). Another potential difficulty in using BHMs for physical-statistical problems is that solving for a set of dynamical equations with a numerical method can be computationally onerous, generally speaking; while this is not a detriment in the work that follows, this can be a prob-

lem for posterior computation. One way to circumvent this issue is to emulate a numerical solver, using techniques as in Hooten et al. (2011). Another methodology that can be used to efficiently solve PDEs using Bayesian numerical analysis comes from Owhadi and Scovel (2017). Finally, Calderhead et al. (2008) suggests methodology to avoid explicitly solving ordinary differential equations by using Gaussian processes.

To put the contributions of this work into context, we briefly review glaciology papers that have employed Bayesian modeling. In Berliner et al. (2008), a Bayesian hierarchical approach is used to model ice streams in one spatial dimension, and an error-correcting process is utilized to account for a simplification in the physical model. A combination of Markov chain Monte Carlo (MCMC) and empirical Bayes methodology is used to fit the model, and basal shear stress and resistive stresses are included. Furthermore, wavelets are used for dimensionality reduction purposes so as to make the computations more feasible. In Pralong and Gudmundsson (2011), a Bayesian model is constructed for an ice stream where the likelihood and prior are Gaussian. The observed data are surface topography, horizontal and vertical surface velocities, and the latent system state is basal topography and slipperiness. The goal is to infer the system state given the observed data, and ultimately a maximum a posteriori (MAP) point estimate is used for inference in conjunction with an iterative method for posterior maximization. Physics is incorporated by solving for the steady state solution with a finite element method (FEM) solver, given the system state. In Brinkerhoff et al. (2016) a flowline model of the SIA is considered with vertically integrated velocities. Gaussian process priors are used for all unknowns, and the Metropolis–Hastings algorithm is used to fit the model. The approach yields convincing results in simulations and a real data set. In Isaac et al. (2015), numerical methods are presented for solving a nonlinear Stokes equation boundary value problem for an ice sheet in Antarctica. The method ultimately uses a low rank approximation to a covariance matrix for the posterior distribution of a basal parameter field. Finally, and perhaps most directly related to this research, in Minchew et al. (2015) interferometric synthetic aperture radar (InSAR) is used to determine velocity fields at Langjökull and Hofsjökull in early June 2012. The velocity directions match the surface gradient, but magnitudes do not appear to coincide with the theoretical predictions of other authors (likely due to the inappropriate modeling of basal sliding).

The same approach within this work can be used for non-SIA problems in cryosphere science, and the Bayesian hierarchical model does not necessitate analytical solutions; the analytical solutions are used for the evaluation of the particular BHM in the paper based upon the SIA. However, in general, the biggest difficulty will be in developing a statistical error-correcting process that appropriately models numerical errors for an arbitrary scenario, where a numerical solver for a different set of dynamical equations is used. In the SIA context, we can rely on prior studies of Bueler et al. (2005) to tell us something about how the numerical errors will look like in the SIA case – i.e., spatial variation in the scale of numerical errors between the dome, interior, and margin. This error pattern will not hold in general for other geometries and systems, and so either different prior studies must be utilized, or if these don't exist, the hierarchical model must be extended to include a more general model for the error-correcting process (e.g., a spatially varying field for the log of the scale of numerical errors with a Gaussian process prior).

The main differentiating contribution of this paper is to utilize the exact analytical solutions from Bueler et al. (2005) to evaluate the BHM employed. An additional novelty is the derivation and utilization of a novel finite difference method for solving the SIA PDE that operates in two spatial dimensions; consequently, the Bayesian model employed also operates in two spatial dimensions, in addition to time. Finally, we explicitly model the errors due to a numerical solver with a spatio-temporal statistical process, which accounts for different scales of spatial variability within the dome, within the interior, and within the margin of the glacier, as well as accumulation of numerical errors forward in time.

4.2 Description of Models

4.2.1 Shallow ice approximation

The physics of glaciers is an extensive topic; hence, only the portions which are most relevant to this paper are described. The reader is pointed to the comprehensive works by Cuffey and Paterson (2010) and van der Veen (2013) for further reading on the subject. PDEs for glaciers are derived from the following considerations. First, glaciers are modeled as very slowly moving and viscous fluids. By applying the principle of mass

conservation, the net ice flux moving in or out of an infinitesimal column of the glacier located at some spatial coordinate, plus the net mass change due to precipitation or melting, yields the change in the height of the column over an infinitesimal time interval. Such a heuristic argument provides a PDE in two dimensions for a glacier, with averaged velocities in two spatial dimensions. The PDE relates the time derivative of the thickness of the glacier to the flux and net mass change (i.e., mass balance). The main assumptions are that ice is isotropic and homogeneous, and also that longitudinal and transverse stress terms can be ignored, which is reasonable when the overall thickness of the glacier is small in comparison to its width. Under these assumptions, the velocity of the ice is made up of two additive components. The first component of the velocity is based upon deformation due to gravity, which acts in the direction of steepest descent of the surface and is a function of the ice viscosity parameter. The second component of velocity also acts along the gradient of the glacier surface and is a function of the basal sliding parameter field. The formulations stem from Glen's flow law (Glen, 1955, 1958) and Weertman's sliding relation (Weertman, 1964).

Written in terms of glacial thickness, $H(x,y,t)$, the SIA PDE is:

$$\begin{aligned}
H_t &= -[\bar{u}H]_x - [\bar{v}H]_y + \dot{b}. \\
-[\bar{u}H]_x &= -[-C_0\gamma(-\rho g H[H+R]_x)H + \frac{2B}{n+2}(\rho g\alpha)^{n-1}H^{n+1}(-\rho g H[H+R]_x)]_x \\
-[\bar{v}H]_y &= -[-C_0\gamma(-\rho g H[H+R]_y)H + \frac{2B}{n+2}(\rho g\alpha)^{n-1}H^{n+1}(-\rho g H[H+R]_y)]_y \\
\alpha &= \sqrt{[H+R]_x^2 + [H+R]_y^2}
\end{aligned}$$

Here $H(x,y,t)$ is the thickness of the glacier at spatial coordinate (x,y) and time t, $\bar{u}$ is the average velocity in the x direction and $\bar{v}$ is the average velocity in the y direction. This model is vertically integrated, and hence only two spatial dimensions are modeled. $R(x,y,t)$ is the bedrock elevation which is assumed to be constant in time, so it can be written as $R(x,y)$; $\dot{b}(x,y,t)$ is the mass balance field, B and $C_0\gamma$ are physical parameters governing the viscosity and basal sliding; ρ governs the mass density of the ice; and finally n is Glen's flow law constant, typically set to 3. Initial conditions (i.e., $H(x,y,0)$) are assumed to be given, and the boundary condition $H \geq 0$ is assumed, just as in Table 2 of Bueler et al. (2005). Additional derivations and details on the SIA

are covered in a variety of sources, including Fowler and Larson (1978), Hutter (1982), Hutter (1983), and Flowers et al. (2005).

It is important to make explicit that there are some limitations of this PDE. Besides ignoring longitudinal and transverse stress terms, the PDE does not model subglacial hydrology, tunneling systems, jökulhlaups, or surges, the dynamics of which are believed to contribute to dynamics of glaciers as a whole. Nonetheless, one hopes these equations may serve as a first approximation for shallow glacier dynamics. In addition to dynamics, another important physical consideration of glaciers is the relationship between temperature and viscosity, which follows an Arrhenius relationship (Cuffey and Paterson, 2010). However, in the context of Icelandic glaciers like Langjökull, this is not consequential since they are temperate (i.e., their temperature is at melting point).

4.2.2 Bayesian hierarchical model

In this section, we provide an overview and set–up of the BHM employed in addition to notation for the key parameters, both statistical and physical. The reader is referred, however, to Table 4.1 for a summary of the model parameters utilized and a schematic illustrating the BHM in Figure 4.5. We use index i to refer to spatial coordinates (for this model space is assumed to be discretized into squares) and index j to refer to time coordinates. Furthermore, the notation $S_{\cdot,j}$ refers to the surface elevation at all spatial coordinates for a particular time index j. Keeping in line with the Bayesian hierarchical modeling framework from Wikle (2016) and Cressie and Wikle (2011), we delineate the models used for the data level, process level, and parameter level. The primary inferential goals are to infer physical process parameters (i.e., ice viscosity and basal sliding) and to predict the height of the glacier at various time points and spatial locations besides those that have been observed (aligned to a grid for which we have bedrock and initial surface height conditions). Within the Bayesian framework, all inferential goals may be achieved by determining the posterior distribution of these quantities (i.e., their probability distributions conditioned on observed data).

At the *data level*, the observed height for each grid point is modeled with a normal distribution (denoted with the notation $\mathrm{N}(\mu,\tau^2)$, where μ is the mean and τ^2 is the variance), where the mean is the physical process value, and the variance is assumed to be known. In particular

it is assumed that $Y_{ij} \sim \mathrm{N}(S_{ij}, \sigma^2)$, where Y_{ij} is the observed surface elevation of the glacier at location i and time index j, S_{ij} is the latent (i.e., unobserved) surface elevation at location i and time index j (equivalent to sum of the glacier thickness and bedrock level), and σ^2 is the variance of the measurement errors for the surface height observations, a fixed and known quantity. The number of observed spatial indices is assumed to be much smaller than the number of total spatial indices modeled at the latent level.

At the *process level*, $S_{.,j} = f(S_0, B, \dot{b}, C_0\gamma, j) + X_j$, where f is a numerical solution to the SIA at time index j, and X_j is an error-correcting process at time index j. A finite difference version of the SIA PDE is described in full detail in Appendix A. In principle, however, the function f may be derived from other numerical solvers. Additionally, it should be made clear that f is the output of a numerical solver for the underlying dynamics Also, S_0 denotes the glacier surface elevation values at the initial time point, which are assumed to be known; e.g., with high precision light detection and ranging (LIDAR) initial conditions provided by the Institute of Earth Sciences at the University of Iceland. $\dot{b}_{.,j}$ is the mass balance field for time index j at all the grid points, which is assumed to be fixed and known for the purpose of this analysis. B is the ice viscosity parameter and $C_0\gamma$ is the basal sliding field, which itself is parametrized with $\mu_{\max}$ as in equation (16) of Bueler et al. (2005) and, furthermore, is static in time. For compact notation, θ is used to refer to B in test cases B-D and $(B, \mu_{\max})$ jointly in test case E.

Since we believe numerical errors will accumulate over time (Bueler et al., 2005), we define the error-correcting process as follows: $X_{j+1} = X_j + \varepsilon_{j+1}$, where ε_{j+1} is $MVN(0, \Sigma)$. (MVN stands for multivariate normal, and the first argument is the mean and the second is the covariance.) Σ is block diagonal, with three blocks for indices corresponding to the margin, interior, and dome of the glacier (the margin is defined as the last grid squares before the glacier drops to 0 thickness, and the dome is the origin grid square), respectively. Each block is defined from a squared exponential kernel with the same length scale, denoted by ϕ, but distinct marginal variances, $\sigma^2_{\text{interior}}$, σ^2_{margin} and σ^2_{dome}. The motivation for using different marginal variance parameters is to account for the widely different errors exhibited at the dome, interior, and margin, as is demonstrated by Bueler et al. (2005) and Jarosch et al. (2013). This error-correcting process leads to a tractable likelihood function, as is

shown in Appendix B.

Finally, at the *parameter level*, B and $\mu_{\max}$ are endowed with truncated normal distributions as priors. B has a normal prior with mean 3.5×10^{-24}, standard deviation 3×10^{-24}, truncated to have support $[1,70] \times 10^{-25}$. $\mu_{\max}$ has a normal prior with mean 3×10^{-11} and standard deviation 1×10^{-11}, truncated to have support $[1,70] \times 10^{-12}$. (Units are $s^{-1}Pa^{-3}$ for ice viscosity and $Pa^{-1}ms^{-1}$ for basal sliding.) The prior supports for B and μ_{max} provide plausible values for temperate ice caps.

It is prudent to discuss the motivations and justifications of the various modeling choices employed in the model previously delineated. The data level is assumed to have independent normal errors with fixed variance; this is justified because of the uniformity of the measuring technology used from site to site (e.g., digital GPS) and symmetry of errors. On the other hand, the precise functional form of the data level is chosen somewhat arbitrarily as a Gaussian, which affords one analytical convenience. Similarly, the error-correcting process at the process level uses a zero mean Gaussian process with a parameterized covariance kernel (e.g., squared exponential), mostly as an analytically manageable way to induce spatial correlation in the error-correcting process. Spatial correlation in numerical errors has been demonstrated, for instance, in Bueler et al. (2005).

Moreover, it is appropriate to consider potential variations of this model for slightly different scenarios; naturally, these could fall into: alternate choices of covariance kernel at the process level (e.g., Matérn, to allow for a less smooth error-correcting process) and varying errors at the data level, to account for compaction or densification that occurs between seasons. For the latter, a suggestion is to use conjugate inverse-gamma distributions for the variances, so that sampling can be accomplished with a Gibbs sampler. Additionally, as aforementioned, one can conceivably use any numerical solver for a PDE at the process level. Future variations may consider utilizing non-zero mean Gaussian processes for the error correction process, which may be more computationally costly yet perhaps more realistic. Generally, this model can be adapted to any science or engineering system that is driven by physically meaningful parameters, whose dynamics are solved by noisy numerical methods, and for which noisy and continuous data are collected with well probed errors.

The mathematical details for how to do posterior computation within this model are given in Appendix B, which includes a derivation of

Parameter Name	Symbol
Time index	j
Spatial index	i
All spatial points for a time index	$., j$
ice viscosity	B
Basal sliding	$C_0\gamma$
Max basal sliding	$\mu_{\max}$
Physical parameters	θ
Measurement error	σ
Error-correcting covariance matrix	Σ
Error-correcting parameters	$(\sigma_{\text{dome}}, \sigma_{\text{interior}}, \sigma_{\text{margin}}, \phi)$
Mass balance field	$\dot{b}_{.,j}$
Initial surface elevation	S_0

Table 4.1. A summary of main parameters and notation utilized.

an approximation to the log-likelihood that allows for computational efficiency. In summary, we compute the posterior of physical parameters directly on a grid since there are at most two physical parameters, and we use samples from the posterior distribution of physical parameters to generate predictions for glacier thickness in the future.

4.3 Experiments to assess the Bayesian hierarchical model

4.3.1 Analytical solutions

In Bueler et al. (2005), analytical solutions to the SIA are presented as benchmarks for numerical solvers of the SIA. As opposed to using other benchmarks such as the EISMINT experiment (Payne et al., 2000), which itself is based on numerical modeling and hence subject to numerical errors, the benchmark solutions provided in this work can be treated as ground truth to compare to. (This is in the sense that these are exact solutions of the SIA, but it must be stressed that the SIA is an approxima-

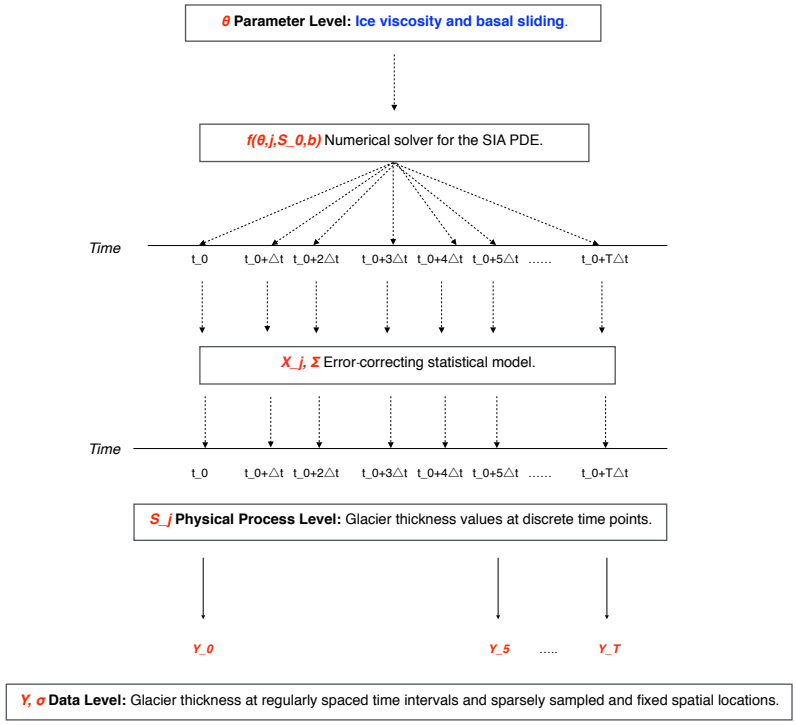

Figure 4.5. Schematic of the physical-statistical BHM that has been constructed based on the SIA PDE. The main parameters and variables for each module of the physical-statistical model are highlighted in red. The main levels of a physical-statistical model shown in Figure 4.4 are displayed here, along with the parameters and variables describing each level.

tion of the true physical dynamics governing a glacier.) These analytical solutions serve as a basis for simulating data sets to validate the Bayesian hierarchical approaches developed in this paper. In other words, the exact analytical solutions provide the latent process in the BHM, conditioning on given initial conditions and mass balance functions. Hence to simulate data from the BHM, one can bypass the need to numerically solve the PDE and introduce errors.

We make use of four analytical solutions from Bueler et al. (2005) that are summarized here, but the reader is referred to the original paper for the exact mathematical formulation and derivation of these analytical solutions. All of the analytical solutions assume a flat bedrock. Test case B includes no mass balance or basal sliding, and, consequently, the motion of the glacier is only attributable to deformation due to gravity. Test case C makes use of a mass balance field that is inversely proportional

to time and directly proportional to thickness, but there is no basal sliding field modeled. Similarly, test case D utilizes a mass balance field with no basal sliding field modeled. In distinction from test case C, however, the mass balance field of test case D is such that the overall solution for glacial thickness is periodic in time. Finally, in contrast to the other tests, test case E has a spatially varying basal sliding field, yet the overall solution is static in time. Note that test A was not utilized in this study because it is a steady state solution without a varying mass balance or basal sliding field.

4.3.2 Simulation study test details

Conditions of the simulation study have been chosen as to closely emulate the data collected at Langjökull ice cap by the Institute of Earth Sciences at the University of Iceland (IES-UI). In particular, 20 years of data are assumed, which is comparable to data provided by the IES. 25 fixed measurement sites are used for bi-annual surface elevation measurements, which are geographically distributed on the glacier in a manner that is comparable to the real data provided by the IES-UI. Figure 4.6 illustrates the locations of these measurement sites on the glacier. Surface elevation measurements for these sites are taken twice a year (i.e., for summer and winter mass balance measurements). The surface elevation measurements are generated by adding Gaussian noise (zero mean, unit variance) to the analytical solutions at the spatio-temporal coordinates of the fixed measurement sites. The choice of unit variance is larger than the errors produced by digital-GPS measurements. Remaining physical parameters were chosen using the values from Bueler et al. (2005) Table 2 to allow for comparisons to this work and the EISMINT I experiment (Payne et al., 2000).

4.4 Results

Validation and diagnostics of the BHM were achieved through a combination of an assessment of posterior probability intervals, a test of the predictive error of thickness values 100 years from the initial time point t_0, and a comparison between observed and expected values for numerical errors based on the error-correcting process utilized. As is discussed in

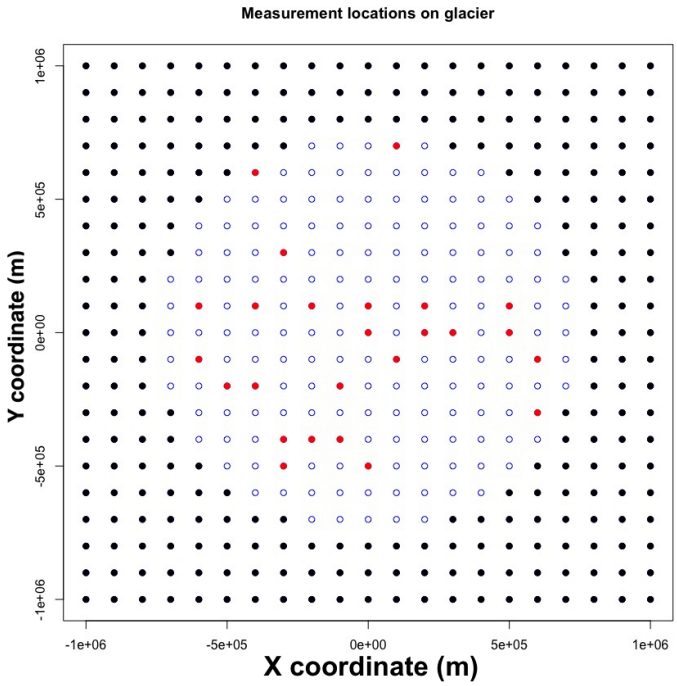

Figure 4.6. An illustration marking the 25 measurement sites on the glacier. This is a top level view of the glacier, where the blue points indicate the glacier, the red points indicate the measurement locations, and the black points indicate locations surrounding the glacier with no glacial thickness.

more detail below, these assessments suggest that the BHM is useful for inference of posterior probability distributions for physical parameters, prediction of future glacial thickness values on the order of 100 years, and the modeling of numerical errors at the margin, interior, and dome of the glacier.

Table 4.2 contains posterior credibility intervals for ice viscosity in test cases B-D. A 3-sd credibility interval was computed with mean +/- 3 standard deviations of the posterior samples. In all of these test cases, the 3-sd credibility interval covers the actual ice viscosity. Furthermore, as is apparent in Table 4.3, the predictive error, relative to thickness values on the order of a kilometer, appears be small overall, particularly at the interior; predictive error is the root mean squared difference between predictions and the exact analytical values for each of the test cases. Note that test E was not included with the predictive checks since it is static in time. Consistent with Bueler et al. (2005) and Jarosch et al.

(2013), however, errors are greatest at the margin and dome of the glacier (evident in Figure 4.8). Nonetheless, the predictive distributions cover the actual thicknesses even at these extremes. This illustrates the utility of the BHM for accounting for errors induced by the numerical solution of the SIA. Additionally, an illustration comparing the posterior and prior distributions for test case D is shown in Figure 4.9.

To investigate the frequentist properties of the posterior probability distribution for ice viscosity (i.e., its performance under repeated sampling of data), 500 simulations were completed under repeated sampling of the surface elevation data at the 25 fixed measurement sites for test cases B-D. The coverage of ice viscosity for a 3-sd interval was computed for each of the simulations, where coverage for a given interval is binary; either the actual parameter value is in the interval or it is not. For test case B, in 499 of 500 simulations the 3-sd credibility interval covered the actual value of ice viscosity. In test cases C and D, the 3-sd credibility interval covered the actual value of ice viscosity in all of the simulations. This suggests that the frequentist coverage probability of the credibility interval is at least 99 percent.

For test case E, one assumes that the field is described by parameterized equation (16) of Bueler et al. (2005). That is, in polar coordinates with radius r and angle Θ:

$$C_0\gamma(r,\Theta) = \frac{\mu_{\max} 4(r-r_1)(r_2-r)4(\Theta-\theta_1)(\theta_2-\Theta)}{(r_2-r_1)^2(\theta_2-\theta_1)^2}$$

for $\theta_1 < \Theta < \theta_2$ and $r_1 < r < r_2$, and $C_0\gamma = 0$ otherwise. In addition to ice viscosity, the inferential object of interest is the scale parameter $\mu_{\max}$. The 3-sd posterior credibility interval for B is $[1,43]$ in units of $10^{-25}s^{-1}Pa^{-3}$, and for $\mu_{\max}$ it is $[1,50]$ in units of $10^{-12}Pa^{-1}ms^{-1}$. The actual values for B and $\mu_{\max}$ are $32\times10^{-25}s^{-1}Pa^{-3}$ and $25\times10^{-12}s^{-1}Pa^{-1}ms^{-1}$, respectively. Hence, the credibility intervals cover both parameters. A figure illustrating the posterior distribution of μ_{max} is given in the supplemental materials.

While the credibility intervals achieved coverage of the actual values of the parameters, it was noticed that the posterior distribution for physical parameters and predictions are biased. Brynjarsdóttir and O'Hagan (2014) exhibit the same phenomenon in a simple physical system with a single physical parameter, and they demonstrate that the bias of a physical parameter posterior distribution reduces as better prior information

is encoded to model the difference between the output of a computer simulator of a physical system and the actual physical process values (i.e., what we have termed as an error-correcting process). To demonstrate that this also holds in the BHM presented in this paper, we consider the following comparison. To assign prior information to the error-correcting process, we consider a discrete parameter set for $\sigma^2_{\text{interior}}$, σ^2_{margin} and σ^2_{dome}: {.1,1,10,100} in units of m^2, which corresponds to different orders of magnitude for variability. In one case, we ignore prior information from Bueler et al. (2005) and put equal probability mass on the parameter space for these parameters. In the second case, we encode more realistic prior information into the scales of errors at the three regions: equal mass on 10 and 100 at the margin, equal mass on .1 and 1 at the interior, and equal mass at 1 and 10 at the dome (all units are m^2). In both cases, the parameter ϕ is fixed at 70 km to place emphasis on the scales of error. The results of inferring the posterior distribution for ice viscosity B are shown in Figure 4.10. Consistent with Brynjarsdóttir and O'Hagan (2014), the posterior distribution of the physical parameter B is much less biased when prior information is encoded into the error-correcting process.

To assess how the posterior distribution for ice viscosity evolves under different sampling plans of the data, we conducted a series of simulations in test case D under varying sampling periods. In particular, we considered data samples once every 10 years, once every 5 years, once a year, and twice a year; the resulting posteriors for ice viscosity are in Figure 4.11. The general pattern is that the bias of the posterior distributions reduces as the period gets shorter, although the posterior becomes more diffuse. The result that some posterior uncertainty does not go away with more collected data is also consistent with the results in Brynjarsdóttir and O'Hagan (2014). The particular period we chose in this analysis (data collected twice a year) was meant to model how the UI-IES glaciology team collects data, that is, twice a year due to summer and winter mass balance measurements.

To assess the accumulating error-correcting process model, we estimated the marginal variances of the error-correcting process for each of the time points with observed data in test case B, by examining the residuals formed by the difference between the numerical solver and the observed data. According to the model, the standard deviation of these

residuals at the interior of the glacier should grow as $\sqrt{\sigma^2 + t\sigma^2_{\text{interior}}}$, where t is the number of time steps (and likewise at the dome and margin). Figure 4.12 shows a match between observed and expected in this regard, and, in particular, the 99 percent confidence bands appear to cover the expected variability as time progresses. Also apparent from this figure is that, as time progresses, the errors at the margin, dome, and interior contribute more error than measurement error, which is on the order of 1 meter. Moreover, this is also evident in Table 4.4, since after 200 time steps from t_0 (i.e., 20 years), the marginal variances will be $200\sigma^2_{\text{interior}}$, $200\sigma^2_{\text{margin}}$, and $200\sigma^2_{\text{dome}}$ based on the accumulating errors model; all of these values exceed 1, the measurement variance.

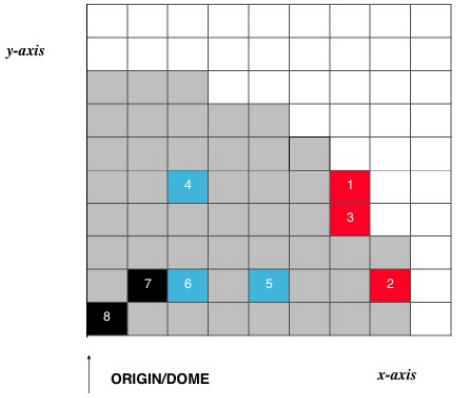

Figure 4.7. Grid map used to interpret the following box-plots in Figure 4.8. Eight randomly chosen grid points are selected for testing predictions; these are not the same as the measurement locations. Only one quadrant of the glacier is shown due to symmetry as is done in Figures 9,10, and 12 of Bueler et al. (2005), and the width of each cell is $10^5 m$. Additionally, the red squares indicate locations at or close to the margin, the blue squares indicate locations that are between the dome and margin of the glacier, and the black squares indicate locations at or close to the dome of the glacier. Moreover, glacier grid squares with non-zero thickness are shaded in grey, as to indicate the glacier location.

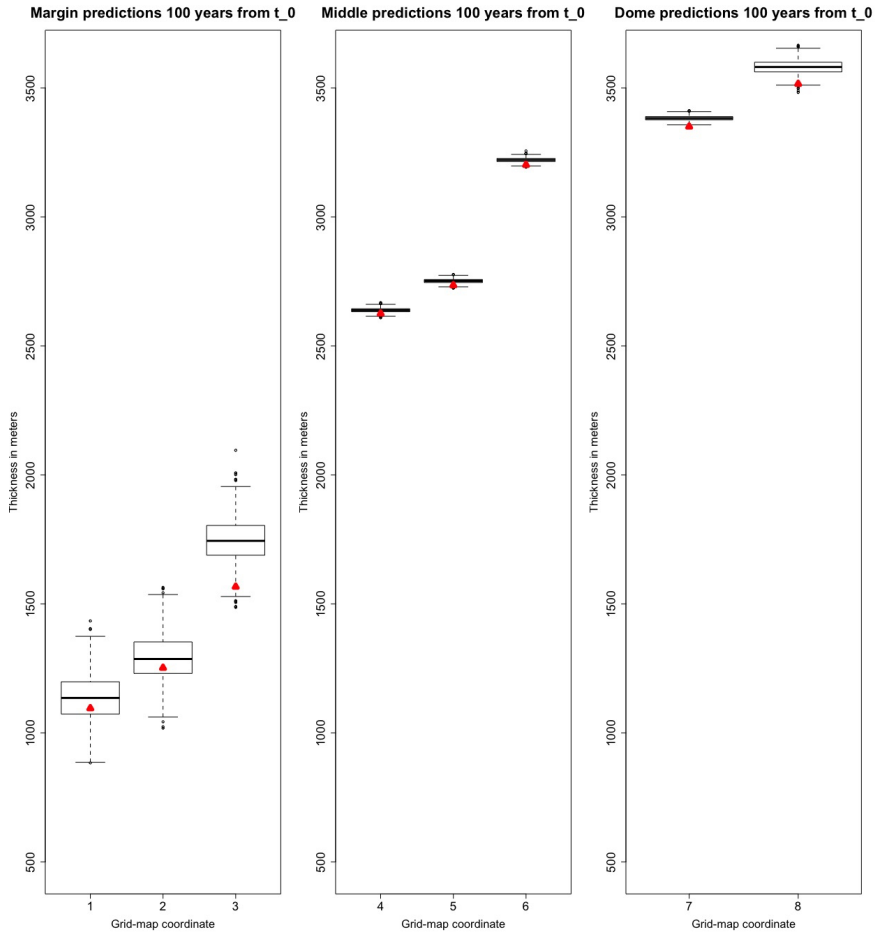

Figure 4.8. Thickness prediction samples 100 years from t_0 for test case B (i.e., no mass balance field or basal sliding). Triangles indicate the actual thickness values from the analytical solution. The first set of plots are close to the margin (red squares of Figure 4.7), the second set of plots are between the dome and margin of the glacier (blue squares of Figure 4.7), and the final set of plots are toward the dome of the glacier (black squares of Figure 4.7). Refer to Figure 4.7 for a grid map to spatially reference each of the box-plots. As can be expected according to Bueler et al. (2005), largest errors occur at the dome and the margin. Note on interpretation: the middle of each box is the median, the interquartile range is denoted by the box, and 1.5 of the interquartile range beyond the first and third quartile is illustrated with the whiskers. Those points that are more than 1.5 of the interquartile range beyond the first and third quartiles are outliers, and they are denoted with circles.

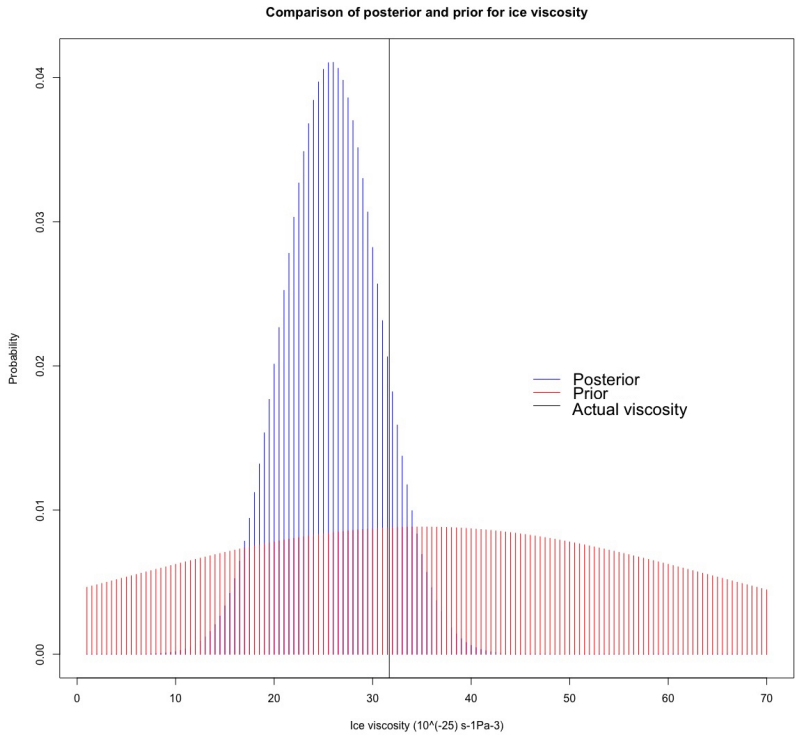

Figure 4.9. Comparison of posterior and prior distributions of ice viscosity for test case D (i.e., mass balance field producing a periodic SIA solution).

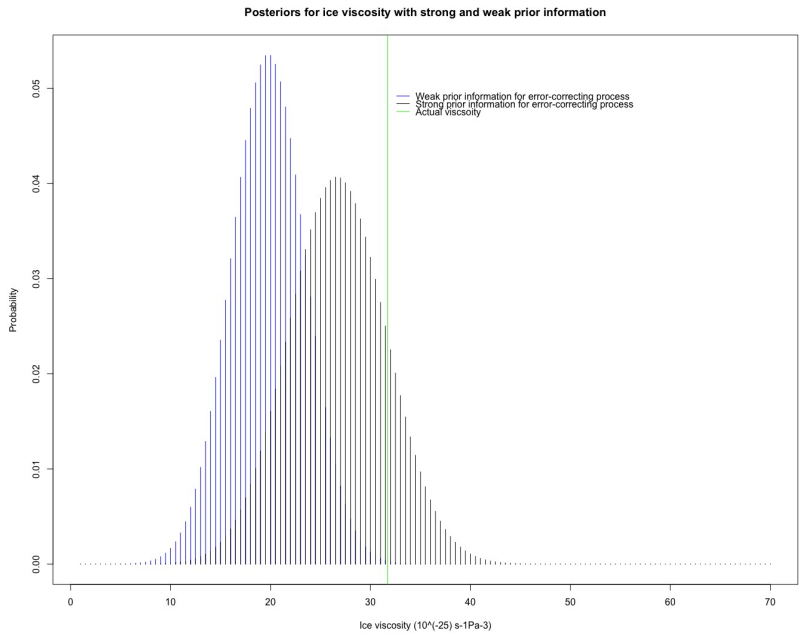

Figure 4.10. A comparison of posteriors under strong and weak prior information for the error-correcting process in test case D (i.e., mass balance field producing a periodic SIA solution); prior information for the error-correcting process leads to a less biased posterior, though with slightly more posterior uncertainty.

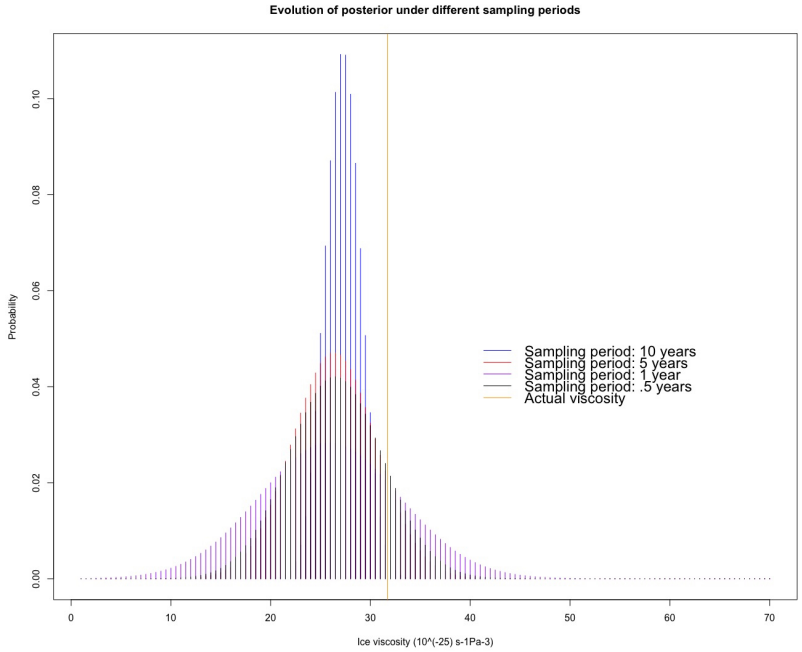

Figure 4.11. A comparison of posteriors in test case D (i.e., mass balance field producing a periodic SIA solution) under different sampling periods: data sampled once every 10 years, every 5 years, once a year, and twice a year. The general trend is that the posterior tends to become less biased as the period of sampling decreases, although the posterior becomes more diffuse. The University of Iceland Institute of Earth Sciences glaciology team takes measurements twice a year for summer and winter mass balance measurements.

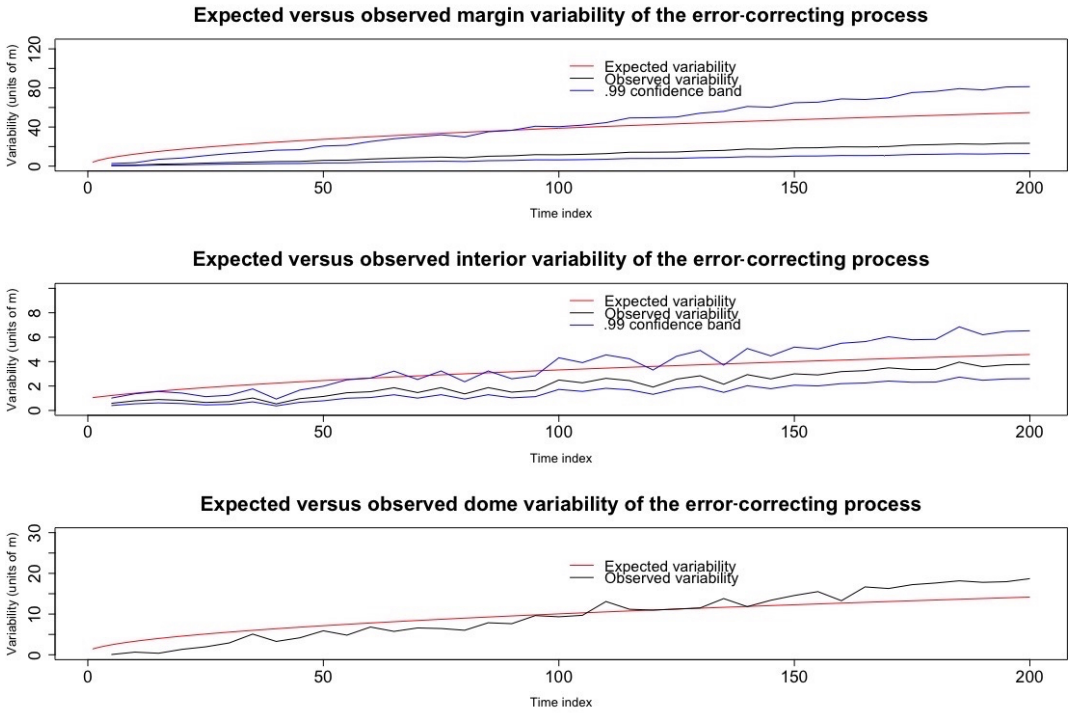

Figure 4.12. An illustration comparing the expected variability of the error-correcting process (as per the Bayesian hierarchical model) to the observed variability of residuals at the interior, margin, and dome for test case B (i.e., no mass balance field or basal sliding). These residuals are the differences between the observed data and the numerical solution.

Test Case	Actual Viscosity	3-sd Credibility Interval
Bueler B	32	$[7,34]$
Bueler C	32	$[5,33]$
Bueler D	32	$[11,42]$
Units	$10^{-25}\ s^{-1}Pa^{-3}$	$10^{-25}\ s^{-1}Pa^{-3}$

Table 4.2. Ice viscosity posterior intervals

Test Case	Dome RMSE	Interior RMSE	Margin RMSE
Bueler B	66	20	75
Bueler C	76	22	82
Bueler D	1.4	17	49
Units	m	m	m

Table 4.3. Results of prediction at $t_0 + 100$ years. RMSE stands for root mean squared error. This is calculated by taking the average of the squared difference between the actual glacial thickness values and predicted glacial thickness values, and then taking the square root.

Test Case	σ^2_{dome}	$\sigma^2_{\text{interior}}$	σ^2_{margin}	ϕ
Bueler B	1	.1	15	71
Bueler C	1	.15	15	64
Bueler D	.1	.1	10	62
Bueler E	.1	.1	10	60
Units	sq. m	sq. m	sq. m	km

Table 4.4. Error-correcting process hyper-parameters; σ^2_{dome} is the error-correcting process variance at the dome, $\sigma^2_{interior}$ is the error-correcting process variance at the interior, σ^2_{margin} is the error-correcting process variance at the margin, and ϕ is the length scale parameter.

4.5 Summary, discussion, and future work

The primary contribution of this work has been to construct a BHM for glacier flow based on the SIA that operates in two spatial dimensions and time, which successfully models numerical errors induced by a numerical solver that accumulate with time and vary spatially. This BHM leads to full posterior probability distributions for physical parameters as well as a principled method for making predictions that takes into account both numerical errors and uncertainty in key physical parameters. Furthermore, the BHM operates in two spatial dimensions and time, which, to our knowledge, is new to the field of glaciology. An additional contribution is the derivation of a novel finite difference method for solving the SIA. When tested using simulated data sets based on analytical solutions to the SIA from Bueler et al. (2005), the results herein indicate that our approach is able to infer meaningful probability distributions for glacial parameters, and, furthermore, this approach makes probabilistic predictions for glacial thickness that adequately account for the error induced by using a numerical solver of the SIA. A future goal is to create an R package for fitting a generalized version of the model used within this work, where the function $f(.)$ is provided by the user. This will allow glaciologists to extend the modeling approach we have developed to other similar scenarios in which the physical dynamics are more complex than the SIA. An additional scenario for which this package can be useful is when the numerical method is not a finite difference method; e.g., a FEM. To this end, we will attempt to utilize emulator inference (Hooten et al., 2011); this will be crucial to ensure that the methodology scales well computationally, since each posterior sample requires a forward PDE solve. Finally, and perhaps most importantly, future work will involve the application of the modeling and methodologies developed within this paper to real data collected by the IES-UI, which includes bedrock elevation and mass balance measurements.

4.6 Appendix A: finite difference method

Here a finite difference scheme is derived for the SIA PDE. The overarching strategy in developing this finite discretization scheme is to take a second-order Taylor expansion for $H(x,y,t)$ with x,y fixed, and then

equate the resultant time derivatives, H_t and H_{tt}, to functions of spatial derivatives by using the original SIA PDE. That is, one starts with the approximation $H(x,y,t+\Delta t) \approx H(x,y,t) + H_t(x,y,t)\Delta t + H_{tt}(x,y,t)\Delta t^2/2$ and uses the first equation of Section 4.2.1 to write H_t and H_{tt} in terms of spatial derivatives. Finally, central differences in space are substituted for the spatial derivatives. This finite difference scheme is motivated by the Lax-Wendroff (Hudson, 1998) method, which is generally better than finite difference methods that use only a single order Taylor expansion (indeed, in the advection-diffusion equation such methods may be unconditionally unstable).

In the following derivations note that the subscripts mean 'derivative with respect to' (e.g., H_t means derivative of H with respect to t).

$$\begin{aligned} H_t &= -[\bar{u}H]_x - [\bar{v}H]_y + \dot{b}, \\ H_{tt} &= -[\bar{u}H]_{xt} - [\bar{v}H]_{yt} + \ddot{b}. \end{aligned}$$

Now we solve for these derivatives in terms of spatial derivatives in $H(x,y,t)$, the glacier thickness, and $R(x,y)$, the bedrock level. The derivation makes repeated use of the differentiation rule for products, the chain rule for differentiation, and equality of mixed partials (e.g., $H_{xt} = H_{tx}$).

$$\begin{aligned} -[\bar{u}H]_x &= -C_0\gamma\rho g T_1 + \frac{2B}{n+2}(\rho g)^n T_2 \\ T_1 &= [2HH_x(H_x+R_x) + H^2(H_{xx}+R_{xx})] \\ T_2 &= T_{21} + T_{22} \\ T_{21} &= [\alpha^{n-1}]_x [H^{n+2}H_x + H^{n+2}R_x] \\ T_{22} &= \alpha^{n-1}[(n+2)H^{n+1}H_x^2 + (n+2)H^{n+1}H_xR_x + H^{n+2}H_{xx} + H^{n+2}R_{xx}] \end{aligned}$$

By symmetry in x and y, $-[\bar{v}H]_y$ can be analogously derived:

$$\begin{aligned} -[\bar{v}H]_y &= -C_0\gamma\rho g T_3 + \frac{2B}{n+2}(\rho g)^n T_4 \\ T_3 &= [2HH_y(H_y+R_y) + H^2(H_{yy}+R_{yy})] \\ T_4 &= T_{41} + T_{42} \\ T_{41} &= [\alpha^{n-1}]_y [H^{n+2}H_y + H^{n+2}R_y] \\ T_{42} &= \alpha^{n-1}[(n+2)H^{n+1}H_y^2 + (n+2)H^{n+1}H_yR_y + H^{n+2}H_{yy} + H^{n+2}R_{yy}] \end{aligned}$$

Derivatives $[\alpha^{n-1}]_x$ and $[\alpha^{n-1}]_y$:

$$[\alpha^{n-1}]_x = \frac{n-1}{2}(S_x^2+S_y^2)^{\frac{n-3}{2}}(2S_xS_{xx}+2S_yS_{yx})$$

$$[\alpha^{n-1}]_y = \frac{n-1}{2}(S_x^2+S_y^2)^{\frac{n-3}{2}}(2S_yS_{yy}+2S_xS_{xy})$$

Now we derive $-[\bar{u}H]_{xt}$:

$$\begin{aligned}
-[\bar{u}H]_{xt} &= -C_0\gamma\rho g T_{1_t} + \frac{2B}{n+2}(\rho g)^n T_{2_t} \\
T_{1t} &= [2H_tH_x^2 + 4HH_xH_{xt} + 2HH_{xt}R_x + 2HH_xR_{xt} + 2H_tH_xR_x + T_1* + T_2*] \\
T_2* &= 2HH_tH_{xx} + H^2H_{xxt} \\
T_1* &= 2HH_tR_{xx} + H^2R_{xxt} \\
T_{2t} &= [T_5 + T_6 + T_7 + T_8] \\
T_5 &= [\alpha^{n-1}]_{xt}H^{n+2}H_x \\
T_6 &= [\alpha^{n-1}]_{xt}H^{n+2}R_x \\
T_7 &= [\alpha^{n-1}]_x[(n+2)H^{n+1}H_tH_x + H^{n+2}H_{xt} + (n+2)H^{n+1}H_tR_x + H^{n+2}R_{xt}] \\
T_8 &= [\alpha^{n-1}]_{xt}H^{n+2}H_x + \alpha_x^{n-1}(n+2)H^{n+1}H_tH_x + \alpha_x^{n-1}H^{n+2}H_{xt} \\
&+ [\alpha^{n-1}]_{xt}H^{n+2}R_x + \alpha_x^{n-1}(n+2)H^{n+1}H_tR_x + \alpha_x^{n-1}H^{n+2}R_{xt} \\
&+ [\alpha^{n-1}]_t(n+2)H^{(n+1)}H_x^2 + \alpha^{n-1}(n+2)(n+1)H^nH_tH_x^2 \\
&+ \alpha^{n-1}(n+2)H^{n+1}2H_xH_{xt} \\
&+ [\alpha^{n-1}]_t(n+2)H^{n+1}H_xR_x \\
&+ \alpha^{n-1}(n+2)(n+1)H^nH_tH_xR_x \\
&+ \alpha^{n-1}(n+2)H^{n+1}H_{xt}R_x \\
&+ \alpha^{n-1}(n+2)H^{n+1}H_xR_{xt} \\
&+ [\alpha^{n-1}]_tH^{n+2}H_{xx} \\
&+ \alpha^{n-1}(n+2)H^{n+1}H_tH_{xx} \\
&+ \alpha^{n-1}H^{n+2}H_{xxt} \\
&+ [\alpha^{n-1}]_tH^{n+2}R_{xx} \\
&+ \alpha^{n-1}(n+2)H^{n+1}H_tR_{xx} \\
&+ \alpha^{n-1}H^{n+2}R_{xxt}
\end{aligned}$$

Note that terms with a time derivative of bedrock such as R_{xt} can be set to 0 since R is assumed to be static in time. However, we keep the

time derivatives for R in the above equation for full generality in case a scenario is revisited where this does not hold. Next we derive $[\alpha^{n-1}]_t$:

$$[\alpha^{n-1}]_t = \frac{n-1}{2}(S_x^2+S_y^2)^{\frac{n-3}{2}}(2S_xS_{xt}+2S_yS_{yt})$$

Next we derive $[\alpha^{n-1}]_{tx}$:

$$\begin{aligned}[\alpha^{n-1}]_{tx} &= \frac{n-1}{2}[\frac{n-3}{2}(S_x^2+S_y^2)^{\frac{n-5}{2}}(2S_xS_{xx}+2S_yS_{yx})(2S_xS_{xt}+2S_yS_{yt}) \\ &+ (S_x^2+S_y^2)^{\frac{n-3}{2}}(2S_{yx}S_{yt}+2S_yS_{ytx}+2S_{xx}S_{xt}+2S_xS_{xtx})]\end{aligned}$$

Next we derive $[\alpha^{n-1}]_{ty}$:

$$\begin{aligned}[\alpha^{n-1}]_{ty} &= \frac{n-1}{2}[\frac{n-3}{2}(S_x^2+S_y^2)^{\frac{n-5}{2}}(2S_xS_{xy}+2S_yS_{yy})(2S_xS_{xt}+2S_yS_{yt}) \\ &+ (S_x^2+S_y^2)^{\frac{n-3}{2}}(2S_{xy}S_{xt}+2S_xS_{xty}+2S_{yy}S_{yt}+2S_yS_{yty})]\end{aligned}$$

Note that $S_{tx} = R_{tx} + H_{tx} = H_{tx}$ since R is assumed to be fixed as a function of t. Note that the same argument holds for other derivatives of S with respect to t. Next we derive $H_{tx}, H_{txx}, H_{ty}, H_{tyy}, H_{tyx}$:

$$\begin{aligned}H_{tx} &= -[\bar{u}H]_{xx} - [\bar{v}H]_{yx} + \dot{b}_{tx} \\ H_{txx} &= -[\bar{u}H]_{xxx} - [\bar{v}H]_{yxx} + \dot{b}_{txx} \\ H_{ty} &= -[\bar{u}H]_{xy} - [\bar{v}H]_{yy} + \dot{b}_{ty} \\ H_{tyy} &= -[\bar{u}H]_{xyy} - [\bar{v}H]_{yyy} + \dot{b}_{tyy} \\ H_{tyx} &= -[\bar{u}H]_{xxy} - [\bar{v}H]_{yyx} + \dot{b}_{tyx}\end{aligned}$$

Hence, these partial derivatives allow us to substitute purely spatial derivatives into the forward in time approximation for H. Without loss of generality, we use a central difference approximation for all spatial derivatives. Furthermore, we used $\Delta_t = .1$ years and $\Delta_x = \Delta_y = 10^5$ m for the analysis in this paper. In total, 441 grid squares were modeled (i.e., 21 by 21) with the dome grid square at the origin. While a coarse grid was chosen for computational convenience, it is expected that numerical errors will go to zero as the grid width goes to zero, as is demonstrated both by Bueler et al. (2005) and Jarosch et al. (2013).

4.7 Appendix B: model fitting

In the following subsections, we go through the key details regarding Bayesian computation for the model used in this work. Assume n total grid points are modeled, of which $m << n$ are observed. Let $X_j \in \mathbb{R}^n$ be the error-correcting process at time j, $S_j \in \mathbb{R}^n$ be the latent glacier surface values at time j, $f(\theta, j) \in \mathbb{R}^n$ be shorthand for the output of the numerical solver at time point j, and ε_j be an independent and identically distributed (IID) multivariate normal noise term at time j with mean 0 and covariance matrix Σ. MVN stands for multivariate normal, and the first argument is the mean and the second is the covariance.) Furthermore, assume that data are collected regularly at every k_{th} time point, such that one observes $Y_k, Y_{2k}, ..., Y_{Nk} \in \mathbb{R}^m$, and the corresponding observation error $Z_k, Z_{2k}, ...Z_{Nk}$ is IID $MVN(0, \sigma^2 I)$. For convenience, we denote Nk as T. Finally, let $A \in \mathbb{R}^{m \times n}$ be a matrix which selects the grid squares of the latent process S that are observed; that is, its rows are unit basis vectors corresponding to those indices that are observed.

4.7.1 Calculating the likelihood $p(Y_k, ..., Y_T|\theta)$

In this subsection, we derive both the likelihood of the observed data: $p(Y_k, ..., Y_T|\theta)$ and an approximation to the likelihood.

Though Section 4.2.2 specifies the BHM in greater detail, the process and data levels of the BHM (i.e., conditioning on θ) are concisely written as follows.

$$\begin{aligned} X_j &= X_{j-1} + \varepsilon_j \\ S_j &= f(\theta, j) + X_j \\ Y_{ck} &= AS_{ck} + Z_{ck} \end{aligned}$$

Assume $j \in 1, 2, ...T$ and $c \in 1, 2, ..N$; hence there are N total spatial vectors observed with a period of length k. Furthermore, X_1 is marginally $MVN(0, \Sigma)$. That is, the process level vectors, S_j, are modeled conditional on the parameter level and the error-correcting process. The data level vectors, Y_{ck}, are generated conditional on the process level S_{ck}. Throughout the following, we condition on θ being fixed.

4.7.2 The exact likelihood

Conditional on θ, the distribution of $(Y_k,...,Y_T)$, viewed as one long random vector, is multivariate normal. Also, conditional on θ, the mean of $(Y_k,...,Y_T)$ is $(Af(\theta,k),..Af(\theta,T))$ because both $(X_k,...,X_T)$ and $(Z_k,...,Z_T)$ have mean 0. It suffices to thus derive the covariance matrix for $(Y_k,...,Y_T)$ conditional on θ. To do this, we note that $Var(Y_{ck}) = Var(AS_{ck}+Z_{ck}) = Var(AS_{ck})+Var(Z_{ck}) = [A(ck\Sigma)A^\intercal]+\sigma^2 I$. Additionally, for $a<b$:

$$\begin{aligned} Cov(Y_a,Y_b) &= Cov(AS_a+Z_a,AS_b+Z_b) \\ &= Cov(AS_a,AS_b) \\ &= Cov(A[f(\theta,a)+X_a],A[f(\theta,b)+X_b]) \\ &= Cov(AX_a,AX_b) \\ &= Var(AX_a) \\ &= [A(a\Sigma)A^\intercal] \end{aligned}$$

Therefore, the covariance matrix for the observed data can be written as $M\otimes(A\Sigma A^\intercal)+\sigma^2 I$, where $M_{ij} = kmin(i,j)$ and $M\in\mathbb{R}^{N\times N}$. This is a useful matrix representation because the inverse of M is band-limited and sparse, for which there exist efficient computationally efficient linear algebraic routines (Rue, 2001).

4.7.3 An approximation to the likelihood

The joint distribution $p(Y_k,...,Y_T|\theta)$ can be written as $p(Y_k|\theta)p(Y_{2k}|Y_k,\theta)$ $...p(Y_T|Y_k,..,Y_{(N-1)k},\theta)$. Since we expect that the data level errors are quite small (on the order of 1m) in comparison to the overall surface elevation measurements (on the order of 1 km), we can approximate $p(S_{(c-1)k}|Y_k,..,Y_{(c-1)k},\theta)$ with $p(S_{(c-1)k}|Y_{(c-1)k},\theta)$. Consequently, $p(Y_{ck}|Y_k,..,Y_{(c-1)k},\theta)$ will be close to $p(Y_{ck}|Y_{(c-1)k},\theta)$. From the above recursive relationship, we can write:

$$Y_{ck} = Y_{(c-1)k}+A[f(\theta,ck)-f(\theta,(c-1)k)]+Z_{ck}-Z_{(c-1)k}+\sum_{j=(c-1)k+1}^{ck} A\varepsilon_j$$

This expression motivates approximating $p(Y_{ck}|Y_k,..,Y_{(c-1)k},\theta)$ as MVN distribution with mean $Y_{(c-1)k}+A[f(\theta,ck)-f(\theta,(c-1)k)]$ and covari-

ance matrix $A(k\Sigma)A^\intercal + 2\sigma^2 I$. A similar expression shows that $p(Y_k)$ is multivariate normal with mean $Af(\theta,k)$ and covariance matrix $A(k\Sigma)A^\intercal + \sigma^2 I$. Nonetheless, we must be clear: $p(Y_{ck}|Y_{(c-1)k},\theta)$ does not exactly follow a MVN with mean $Y_{(c-1)k} + A[f(\theta,ck) - f(\theta,(c-1)k)]$ and covariance matrix $A(k\Sigma)A^\intercal + 2\sigma^2 I$; this is because $Z_{(c-1)k}$ and $Y_{(c-1)k}$ are dependent. A simple example illustrating this approximation is presented in the supplemental materials.

4.7.4 Posterior computation

Posterior inference is accomplished with grid sampling (Gelman et al., 2013); this approach directly computes the posterior distribution, $p(\theta|Y_k,...,Y_T)$ of the parameter, proportional to $p(Y_k,...,Y_T|\theta)p(\theta)$, on a grid of plausible values. The likelihood is derived in the previous subsection. Parameters for the error-correcting process are selected using knowledge elicited from the studies of Bueler et al. (2005). To verify the sensitivity of grid sampling to the grid width, three grid widths for B are considered: .25, .50, and 1, and the grid's range is from [1,70] (all in units of $10^{-25}\ s^{-1}Pa^{-3}$). The summary statistics for generating 10^6 posterior samples from more to less fine $(.25,.50,1)$ are given below:

- Min: (5.25,5.00,6.00)
- 1st Quartile: (23.8,23.5,24.0)
- Median: (27.0,26.5, 27.0)
- Mean: (27.1,26.7,27.1)
- 3rd Quartile: (30.5,30.0,30.0)
- Max: (51.50,49.0,51.0)

The similarity of summary statistics across grid widths indicates that the posterior samples are not very sensitive to grid width; a grid width of .50 was used for the analyses within. Moreover, the posterior samples in this check were generated for test case D (i.e., mass balance field producing a periodic solution to the SIA).

4.7.5 Predictions of glacial surface elevation

In this section, we give details for how to make predictions under the proposed Bayesian model. Denote $S_{T_{end}} \in \mathbb{R}^n$ for future glacier elevation values we want to make a prediction for at time point T_{end}. Our goal is to approximate the posterior predictive distribution $p(S_{T_{\text{end}}}|Y_k,...Y_T)$. To make this computationally simple, our first assumption (as in the computation of the likelihood) is to suggest that $p(S_T|Y_k,...Y_T,\theta)$ is approximately equivalent to $p(S_T|Y_T,\theta)$. This is because relative to the overall glacier surface elevation values (an average of about 2000 m), the measurement errors are small, on the order of 1 m. Moreover, based on the model specified above, we know that $S_{T_{\text{end}}} = X_T + \sum_{j=T+1}^{T_{\text{end}}} \varepsilon_j + f(\theta, T_{\text{end}})$. This suggests the following iterative procedure to generate a posterior sample for the prediction of $S_{T_{\text{end}}}$: for each independent sample θ_l from $p(\theta|Y_k,...,Y_T)$, generate a sample from a multivariate normal whose mean is 0 and covariance given by $(T_{\text{end}} - T)\Sigma$, add the sample to $f(\theta_l, T_{\text{end}})$, and then add this sum to a sample from $p(X_T|\theta = \theta_l, Y_T)$.

We must then determine how to sample from the distribution of $p(X_T|\theta = \theta_l, Y_T)$. Let $X_{\text{Tobs}} \in \mathbb{R}^m$ be a subvector of X_T corresponding to the indices that are observed at the data level, and $X_{\text{Tpred}} \in \mathbb{R}^{n-m}$ be a subvector of X_T corresponding to unobserved indices. The distribution for $p(X_{\text{Tobs}}|\theta, Y_T)$ is multivariate normal due to conjugacy. The precision, denoted by Q_{obs}, is $\sigma^{-2}I + [A(T\Sigma)A^\intercal]^{-1}$. The mean, denoted by μ_{obs}, is $Q_{\text{obs}}^{-1}(\sigma^{-2}IY_T + [A(T\Sigma)A^\intercal]^{-1}Af(\theta,T)) - Af(\theta,T)$. $p(X_{\text{Tpred}}|X_{\text{Tobs}}, \theta, Y_T)$ is multivariate normal, whose mean and variance can be derived with the well-known conditional multivariate normal formula, as in Theorem 2.44 of Wasserman (2010). That is, the mean is $T\Sigma_{\text{pred,obs}}Q_{obs}$ and the variance is $T\Sigma_{\text{pred,pred}} - T\Sigma_{\text{pred,obs}}Q_{obs}T\Sigma_{\text{obs,pred}}$. Here, $\Sigma_{\text{pred,obs}}$ is the submatrix of Σ that contains the rows of Σ that correspond to the indices that are to be predicted, and the columns correspond to the indices which are observed. $\Sigma_{\text{obs,pred}}$ is analogously defined.

5 Paper 2

A hierarchical spatio-temporal statistical model motivated by glaciology

Gopalan, G., Hrafnkelsson, B., Wikle, C.K., Rue, H., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F.: A Hierarchical Spatio-Temporal Statistical Model Motivated by Glaciology. Published in the Journal of Agricultural, Biological, and Environmental Statistics, 2019. *Winner of the 2019 American Statistical Association (ASA) Section on Bayesian Statistical Science (SBSS) Laplace Award.*

Abstract: In this paper, we extend and analyze a Bayesian hierarchical spatio-temporal model for physical systems. A novelty is to model the discrepancy between the output of a computer simulator for a physical process and the actual process values with a multivariate random walk. For computational efficiency, linear algebra for bandwidth-limited matrices is utilized, and first-order emulator inference allows for the fast emulation of a numerical partial differential equation (PDE) solver. A test scenario from a physical system motivated by glaciology is used to examine the speed and accuracy of the computational methods used, in addition to the viability of modeling assumptions. We conclude by discussing how the model and associated methodology can be applied in other physical contexts besides glaciology.

5.1 Introduction

Scientists and engineers often study a physical system with the goal of making spatio-temporal predictions (e.g., temperature or glacier thickness) and inferring unknown quantities governing the system (e.g., atmospheric density or ice viscosity). This system's dynamics can often be phrased in terms of spatio-temporal partial differential equations (PDEs) that are based on approximations. The scientist or engineer may also be able to simulate the physical system with a computer simulator, such as a

numerical PDE solver, which is subject to imperfections (e.g., numerical error). Moreover, the scientific constants entering into the system's dynamical equations such as density, friction, or viscosity may not be known precisely, but their range can be constrained to some set of plausible values. Additionally field data, though potentially scarce and noisy, can be incorporated into the analysis.

Such scenarios can be modeled with a variant of a Bayesian hierarchical spatio-temporal model that was introduced in Gopalan et al. (2018) for glacial dynamics, if considered more generally. We delineate three methods to make posterior inference efficient: The first is to utilize bandwidth-limited linear-algebraic routines for likelihood evaluation (Rue, 2001), the second is to utilize an embarrassingly parallel approximation to the likelihood, and the third is to use first-order emulators (Hooten et al., 2011) for speeding up computer simulators. Though our modeling and numerical results are still within a glaciology context, we conclude with a discussion of how the model can be applied to other physical scenarios. Before introducing the Bayesian hierarchical model and associated methodology for computationally efficient posterior inference, it is appropriate to summarize relevant statistical literature developed over the last two decades.

Bayesian hierarchical modeling for geophysical problems was introduced in Berliner (1996) and Wikle et al. (1998), and summarized in Berliner (2003), Cressie and Wikle (2011), and Wikle (2016). In this modeling approach, prior distributions are specified for physical parameters of interest, a physical process is modeled at the intermediary, latent level (conditional on the physical parameters), and the data collection process is modeled conditional on the latent physical process values. Both numerical error and model uncertainty can be incorporated at the process level, while measurement errors can be modeled at the data level. This approach has been applied in a variety of scientific contexts, including the study of ozone concentrations (Berrocal et al., 2014), sediment loads at the Great Barrier Reef (Pagendam et al., 2014), precipitation in Iceland (Sigurdarson and Hrafnkelsson, 2016), Antarctic contributions to sea level rise (Zammit-Mangion et al., 2014), and tropical ocean surface winds (Wikle et al., 2001) (among many others). In Gopalan et al. (2018), the motivating example for the work in this paper, a Bayesian hierarchical model for shallow glaciers based on the shallow ice approximation (SIA) PDE was developed and evaluated.

Kennedy and O'Hagan (2001) suggest constructing Bayesian statistical models that incorporate the output of a computer simulator of a physical process, such as a numerical solver for the underlying system of PDEs. Fundamental to their approach is the inclusion of a specific term that represents the deviation between the output of a computer simulator and the actual process values, known as *model discrepancy* or *model inadequacy*. This framework is developed in Higdon et al. (2004), Higdon et al. (2008), and Brynjarsdóttir and O'Hagan (2014). In particular, Higdon et al. (2008) use a Bayesian model along with a principal components based approach for reducing the computational overhead of running a computer simulation with high dimensional output multiple times (an approach termed as *emulation*). Brynjarsdóttir and O'Hagan (2014) note that the prior for model discrepancy must be chosen carefully to mitigate bias of physical parameters and predictions. In particular, as more prior information is incorporated into a model discrepancy term through a constrained Gaussian process (GP) prior over a space of functions, the less biased inferences and predictions tend to become. The notions of an emulator, a computer simulator, and model discrepancy enter naturally into the aforementioned Bayesian hierarchical framework. Conditional on physical parameters coupled with initial and/or boundary conditions, the physical process values at the latent level can be written as the sum of a computer simulator or emulator term and a model discrepancy term.

To be precise, let us assume that the physical process S can be indexed through time, i.e., as S_j, and S_j is a vector where each element corresponds to a distinct spatial location. One can specify the process level conditional on physical parameter θ as

$$S_j = f(\theta, j) + \delta(j) \tag{1}$$

where $\delta(.)$ is a vector valued model discrepancy function that is independent of θ, and $f(\theta, j)$ is the output of a computer simulation or emulator for physical parameter θ at time index j. If, for instance, at each time point j an observation Y_j of S_j is made with associated measurement error η_j, then observations can be written as

$$Y_j = f(\theta, j) + \delta(j) + \eta_j, \tag{2}$$

which is analogous to Eq. 5 of Kennedy and O'Hagan (2001).

In Kennedy and O'Hagan (2001), $\delta(.)$ is a fixed but unknown function independent of θ that is learned with a GP prior distribution. Similarly,

$\delta(.)$ has a constrained GP prior in Brynjarsdóttir and O'Hagan (2014). The approach in this paper instead assumes a temporally indexed stochastic process (with spatial correlation) that follows a multivariate random walk, rather than a deterministic function. Additionally, in Liu and West (2009), the authors frame a computer emulator of time series run under multiple inputs as a dynamic linear model (DLM). As part of their approach, they allow for time varying auto-regressive coefficients that follow a random walk process, to embed non-stationarity into the model.

While the approach taken in this paper most closely follows the above literature (i.e., Bayesian hierarchical modeling, model discrepancy, and emulation), we briefly review literature in probabilistic numerics and Bayesian numerical analysis; the emphasis in Bayesian numerical analysis is to use probabilistic methods to solve numerical problems, whereas, in the Bayesian hierarchical setup, one is also interested in inference of scientifically relevant parameters and predictions of the physical process. In Conrad et al. (2017), a probabilistic ordinary differential equation (ODE) solver is developed that adds stochasticity at each iteration; conditions for the convergence of this method to the ODE solution are given. Chkrebtii et al. (2016) utilize GPs for solving ODEs; moreover, Calderhead et al. (2008) use a GP regression based method to avoid explicitly solving nonlinear ODEs when performing inference for parameters that provides computational speed ups; additionally, Owhadi and Scovel (2017) present a gamblet-based solver that comes with provably computationally efficient solutions to PDEs. The approach is derived from a game theoretic and stochastic PDE framework.

In the spatio-temporal model described in this paper, stochasticity is induced with an error-correcting process that is separated from the numerical solution. In general, another way to achieve this is to define a stochastic process by equating a PDE to a white noise term – that is, the solution X to a stochastic partial differential equation (SPDE) $L[X] = W$, where L is a differential operator and W is a white noise process (indexed by spatio-temporal coordinates). For instance, a fractional Laplacian operator yields the Matérn covariance function (Whittle, 1954, 1963; Lindgren et al., 2011). We employ the former approach mainly because it is difficult to derive exact covariance functions for arbitrary differential equations (e.g., in the presence of nonlinearities), though we highlight the utility of the latter approach in situations where an analytical covariance function can be derived exactly.

A major feature of this work is to represent the discrepancy between real physical process values and the output of a computer simulator for these physical process values as a multivariate random walk; typically, model discrepancy is endowed with a GP prior or a constrained GP prior over a space of functions as in Kennedy and O'Hagan (2001) and Brynjarsdóttir and O'Hagan (2014). A random walk model is motivated by the notion that the inaccuracy of a computer simulator is most likely going to increase forward in time; conveniently, a random walk's variance grows through time as well. Along with this model is the development of two ways for making computations faster: the first is harnessing first-order emulator inference (Hooten et al., 2011) for speeding up the computation of a numerical solver, and the second is the use of bandwidth-limited numerical linear algebra (Rue, 2001) for computing the likelihood efficiently. Moreover, in the regime of a high signal-to-noise ratio, an embarrassingly parallel approximation to the likelihood can be employed. Finally, methodology to fit a spatial Gaussian field for the log of the scale of numerical errors is discussed.

We must also be clear about what distinguishes this work from its predecessor, Gopalan et al. (2018). This includes the use of emulators, probing higher order random walks besides order 1, derivation of sparsity and computational complexity of log-likelihood evaluation, empirical run time results, and methodology to fit an error-correcting process when little prior information is available. The structure of this paper is as follows: First a test system from glaciology is described. Then the statistical model that is the focus of this work is presented in detail (in the context of the glaciology test case), followed by the exact and approximate likelihood. Then this model is analyzed in terms of computational run time and accuracy of inference, based on the test system from glaciology; moreover, the random walk error-correcting process is assessed with residual analysis. Afterward, we discuss how the model and associated methodology can be applied to other physical scenarios, and conclude by summarizing the model, method, and limitations of the approach.

5.2 Description of a test system from glaciology

Before delving into the specifics of the Bayesian hierarchical model and computational subtleties, we begin with a brief discussion of glaciology. Glaciology is the study of physical systems consisting mostly of ice and snow. This broad definition includes the study of the crystalline nature of ice, the transformation and compaction of snow into ice, the dynamics of the flow of ice and water in a glacier, the relationships between fundamental quantities like viscosity, temperature, and pressure, the relationships between precipitation and meteorology with said ice systems, the interaction of ice systems with other geological systems such as volcanoes and bedrock, and so on. As such, glaciology synthesizes elements from a multitude of scientific disciplines including continuum mechanics, fluid mechanics, hydraulics, chemistry, and meteorology.

Bueler et al. (2005) introduce analytical solutions for the SIA PDE, a commonly used model for the dynamics of glaciers (Fowler and Larson, 1978; Hutter, 1982; Flowers et al., 2005; Cuffey and Paterson, 2010; van der Veen, 2013; Brinkerhoff et al., 2016; Guan et al., 2016; Gopalan et al., 2018). Based on the principle of conservation of mass, the SIA dictates that glacier flow is in the direction of the (negative) gradient of the glacier surface and is due to gravity and basal sliding (also referred to as friction or drag if in the direction of the positive gradient). While an explanation of the SIA PDE is given in Gopalan et al. (2018), our focus is on ice viscosity, B. Intuitively, this parameter controls the softness of the ice. The other main physical parameter, which is not the subject of this paper, is $C_0\gamma$. This controls basal sliding or friction.

For the analysis that follows, we focus on a periodic solution to the SIA in which the thickness of the glacier oscillates through time; $H(r,t)$, the thickness of the glacier as a function of two dimensional space (in polar coordinates) and time, is

$$H(r,t) = H_s(r) + P(r,t), \tag{3}$$

$$P(r,t) = C_p \sin(2\pi t/T_p)\cos^2\left[\frac{\pi(r-0.6L)}{.6L}\right]; \text{if } 0.3L < r < .9L, \tag{4}$$

$$P(r,t) = 0; \text{if } 0 \leq r \leq 0.3L \text{ or if } r \geq 0.9L. \tag{5}$$

In Eq. 3, H_s is a static initial profile of the glacier (i.e., a dome as in

Eq. 21 of Bueler et al. (2005)), P is a perturbation (e.g., precipitation) function, L is the margin length, C_p is the magnitude of the periodic perturbation, and T_p is the period of the perturbation. Bueler et al. (2005) derive a mass balance function that achieves this periodic solution for the SIA PDE. Qualitatively, this test case appears like a dome with a periodic oscillation in thickness around an annulus defined by $0.3L < r < .9L$. In Figure 5.13, an illustration of the oscillations of glacier thickness through time is displayed.

The value of each surface elevation measurement is the value of the exact analytical solution above added to a zero-mean Gaussian random variable with standard deviation of 1 meter, larger than errors of the digital-GPS instruments employed by the UI-IES. We use the same values of parameters as in Bueler et al. (2005) to make for easier comparison to that work and the EISMINT experiment. In particular, $H_0 = 3600$ m, $L = 750$ km, $C_p = 200$ m, and $T_p = 5000$ years.

Employing the same set up as Gopalan et al. (2018), glacial surface elevation measurements are assumed to be collected for 20 years, twice a year, and at 25 fixed spatial locations across the glacier, to emulate how the glaciology team at the University of Iceland Institute of Earth Sciences (UI-IES) collects data at Icelandic glaciers (e.g., see Figure 5.14 illustrating Langjökull and the mass balance measurement sites).

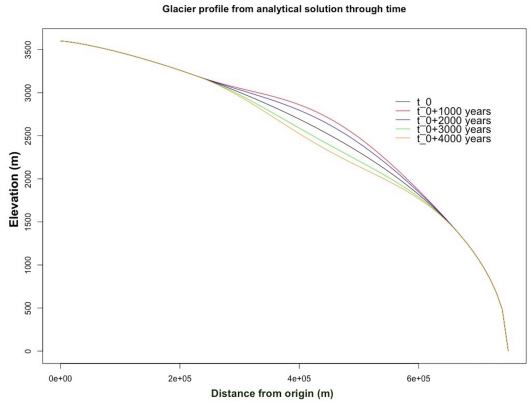

Figure 5.13. An illustration of the periodic oscillatory exact solution to the SIA PDE that is used for the analysis. Since the solution is radially symmetric, only a radial cross section is illustrated. This solution is stationary except for an annulus defined by $0.3L < r < .9L$*, where L is 750 km; in the annulus, the glacier thickness vibrates back and forth periodically, as illustrated.*

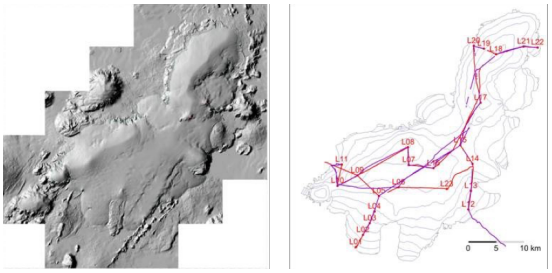

Figure 5.14. A digital elevation map of Langjökull along with measurement sites demarcated on the right, provided by the University of Iceland Institute of Earth Sciences (UI-IES). Langjökull is Iceland's second largest glacier by area, at 900 *sq. km, and its mean thickness is 210 meters above sea level (Björnsson and Pálsson, 2008), so Langjökull is shallow.*

5.3 The hierarchical spatio-temporal model and its properties

Now that we have acquainted the reader with some facts about glaciology and the particular test case used for the analysis in this paper, we next delineate the hierarchical spatio-temporal model that is the focus of this work, by specifying its variables, parameters, and properties, including efficient computation of the likelihood and connections to other modeling frameworks. For the sake of specificity of the presentation, the glaciology example is referenced, similarly to the set up in Gopalan et al. (2018). We assume that n spatial locations are modeled at the latent level, and m of those locations are observed, where m is typically much smaller than n. We use the index j to refer to time indices and i to refer to spatial indices; while space and time are discretized, the differences between successive time and spatial points can be made as small as desired depending on the context of the application and computational resources available. Throughout, we use bolded notation for vectors and uppercase, unbolded, and non-italic notation for matrices. All other mathematical symbols are scalars.

We introduce the Bayesian hierarchical model in the parameter, pro-

cess, data level framework of Berliner (1996). We denote the physical parameters as θ and initial and/or boundary conditions for the physical process as ϕ. At the parameter level, one possibility is to use a truncated normal distribution for θ if the support of the parameter value can be constrained, as was done in Gopalan et al. (2018), where θ represented ice viscosity. However, more generally, the distribution can be specified based on domain knowledge or expertise. We denote the output of a computer simulator, which could be either a numerical solver or an emulator, at time j with the notation $f(\theta,\phi,j)$, which, in full generality, is an element of $\mathbb{R}^n$. While some values could be negative (e.g., temperature), in many cases the computer simulator output can be restricted to the nonnegative real numbers. For a specific example, in Appendix A of Gopalan et al. (2018), $f(\theta,\phi,j)$ is a second-order finite difference solver for glacier thickness, which is constrained to be nonnegative based on a boundary condition. Evidence for a nonnegative support for the physical process, in glaciology, can be found in Gopalan et al. (2018). Particularly, this is evident in Figure 6 of that paper, which shows the process (i.e., glacier thickness) predictions across the glacier, and the distributions are all greater than zero. Specifically, the minimum of the smallest box-plot is more than 750 m. Nonetheless, the reader is suggested to think carefully about whether a negligible amount of probability mass is below zero in different applications (e.g., temperature models).

The process level of the model, conditional on θ and ϕ, can be written as:

$$X_j = X_{j-1} + \varepsilon_j, \tag{6}$$
$$S_j = f(\theta,\phi,j) + X_j, \tag{7}$$

where X_0 is a vector of zeros.

In the above expressions, ε_j is $MVN(0,\Sigma)$ and independent of ε_l for $j \neq l$. Furthermore, X_j, ε_j, $f(\theta,\phi,j)$, and (consequently) S_j are members of $\mathbb{R}^n$. In Gopalan et al. (2018), $\{X_1, X_2, ...,\}$ was referred to as an error-correcting process because it was meant to represent the difference between the numerical solver and the exact solution to the SIA PDE. Note that in Gopalan et al. (2018), S_j referred to glacial thickness at a particular time point, where each component referred to the glacial thickness at a particular grid point. In more generality, the error-correcting statistical process can be a random walk of higher order;

a multivariate RW process of order q ($RW(q)$) is given by:

$$X_j + \sum_{p=1}^{q} (-1)^p \binom{q}{p} X_{j-p} \quad = \quad \varepsilon_j \tag{8}$$

where ε_1,..., ε_q are independent and marginally $MVN(0,\Sigma)$. This form of a higher order random walk is a multivariate extension of the integrated auto-regressive process given in Chapter 5.6 of Madsen (2007). For $q = 2$, this corresponds to RW(2) of Rue and Held (2005).

At the data level, it is assumed that data are regularly sampled at every k-th time point, so that one observes $Y_k, Y_{2k}, ..., Y_{Nk} \in \mathbb{R}^m$; in the glaciology test case, the variables Y referred to glacial surface elevation measurements, and k was set to 5, to represent the fact that the glaciologists take a set of measurements in the summer and winter, or twice a year. The corresponding observation errors $\eta_k, \eta_{2k}, ..., \eta_{Nk}$ are IID $MVN(0, \sigma^2 \mathrm{I})$, and represent GPS measurement errors in the glaciology example. We define the matrix $\mathrm{A} \in \mathbb{R}^{m \times n}$ to be such that its rows are unit basis vectors (i.e., an incidence matrix as in Cressie and Wikle (2011)). That is, $\mathrm{A}_{ab} = 1$ if and only if the bth index of the process level vector S has been observed, and $\mathrm{A}_{ab} = 0$ for all other entries. Then the data level model, conditional on the process S, is

$$Y_{ck} \quad = \quad \mathrm{A}S_{ck} + \eta_{ck}, \tag{9}$$

where we assume that $j \in \{1, 2, ..., T\}$ and $c \in \{1, 2, ..., N\}$, so there are N total observed spatial vectors, observed with a period of length k.

Conditional on θ, ϕ, and a computer simulator, the model can be thought of as a hidden Markov model (HMM) (Baum and Petrie, 1966); the latent physical process evolves according to a RW(1) process added to a numerical solution, and it is observed indirectly with Gaussian noise. It can also be thought of as a conditional general state space model. This is because, conditioning on θ, ϕ, and a computer simulator, one can write:

$$S_j \quad = \quad S_{j-1} + [-f(\theta, \phi, j-1) + f(\theta, \phi, j)] + \varepsilon_j, \tag{10}$$

$$Y_{ck} \quad = \quad \mathrm{A}S_{ck} + \eta_{ck}. \tag{11}$$

Here, the state evolves linearly with a time-dependent offset term: // $[-f(\theta, \phi, j-1) + f(\theta, \phi, j)]$. The notation ck is used in Eq. 11 to indicate that observations of the process are only observed every kth time

point, whereas the latent process evolves at every time step j. The reader who is interested in further understanding the connection between Gaussian processes and state space models may consult Solin and Särkkä (2014).

5.3.1 Exact likelihood

An advantage of using this model is that the likelihood, $p(Y_k, Y_{2k}, ..., Y_{Nk}|\theta, \phi)$, can be computed exactly in an efficient manner. It can also be approximated in a way that leads to embarrassingly parallel computation when the signal-to-noise ratio is high. The next several sections provide more details for these considerations. The likelihood of the model, $L(\theta, \phi) = p(Y_k, Y_{2k}, ..., Y_{Nk}|\theta, \phi)$, has a multivariate normal PDF form:

$$L(\theta, \phi) = \frac{1}{(2\pi)^{(mN)/2}|\Sigma_l|^{1/2}} \exp^{-(Y-\mu_l)^T \Sigma_l^{-1}(Y-\mu_l)/2}, \quad (12)$$

where the mean is:

$$\mu_l = (\mathrm{A}f(\theta, \phi, k), ..., \mathrm{A}f(\theta, \phi, Nk)), \quad (13)$$

and the covariance matrix is:

$$\Sigma_l = \mathrm{U} \otimes \mathrm{V} + \sigma^2 \mathrm{I}, \quad (14)$$

where $\mathrm{U}_{ab} = k\min(a, b)$ with $\mathrm{U} \in \mathbb{R}^{N \times N}$, and $\mathrm{V} = \mathrm{A}\Sigma\mathrm{A}^\intercal$. Also, the symbol $\otimes$ stands for the Kronecker product. Y_{ck} is multivariate normal (conditioning on θ and ϕ) as a direct consequence of Equations 7 and 9, noting that X_{ck} and η_{ck} are independent conditional on θ and ϕ. Moreover, the linearity property of expectations can be used to show that the mean of Y_{ck} is $E[AS_{ck} + \eta_{ck}] = E[AS_{ck}] + E[\eta_{ck}] = E[Af(\theta, \phi, ck) + AX_{ck}] + E[\eta_{ck}] = E[Af(\theta, \phi, ck)] + E[AX_{ck}] + E[\eta_{ck}] = Af(\theta, \phi, ck) + 0 + 0$ (again, conditional on fixed θ and ϕ fixed). Appendix A contains more details of the covariance matrix.

Since evaluating Eq. 12 requires the calculation of the inverse of the matrix Σ_l and its determinant, these must be calculated efficiently (generally this takes $O(N^3 m^3)$ operations, which can grow very quickly with more space and time observations). Since U^{-1} is tridiagonal, the bandwidth of U^{-1} is 1, and the band-limited nature of U^{-1} allows us to compute Σ_l^{-1} and $|\Sigma_l|$ in $O(Nm^3)$ time (Rue, 2001; Golub and Van Loan,

2012). More details for this derivation are given in Appendix A. While using band-limited matrix algebra routines can improve computation, in the next subsection we derive an approximation to the likelihood that is embarrassingly parallel and can therefore accelerate computation even more.

5.3.2 An approximate likelihood

Here we show how to approximate the likelihood in a way that leads to embarrassingly parallel computation. The likelihood $p(Y_k,...,Y_{Nk}|\theta,\phi)$ can be equivalently written as $p(Y_k|\theta,\phi)p(Y_{2k}|Y_k,\theta,\phi)...$ $p(Y_{Nk}|Y_k,..,Y_{(N-1)k},\theta,\phi)$. First note that:

$$Y_k \quad = \quad \mathrm{A}f(\theta,\phi,k)+\eta_k+C_k, \tag{15}$$

where $C_k=\sum_{j=1}^{k}\mathrm{A}\varepsilon_j$. Hence, $p(Y_k|\theta,\phi)$ is multivariate normal with mean $\mathrm{A}f(\theta,\phi,k)$ and covariance matrix $\mathrm{A}(k\Sigma)\mathrm{A}^\intercal+\sigma^2\mathrm{I}$. More generally, we have the relationship:

$$Y_{ck} \quad = \quad Y_{(c-1)k}+\mathrm{A}[f(\theta,\phi,ck)-f(\theta,\phi,(c-1)k)]+\eta_{ck}-\eta_{(c-1)k}+C_{ck},$$

where $C_{ck}=\sum_{j=(c-1)k+1}^{ck}\mathrm{A}\varepsilon_j$. Thus we can approximate $p(Y_{ck}|Y_k,..,Y_{(c-1)k},\theta,\phi)$ as a MVN distribution with mean $Y_{(c-1)k}+\mathrm{A}[f(\theta,\phi,ck)-f(\theta,\phi,(c-1)k)]$ and covariance matrix $\mathrm{A}(k\Sigma)\mathrm{A}^\intercal+2\sigma^2\mathrm{I}$. Nonetheless, to clarify, $p(Y_{ck}|Y_k,..,Y_{(c-1)k},\theta,\phi)$ is not exactly a MVN with mean $Y_{(c-1)k}+\mathrm{A}[f(\theta,\phi,ck)-f(\theta,\phi,(c-1)k)]$ and covariance matrix $\mathrm{A}(k\Sigma)\mathrm{A}^\intercal+2\sigma^2\mathrm{I}$ because $Y_{(c-1)k}$ and $\eta_{(c-1)k}$ are dependent. However, when the magnitude of the observation error $\eta_{(c-1)k}$ is much smaller in comparison to the magnitude of the observation $Y_{(c-1)k}$, and for $Z\sim MVN(0,\sigma^2\mathrm{I})$ with Z independent of $Y_{(c-1)k}$, $Y_{(c-1)k}-\eta_{(c-1)k}\approx Y_{(c-1)k}-Z$.

This approximation is embarrassingly parallel because each of the N terms in the product form of the likelihood $p(Y_k,...,Y_T|\theta,\phi)=p(Y_k|\theta,\phi)$ $p(Y_{2k}|Y_k,\theta,\phi)...p(Y_{Nk}|Y_k,..,Y_{(N-1)k},\theta,\phi)$ (or sum, if computing the log-likelihood) can be evaluated independently of each other. Therefore, in parallel, the computation comes down to evaluating a multivariate normal PDF of dimension m – this can be done in $O(m^3)$.

5.3.3 Computational complexity summary

If no attention is paid to the structure of Σ_l, the cost of evaluating $L(\theta, \phi)$ is limited by the evaluation of Σ_l^{-1} and $|\Sigma_l|$, which generally takes $O(N^3 m^3)$ operations. However, the exact likelihood evaluation can be reduced to $O(Nm^3)$ using band-limited numerical linear algebra. The computational complexity of the approximation is also $O(Nm^3)$ (if no parallelism is used). While an exact likelihood is preferred to an approximation, a benefit of the approximation is that it is embarrassingly parallel – if parallelized, the time complexity is that of evaluating a multivariate normal PDF of dimension m, which is $O(m^3)$. Nonetheless, there also exist parallel versions of sparse Cholesky decomposition, for instance in Gupta and Kumar (1994). Empirical comparisons of the exact and approximate likelihood computations are presented in Section 5.4.5.

5.4 Analysis of the model and associated methodology

The purpose of this section is to motivate the various modeling choices introduced in this paper using the previously described test system from glaciology, both in terms of computational run time and quality of inferences. In particular, we compare an emulator based posterior inference to finite difference method based posterior inference, motivate the use of the random walk error-correcting process with residual analysis, examine the impact of prior information encoded into the error-correcting process on the bias of posterior distributions for physical parameters, and compare the run-time and accuracy of the likelihood approximation versus the exact likelihood. The physical parameter of interest in these examples is ice viscosity, B, whose actual value is the same as Bueler et al. (2005), Payne et al. (2000), and Gopalan et al. (2018): 31.7×10^{-25} in units of $s^{-1}Pa^{-3}$.

Consistent with Gopalan et al. (2018) is the choice of settings for the numerical PDE solver: a 21 by 21 grid (so $n = 441$) is used with $\Delta_x = \Delta_y = 10^5$ m and $\Delta_t = .1$ years. Note that, consequently, the number of simulator runs (25) is much smaller than the dimensionality of the output of the solver (441).

5.4.1 An emulator compared to a numerical PDE solver

In this section, we conduct an empirical study to examine how a first-order spatio-temporal emulator (i.e., an emulator based on the method in Appendix B) compares to a numerical solver of the PDE, both in terms of run-time of computations and posterior inference of ice viscosity. While the precise technical details for constructing a first-order spatio-temporal emulator are given in Appendix B, the idea is to approximate the numerical solver output for each time point that there is collected data. To do this, we train an emulator using the following values for ice viscosity: $\{10, 12.5, 15.0, ..., 70.0\}$ in units of $10^{-25}s^{-1}Pa^{-3}$, a grid of values that is intentionally coarser than the values used for posterior computation, since in this case the emulator must be used for parameter values not in the training set. We used the `rbenchmark` (Kusnierczyk, 2012) package to benchmark the run-time of the log-likelihood of the model evaluated at the actual parameter value computed with a numerical solver versus a first-order spatio-temporal emulator, using a MacBook Pro early 2015 model with a 2.7 GHz Intel Core i5 processor and 8 GB 1867 MHz DDR3 memory. The emulator version performs 14.5 times faster (.354 seconds for the emulator based log-likelihood versus 5.148 seconds for the numerical solver based log-likelihood). We also generated samples from the posterior distribution of ice viscosity with grid sampling (grid [10,70] inclusive with grid width .50 in units of $10^{-25}s^{-1}Pa^{-3}$), using both the numerical PDE version and the emulated version. The summary statistics of 10^6 posterior samples for ice viscosity using both the emulator and numerical solver are given in Table 5.5. Qualitatively, the summary statistics are similar.

The principle behind choosing the ice viscosity parameter values in the training set is to fill the space of the support for ice viscosity, but not to choose a grid as fine as the one used for posterior sampling. (Such an approach would be circular, in that the emulator would just be generating predictions inside of the training set.) However, such a heuristic will not be feasible as the number of parameters grows beyond one parameter (the number of design points would need to grow exponentially in the number of dimensions). In such cases, we suggest using other space-filling designs: notably, a latin hypercube design has been used extensively in the computer experiments literature, for instance in Higdon et al. (2008).

Test Case	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Emulator	15.0	26.5	27.0	27.4	29.0	38.5
Numerical	15.0	24.5	26.5	26.3	28.0	37.5

Table 5.5. Summary statistics of 10^6 *posterior samples of the ice viscosity parameter using an emulator for the SIA and a numerical solver for the SIA; qualitatively, these posterior samples are similar. Units are in* $10^{-25}s^{-1}Pa^{-3}$.

5.4.2 Assessing a random walk

The choice of using a random walk to correct for deviations between the output of a computer simulator and the actual physical process values has a few important motivations:

1. The inaccuracy of a spatio-temporal computer simulation is most likely going to increase as it is run further into the future. Conveniently, a random walk's variance increases with time – for example a RW(1) has marginal variance $j\Sigma$ at time j.

2. As shown in Appendix A and Section 5.3.1, the likelihood involves band-limited matrices, for which there exist specialized numerical linear algebra routines. However, there is a trade-off in bandwidth and the order of the random walk utilized.

3. Spatial correlations in the inaccuracies of a computer simulation can be captured with the covariance matrix Σ.

In addition to these motivations, the purpose of this section is to empirically assess how a random walk model performs for correcting the output of a numerical SIA PDE solver. To do this, we use the analytical SIA solution as a gold standard. This is a simplification in the sense that the real glacial dynamics will not follow the SIA PDE and therefore the analytical SIA solution exactly, but nonetheless this is a way to check the veracity of the random walk error model in some capacity – at the very least, as a model for numerical error but not model uncertainty.

Figure 5.15 displays the differences between the analytical SIA PDE solution for glacial thickness and the numerical SIA PDE solution for glacial thickness at all of the glacier grid points, run forward for 5000

time steps (i.e., 500 years). More precisely, the points in blue are at the margin of the glacier, the points in red are at the interior, and the points filled in black are close to the top (also referred to as the dome) of the glacier. Recall from Figure 5.13 that the glacier looks like a shallow ellipsoid sliced in half (in the x-y plane), and so the panel on the right of this figure is a top view of the glacier grid points, which looks like a circle of radius 750 km projected onto the x-y plane. In comparison, the height is 3600 m.

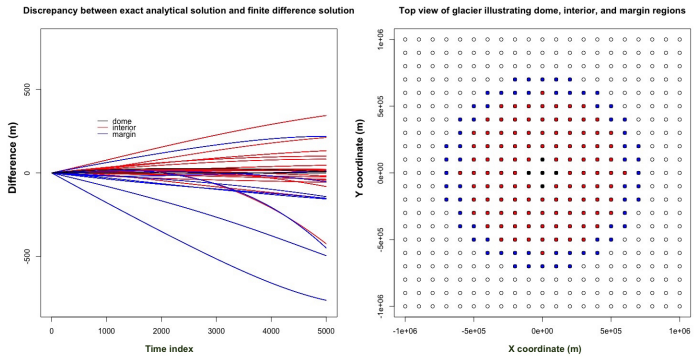

Figure 5.15. An illustration of the difference between the exact analytical solution and the numerical solution for the SIA PDE. On the right panel is a top view of the glacier, whose shape looks like a dome, and therefore the projection on to the x-y plane is a circle. The blue points signify the margin of the glacier (where it drops down to zero thickness), the red points are at the interior of the glacier, and the black points are toward the top of the glacier. The points that are not filled in signify the border of the glacier, where there is no ice thickness. On the left panel the discrepancies between the analytical SIA PDE solution and the numerical SIA PDE solution for all grid points are shown. Specifically, the color of each path corresponds to the grid points on the right panel. Additionally, the paths are shown for 500 years, or 5000 time steps.

The differences are all very smooth (i.e., continuous) functions of time, implying that the numerical SIA PDE solver is producing continuous output as well – we know that the analytical solution is continuous based on the functional form in Eqs. 3-5. Thus, it appears that a random walk of at least a few orders is necessary to represent these differences. Moreover, as expected from Bueler et al. (2005), the largest errors occur at the margin, whereas the interior and dome differences are less extreme.

To assess if a random walk model is appropriate, for each time point

j and for orders 1-7, we computed residuals, in other words, the left hand side of Eq. 8, which should theoretically be distributed like ε_j (i.e., independent $MVN(0,\Sigma)$ random variables). To compute X_j, we take the difference $S_j - f(\theta, j)$, where S_j is the analytical glacial thickness solution to the SIA PDE at time j (i.e., the real physical process for the purpose of this analysis), and $f(\theta, j)$ is the numerical glacial thickness solution to the SIA PDE at time j. We examine the residuals for two randomly selected grid points of the glacier (one at the interior and one at the margin) in Figure 5.16 and 5.17.

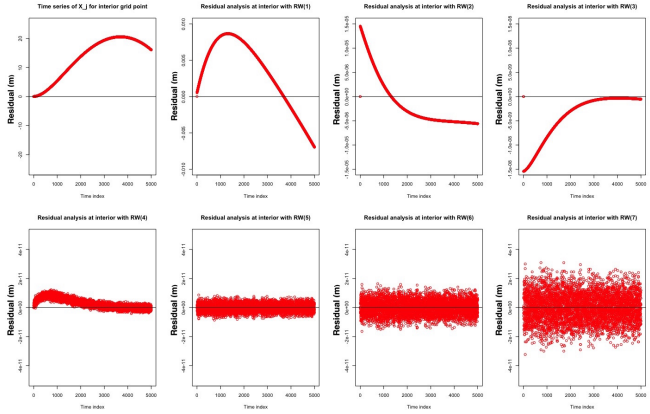

Figure 5.16. This figure displays residuals in units of meters (i.e., the term ε_j in Eq. 8) for RW(q) of orders 1-7 for a randomly selected interior grid point. The first four panels display values on different scaled y-axes to better show the shapes, whereas the bottom four panels have the same scaling for the y-axis to be able to compare across the figures. RW(5) and above look like white noise processes, though RW(5) has the smallest variance.

A few important observations should be emphasized based on the empirical analysis displayed in these figures. The first is that a single order random walk substantially filters the discrepancy; for the interior grid point, it is reduced from the order of 10 m to the order of .01 m (1000 times reduction in magnitude), and for the margin grid point from the order of 100 m to .05 m (more than 1000 times reduction). Additionally, for both the interior and margin grid points, it appears that RW(5) is optimal in the sense that the residuals closely resemble a white noise process and have the smallest variance. While the residuals from higher order RW processes also resemble white noise, the magnitude of the noise is larger. Nonetheless, we believe that real physical processes will not

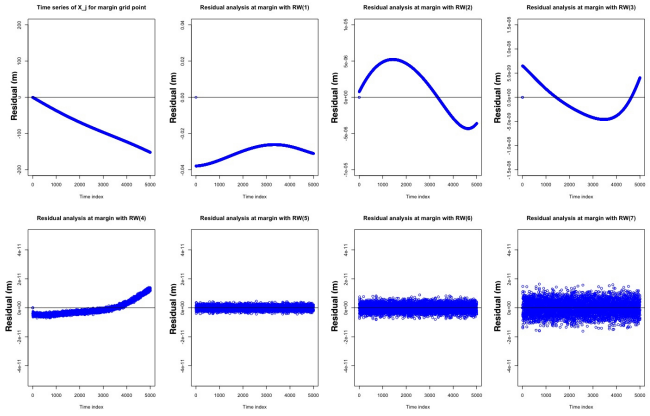

Figure 5.17. This figure also displays residuals in units of meters (i.e., the term ε_j in Eq. 8) for RW(q) of orders 1-7 for a randomly selected margin grid point. Just like the previous figure, the first four panels display values on different scaled y-axes to better show the shapes, whereas the bottom four panels have the same scaling for the y-axis to be able to compare across the figures. Just as in the previous figure, RW(5) and above look like white noise processes, though RW(5) has the smallest variance.

always be as smooth as the analytical SIA PDE solutions, and hence it is likely that a lower order RW process will be preferred for real scenarios.

5.4.3 Reducing bias for the posterior distribution of θ

In Brynjarsdóttir and O'Hagan (2014), when prior information about the model discrepancy term is introduced in a simple physical system (i.e., a constrained GP over a space of functions), the bias of a posterior distribution of a physically relevant parameter reduces. We have found that a very similar phenomenon occurs in the glaciology test case, a result that was pointed out in Gopalan et al. (2018). Specifically, in Bueler et al. (2005), it is shown that there is large spatial variation in the scale of deviations between the exact solution to the SIA and a numerical finite difference solver of the SIA. Specifically, there is spatial variation between the dome, interior, and margin of a glacier, with deviations at the margin being markedly larger than at the interior and dome. To investigate the effect of such prior information, we choose the matrix Σ to be such that it is block diagonal with 3 blocks, Σ_{int}, Σ_{dome}, and Σ_{margin}. Each of these blocks is derived from a square exponential covariance kernel

with the same length-scale parameter $\phi = 70$ km, but differing variance parameters σ^2_{int}, σ^2_{dome}, and σ^2_{margin}. If we ignore prior information from Bueler et al. (2005), we assume that there is an equal prior probability that each of σ^2_{int}, σ^2_{dome}, and σ^2_{margin} is in the set $\{.1, 1, 10, 100\}$ in units of m^2. If we use prior information from Bueler et al. (2005), we instead assume equal prior probability on $\{.1, 1\}$ for σ^2_{int}, $\{1, 10\}$ for σ^2_{dome}, and $\{10, 100\}$ for σ^2_{margin} (again all units are m^2). As shown in Gopalan et al. (2018), the posterior for ice viscosity is less biased in the case that incorporates prior information for the scale of errors; this phenomenon is explored again in the next section.

While in the above discussion we have not been precise about the term bias, the following ought to make this notion more rigorous. Let θ_0 be the true parameter, and $\hat{\theta}$ be an estimator of θ_0. The frequentist definition of bias is usually $E[\hat{\theta} - \theta_0]$, where the expectation (i.e., average) is taken over the sampling distribution, $p(Y|\theta_0)$. The Bayesian notion of bias used informally in the preceding paragraph (and essentially the same notion as in Brynjarsdóttir and O'Hagan (2014)) is $b(Y, \theta_0) = E[\theta - \theta_0]$, where the expectation (i.e., average) is taken with respect to the posterior distribution of θ, $p(\theta|Y)$. Consider $E[b(Y, \theta_0)]$, where the (outer) expectation is taken with respect to the sampling distribution. Then $E[b(Y, \theta_0)] = E[E[\theta - \theta_0]] = E[E[\theta] - \theta_0] = E[\hat{\theta} - \theta_0]$, which is the frequentist bias. In other words, the frequentist bias is equivalent to the average of $b(Y, \theta_0)$ over the sampling distribution, if the posterior mean is chosen as an estimator. In the glaciology test case, we have (informally) not noticed much variability in the posterior for ice viscosity over repeated sampling of the data, and hence the distinction between Bayesian bias and frequentist bias is not significant.

The reader may wonder why a fixed θ_0 was assumed in the preceding paragraph, despite that a Bayesian model has been presented in this paper. In fact, it is typical to assume that the actual value of a parameter is fixed, despite ascribing a probability distribution to it in the form of a prior or posterior. Conceptually, such a probability distribution is a representation of a modeler's uncertainty regarding the fixed, unknown value of the parameter. For more on this interpretation of Bayesian statistics, the reader can consult results of statistical decision theory (e.g., on admissibility) in Lehmann and Casella (2003) and Robert (2007). This viewpoint is also taken in Bayesian asymptotic analysis, such as the Bernstein-von Mises theorem (van der Vaart, 2000; Shen and Wasserman,

2001).

5.4.4 Inferring Σ

The covariance matrix Σ, first introduced after Equation 7 in Section 5.3, determines the spatial correlation inherent in the error-correcting process, X. Since spatial correlation in the error-correcting process is important to model (which is particularly evident in the glaciology example of Bueler et al. (2005)), we need to discuss how Σ ought to be specified. Choosing Σ can be difficult if no or little prior information is available, and in such a case, we suggest:

$$\Sigma \quad = \quad \text{diag}(v)\text{R diag}(v),$$

where $\log(v) \sim MVN(\mu_v, \Sigma_v)$, Σ_v is derived from a GP kernel such as squared exponential or Matérn kernel, and R is a correlation matrix also derived from a GP kernel. To avoid non-identifiability and complexity of inference, it is suggested to pre-specify the parameters of these GP kernels. This approach is similar to the modeling strategy employed in Geirsson et al. (2015). The intuition behind this approach is that the term v encodes spatial variability in the scale of deviations between the output of a computer simulator and the true physical process, and spatial correlation in these deviations is strongly enforced with non-diagonal terms in both Σ_v and R.

Figure 5.18 illustrates a map of the mean posterior field for the variances of the error-correcting process, where the area of each circle is proportional to the inferred posterior mean of variance; due to a multivariate normal prior on $\log(v)$, elliptical slice sampling is used as the method for posterior sampling (Murray et al., 2010). Consistent with Bueler et al. (2005), the variances tend to increase at the margins and are smaller at the interior. Additionally, the scaled differences between the analytical solution and numerical solver at the final time point the simulator is run (where scaling is inverse of the posterior mean of standard deviation) should theoretically approach a mean zero normal distribution according to the model. The p-value for an Anderson-Darling test is .436, suggesting that the scaled differences between the analytical solution and numerical solver are consistent with a normal distribution. Moreover, the sample mean for these scaled differences is .079 and the sample standard deviation is .409.

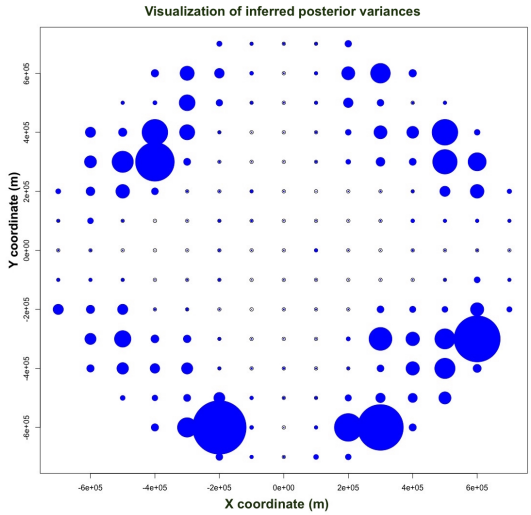

Figure 5.18. Inferred posterior variance field of the error-correcting process, where the area of each circle is proportional to the variance at the grid point centered at the circle. Qualitatively, this field behaves as one would expect from the the work of Bueler et al. (2005), where the authors demonstrate that numerical inaccuracies for the SIA PDE are greatest toward the margin, but much smaller at the interior of the glacier.

As is discussed in the previous subsection, prior information for Σ has an effect on the inference of physical parameters (i.e., ice viscosity), and in particular, a lack of prior information can lead to a very biased posterior distribution for physical parameters. To compare the fitted Σ using a GP field against the Σ matrices discussed in the previous section, we show in 2.4.4 a comparison of posterior inference for the ice viscosity parameter for three choices of Σ. The first choice of Σ is the posterior mean of samples assuming the structure $\Sigma = \text{diag}(v)\text{Rdiag}(v)$, with $\log(v) \sim MVN(\mu_v, \Sigma_v)$. In the second and third scenarios, Σ is block diagonal with three variance parameters for each of the three blocks. A weakly informative case assumes that $\sigma^2_{int} = \sigma^2_{dome} = \sigma^2_{margin} = .1$, whereas a more informative case (using prior information from Bueler et al. (2005)) has $\sigma^2_{int} = \sigma^2_{dome} = .1$ and $\sigma^2_{margin} = 10$ (all units are m^2). The scenario for weak prior information for Σ results in a very biased posterior distribution whose support does not cover the actual parameter value (31.7×10^{-25} in units of $s^{-1}Pa^{-3}$) – the maximum in this case is 26.5×10^{-25} in units of $s^{-1}Pa^{-3}$. While the (absolute) biases of the

posterior for ice viscosity for the GP field version compared to the prior information from Bueler et al. (2005) are comparable (5.09 versus 4.01 in units of $10^{-25}s^{-1}Pa^{-3}$), the posterior variance is markedly larger in the former case. This result suggests that prior knowledge from a domain expert is likely to be useful in determining Σ, though in a case when that does not exist, the methodology described in this section is an adequate alternative.

Test Case	Min	1st Quartile	Median	Mean	3rd Quartile	Max
GP Σ	10.0	21.0	36.0	35.7	50.5	70.0
Strong prior	18.0	25.0	26.5	26.6	28.0	35.5
Weak prior	12.5	18.5	19.5	19.5	20.5	26.5

Table 5.6. Summary statistics of 10^6 *posterior samples of the ice viscosity parameter under three versions of* Σ*. While the weakly-informative case leads to a very biased posterior, the biases for the ice viscosity posterior in the first two* Σ *matrices are comparable. Nonetheless, the posterior variance is much less in the case with prior information from Bueler et al. (2005).*

5.4.5 Exact versus approximate likelihood

In Section 5.3.1, we showed an exact way to calculate the model likelihood as well as an approximation in Section 5.3.2. In this subsection, our purpose is to compare these two methods of likelihood computation in terms of run-time and posterior inference. Using a MacBook Pro early 2015 model with a 2.7 GHz Intel Core i5 processor and 8 GB 1867 MHz DDR3 memory (as before), one component of the log-likelihood approximation (which can be computed in an embarrassingly parallel fashion with the other components of the sum) takes .0179 s, whereas the full log-likelihood calculation, as in Section 5.3.1, is .354 seconds (in both cases, using a first-order emulator). The results of comparing posterior samples for the ice viscosity parameter are given in Table 5.7 – thus, while the mean, median, first, and third quartiles are comparable, the approximate version has larger posterior uncertainty than the exact version as is evidenced by the wider tails. These results suggest that, while there is likely a computational speed-up afforded by using the approximation (i.e., at least an order of magnitude), the price to pay is

increased posterior uncertainty.

Test Case	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Exact	15.0	26.5	27.0	27.4	29.0	39.5
Approx.	10.0	26.0	28.0	27.7	31.0	52.5

Table 5.7. Summary statistics of 10^6 *posterior samples of the ice viscosity parameter using an exact likelihood and a likelihood approximation (units are in* $10^{-25}s^{-1}Pa^{-3}$*). While the 1st quartile, median, mean, and 3rd quartile are similar, the tails in the approximation are much wider.*

5.5 Generality of the model and methodology

Though we have tested the model and methodology in the previous sections in the context of a glaciology example, it should be noted that they can be used in other physical systems with similar components. In essence, this modeling and methodology can be applied in scenarios where:

1. A computer program (i.e., *computer simulator*) is available to simulate a continuous physical process through space and time, but there is a deviation between the output of the computer simulator and the actual physical process.

2. The deviations between the computer simulator output and the actual physical process values tend to grow with time and exhibit spatial correlation structure.

3. Measurements of the physical process are available, but they are potentially scarce both in space and time.

4. Physical parameters governing the physical process are uncertain but can be constrained with domain knowledge for the random walk error covariance (i.e., Σ).

Recall that at the process level, the model stipulates that:

$$S_j = f(\theta, \phi, j) + X_j. \tag{16}$$

To apply the same setup to another physical scenario, a different version of $f(.,.,.)$, such as a numerical PDE solver for another system of spatio-temporal PDEs besides the SIA, can be used. However, while $f(.,.,.)$ will need to be tailored to another physical scenario based on a different numerical scheme or physical model, the X_j term would be modeled in the same way (i.e., with a random walk).

5.6 Conclusion

The objective of this work has been to set forth a versatile physical-statistical model in the Bayesian hierarchical framework that incorporates a computer simulator for a physical process, such as a numerical solver for a system of PDEs. Posterior inference for physical parameters (and, consequently, posterior predictions of the physical process) can be computationally demanding within this model, since each evaluation of the likelihood requires a full PDE solve and computing the inverse and determinant of a large covariance matrix. Therefore, we have set forth two main ways to speed up computation: first is the use of bandwidth-limited linear algebra in a manner similar to Rue (2001) for quickly handling the covariance matrix in the likelihood, and the second is the use of spatio-temporal emulation in a manner similar to Hooten et al. (2011) to emulate a PDE solver that is expensive to evaluate. An additional method for speeding up computation is to approximate the likelihood in a way that leads to embarrassingly parallel computation. The utility of this model and corresponding inference methodology is demonstrated with a test example from glaciology.

A unique feature of this work is how we represent the discrepancy between a computer simulator for a physical process and the real physical process values. One approach, as in Kennedy and O'Hagan (2001) and Brynjarsdóttir and O'Hagan (2014), is to assume that this is a fixed yet unknown function that can be learned with a GP (or constrained GP) prior distribution over a space of functions. Instead, we assume that this discrepancy is a spatio-temporal stochastic process (i.e., a random walk), which is motivated by the fact that a computer simulation is likely to become less accurate as it is run further forward in time, as well as exhibit some degree of spatial correlation in inaccuracies. An interesting consequence of this modeling decision is that linear algebraic routines

for band-limited matrices can be utilized for evaluating the likelihood of the model in an efficient manner. Another interesting artifact of this approach is that when prior information is used for the random walk's error term (i.e., in Σ), the bias for the posterior distribution of θ is reduced. The same phenomenon is exhibited in the work of Brynjarsdóttir and O'Hagan (2014), where a constrained GP prior over a space of functions ends up reducing the bias of the physical parameter posterior distribution.

Despite that the model and methodology appear to perform well in the analysis of this paper, it is important to comment on some potential drawbacks of the approach, particularly when applied to other physical contexts. In this paper, emulation works adequately with a single parameter, though emulators do not always work well in other applications or higher dimensional parameter spaces. For example, Salter et al. (2019) document some shortcomings of a principal components based emulator in climate modeling. The second main computational advantages stem from log-likelihood evaluation speed-ups. The use of bandwidth-limited matrix algebra for the exact log-likelihood can be used so long as the model holds, which may not always be the case (e.g., with a non-Gaussian data distribution). Additionally, the log-likelihood approximation holds when the measurement errors are small relative to the signal modeled, which depends on the measurement instruments used to collect the data. For instance, on common geophysical scales of thousands of meters, light detection and ranging (LIDAR) or digital-GPS data have maximum errors on the order of a meter.

Additionally, if it is not possible to program the computer simulator to produce output at the data measurement locations, there are essentially two main ways to handle such a scenario. The first is to use spatial kriging to predict the value of the computer simulator at the spatial locations where data are collected, given the output of the computer simulator at the grid points. A simpler approach is to use inverse-distance weighting of the simulator output at the nearest neighbors; that is, take a weighted average of the four nearest grid points of the simulator, where the weights are proportional to the inverse of distance. Such an approach, for example, has been used in Geirsson et al. (2015).

Future research will include predicting Langjökull glacier surface elevation using the modeling and methodology within this paper, based on actual data collected by the UI-IES.

5.7 Appendix A: the exact likelihood and computation

As was shown in Appendix B of Gopalan et al. (2018), the covariance matrix for the observed data can be written as $\mathrm{U} \otimes \mathrm{V} + \sigma^2 \mathrm{I}$, where $\mathrm{U}_{ab} = k \min(a,b)$ with $\mathrm{U} \in \mathbb{R}^{N \times N}$, and $\mathrm{V} = \mathrm{A}(\Sigma)\mathrm{A}^\intercal$. It can be verified that U^{-1} is tridiagonal, so it has bandwidth one – more specifically:

$$\mathrm{U}^{-1} = k^{-1} \begin{bmatrix} 2 & -1 & 0 & \dots & & \\ -1 & 2 & -1 & 0 & \dots & \\ 0 & -1 & 2 & -1 & 0 & \dots \\ 0 & 0 & \ddots & \ddots & \ddots & \\ 0 & \dots & & -1 & 2 & -1 \\ 0 & \dots & & & -1 & 1 \end{bmatrix}.$$

One useful property of the Kronecker product is that $(\mathrm{U} \otimes \mathrm{V})^{-1} = \mathrm{U}^{-1} \otimes \mathrm{V}^{-1}$. Therefore:

$$\begin{aligned} (\mathrm{U} \otimes \mathrm{V})^{-1} &= \mathrm{U}^{-1} \otimes \mathrm{V}^{-1} \\ &= k^{-1} \begin{bmatrix} 2\mathrm{V}^{-1} & -\mathrm{V}^{-1} & 0 & \dots & & \\ -\mathrm{V}^{-1} & 2\mathrm{V}^{-1} & -\mathrm{V}^{-1} & 0 & \dots & \\ 0 & -\mathrm{V}^{-1} & 2V^{-1} & -\mathrm{V}^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \ddots & \ddots & \\ 0 & \dots & & -\mathrm{V}^{-1} & 2\mathrm{V}^{-1} & -\mathrm{V}^{-1} \\ 0 & \dots & & & \mathrm{V}^{-1} & \mathrm{V}^{-1} \end{bmatrix}, \end{aligned}$$

whose bandwidth is $O(m)$.

Let us denote $\mathrm{U} \otimes \mathrm{V}$ as W. By the matrix inversion lemma, it follows that $(\sigma^2 \mathrm{I} + \mathrm{W})^{-1} = \sigma^{-2}\mathrm{I} - \sigma^{-2}(\mathrm{W}^{-1} + \sigma^{-2}\mathrm{I})^{-1}\mathrm{I}\sigma^{-2}$. The matrix $\mathrm{W}^{-1} + \sigma^{-2}\mathrm{I}$ has bandwidth $O(m)$ since W^{-1} has bandwidth $O(m)$ as shown previously, so this expression can be computed in $O(Nm^3)$ (Rue, 2001; Golub and Van Loan, 2012).

Similarly, by the matrix determinant lemma, $\log[\det(\sigma^2 \mathrm{I} + \mathrm{W})]$ is $\log[\det(\mathrm{I} + \sigma^2 \mathrm{W}^{-1})\det(\mathrm{W}^{-1})^{-1}] = \log[\det(\mathrm{I} + \sigma^2 \mathrm{W}^{-1})]$-$\log[\det(\mathrm{W}^{-1})]$. Since both terms are log-determinants of square matrices of dimension Nm and bandwidth $O(m)$, this can be calculated in $O(Nm^3)$ due to the efficient Cholesky factorization of band-limited matrices (Rue, 2001; Golub and Van Loan, 2012).

5.8 Appendix B: first-order spatio-temporal emulators

In the examples of this paper, the function $f(.,.,.)$ (i.e., the computer simulator) can take one of two forms: a numerical PDE solver for the SIA, or an emulator constructed from the numerical PDE solver for the SIA. The numerical method for solving the SIA PDE is as given in Gopalan et al. (2018), and the emulator is constructed based on the finite difference solver in a manner as suggested in Hooten et al. (2011), termed first-order emulation.

That is, we start with a set of plausible values for ice viscosity: $\{\theta_1, \theta_2, ..., \theta_p\}$ and, for each time point there is collected data ck, we store a matrix M_{ck}, where the q-th column of matrix M_{ck} is the output of the numerical solver using parameter value θ_q after running for ck time steps forward. Thus, each matrix M_{ck} is of dimension n by p, and without essential loss of generality we can assume that the number n is much larger than p, and each matrix M_{ck} is of rank p.

For each matrix, M_{ck}, we compute a singular value decomposition (SVD), $\mathrm{U}_{ck}\mathrm{D}_{ck}\mathrm{V}_{ck}^{\mathsf{T}}$. The goal is to find a (vector valued) function $v_{ck}(\theta*)$ such that the emulated output at time ck for parameter value $\theta*$ is $\mathrm{U}_{ck}\mathrm{D}_{ck}v_{ck}(\theta*)$. To find the q-th element of v_{ck}, we train a random forest (Breiman, 2001; Liaw and Wiener, 2002) with $(\theta_1, (V_{ck}^{\mathsf{T}})_{q1})$, $(\theta_2, (V_{ck}^{\mathsf{T}})_{q2}), ..., (\theta_p, (V_{ck}^{\mathsf{T}})_{qp})$ as training data, where $(V_{ck}^{\mathsf{T}})_{q1}$ is the first element of the q-th right singular vector, $(V_{ck}^{\mathsf{T}})_{q2}$ is the second element of the q-th right singular vector, and so on. Not all of the right singular vectors need be used in emulation, and a heuristic such as an elbow-scree plot or the randomization procedure of Friedman et al. (2001) can be used to determine the number of right singular vectors to keep. However, if the number of simulator runs (p) is much smaller than the dimensionality of the output (n), all of the right singular vectors can be utilized with computational savings, as is done in the experiments of this paper.

We have assumed the initial conditions and boundary conditions are known, since this is the case in the glaciology problems we have studied, where the boundary condition is that glacial thickness is nonnegative, and the initial glacier profile (i.e., a dome) is known. In general, however, ϕ may be incorporated into the analysis above by considering θ and ϕ jointly. Additionally, a variant is to directly emulate the likelihood

function. However, since there is flexibility in the choice of Σ (which enters into the likelihood), unless one is set on using a particular value of Σ, it is sensible to emulate the numerical solver as opposed to retraining a likelihood emulator for each potential choice of Σ.

6 Paper 3

Spatio-temporal statistical modeling of Langjökull

Gopalan, G., Hrafnkelsson, B., Aðalgeirsdóttir, G., and Pálsson, F.: Spatio-temporal statistical modeling of Langjökull, to be submitted to the Annals of Applied Statistics, 2019.

Abstract: The Bayesian hierarchical model and methodologies of Papers 1 and 2 are applied to Langjökull, a main glacier of Iceland. The data consist of a 100 m resolution digital elevation map, a 100 m resolution map of bedrock topography, and mass balance and surface elevation measurements at 22-25 sites collected twice a year, from 1997 to 2015. Mass balance predictions are attained with a spatio-temporal statistical model implemented in R-INLA. In contrast to the simulation studies of Paper 1, the posterior for ice viscosity concentrates around a single value (5.05×10^{-24}, $s^{-1}Pa^{-3}$), a value that is within the expected range of values in the glaciology literature. Additionally, the prediction intervals capture 20 of 22 surface elevation measurements during spring 2015, having been trained on surface elevation data from 1997-2014.

6.1 Introduction

About 10 percent of Iceland's area is covered by glaciers (Björnsson and Pálsson, 2008), and the glaciers of Iceland provide a natural laboratory to study the ramifications of a changing climate. The subject of this paper is Langjökull, which is Iceland's second largest glacier by area (and third by volume). Langjökull is 900 km^2 in area, 190 km^3 in volume, 210 m mean thickness, with a max thickness of about 650 m.

To provide context for the analysis contained in this paper, the predecessors of this work are summarized in the following paragraphs. Paper 1 first introduces a prototypical Bayesian hierarchical model (BHM) involving a computer simulator of glacial dynamics (a novel finite difference solver to the shallow ice approximation partial differential equation, or

SIA PDE). Subsequently, Paper 1 evaluates the BHM with simulation studies involving analytical solutions to the SIA PDE in idealized test scenarios. A conclusion of this work is that posterior inference for ice viscosity and predictions for glacial surface elevation can be biased, though this bias can be reduced by incorporating prior information into a model discrepancy term (what we called an error-correcting process for reasons discussed in Paper 2).

Paper 2 focuses on computational improvements to the model introduced in Paper 1, particularly by using bandwidth-limited matrix algebra and first-order emulation for speeding up the computer simulator of glacial dynamics. An approximation to the log-likelihood is also discussed. Moreover, higher-order random walk models are assessed for representing model discrepancy, again in a test scenario with an analytical solution to the SIA PDE. A random walk is a natural choice, since the uncertainty of a computer simulator is likely to increase as it is run forward in time. A conclusion of Paper 2 is that first-order emulation and bandwidth-limited linear algebra can significantly improve computation, and a random walk error-correcting process is a good representation of model discrepancy.

As the final part in this sequence of papers, Paper 3 applies the Bayesian hierarchical model and associated methods from the previous two papers to real data from Langjökull, as opposed to simulations based on analytical PDE solutions. As in Paper 1, the primary inferential goals are to calculate the posterior distribution of ice viscosity, as well as to make probabilistic forecasts of glacier surface elevation. A key scientific quantity necessary for achieving these goals is mass balance (MB) – this is the rate of change of mass (through melting or the accumulation of snow) at the glacier surface with respect to time. Unlike Paper 1 and Paper 2, in Paper 3 MB is not available at every grid point along the glacier, but at 22-25 fixed sites. Therefore, a subgoal is to make MB predictions for the entirety of the glacier, using the MB measurements at these 22-25 sites.

This paper is structured as follows. First, we delineate the Langjökull data sources available from the UI-IES: 1) 100 meter resolution initial surface elevation conditions at 1997 derived differential GPS measurements, 2) GPS surface elevation measurements at 22-25 measurement locations for 1997-2015 measured twice a year, 3) summer and winter MB measurements at the measurement sites, 4) 100 meter resolution

bedrock measurements with radio-echo sounding instruments. Second, we describe MB predictions derived from R-INLA, briefly reviewing its methodology and the use of penalized-complexity (PC) priors. Third, we recapitulate the BHM from Papers 1 and 2, and follow the same methodology to produce a posterior distribution for ice viscosity and probabilistic predictions for ice viscosity. Finally, we conclude by discussing some future directions of work, including the application of the BHM to infer bedrock topography, and incorporating uncertainties for MB predictions from R-INLA.

6.2 Data

As previously discussed, MB (i.e., rate of change of mass with respect to time) is a critical quantity that is necessary for glacial-dynamical equations, such as the form of the SIA PDE that is stated in Section 4.2.1. For the purposes of the current work, MB is needed for all spatial points on the 100 m resolution grid and all years between 1997-2015, as input to the PDE solver we use in the Bayesian hierarchical model of Papers 1 and 2.

In glaciology, MB is a pivotal quantity defined as the rate of change of mass at the surface at a particular glacier location. This change in mass is essentially attributable to precipitation, snow drift, and the melting of ice. Since MB varies both temporally and over spatial regions of the glacier, this is a function of spatio-temporal coordinates, i.e., x, y, and t. As is delineated in more detail in Paper 1, the dynamics of glaciers are essentially due to deformation caused by gravity, interaction of the ice with bedrock, and MB.

The glaciology team at the University of Iceland Institute of Earth Sciences (UI-IES) has measured MB at the Icelandic glaciers twice a year annually, and for Langjökull since 1997. To measure summer balance, a measurement is taken in October to early November (i.e, change in mass over the summer, usually due to snow and ice melt), and to measure winter MB (i.e., change in mass over the winter time, usually due to snow accumulation), a measurement is taken between late April and mid-May. Measurements are accomplished using a stratigraphic method, by measuring the depth to the previous summer's layer by drilling ice cores (in the case of winter MB). The net MB for the year is the sum of the

winter and summer MB values, in units of meters (water equivalent) per year. Water equivalent means that the value is calculated using the density of water, so that measurements across different years are comparable on the same unit. The measurements are taken at 22-25 fixed measurement sites, which are chosen to cover the range of elevation of Langjökull and lie along major flow lines of the glacier. While in Pálsson et al. (2012) a manual interpolation (which is informed with spatial kriging) method is used to produce MB predictions over other portions of the glacier not measured, in this work we will use a Bayesian spatial statistical model with the R-INLA software, which is discussed in more detail in the next section.

In addition to MB measurements, the UI-IES has collected surface elevation data. In particular, during 1997, the UI-IES surveyed the glacier surface using a differential GPS methodology along lines of length approximately 1 km (see Figure 3 of Pálsson et al. (2012)), with an estimated 1.06 m standard deviation in errors. The resultant data are of resolution 100 m. In addition to this high resolution data, the UI-IES collects digital-GPS measurements of surface elevation at the MB measurement sites twice a year.

Simultaneously, during the 1997 campaign to map surface elevation at Langjökull, the UI-IES conducted radio-echo sounding to determine bedrock elevations – the glacier rests upon bedrock with a widely varying topography. The radio-echo sounder was designed by members of the University of Iceland Science Institute (Sverrisson, Marteinn and Jóhannesson, Ævar and Björnsson, Helgi, 1980). It consists of transmitting and receiving dipole antennas that are 30 m long: the transmitter issues a .2 microsecond pulse at a frequency of 1 kHz, and the receiver passes the reflected signal through a band-pass filter of 2-5 MhZ. More details and method are given by Björnsson (1986). The result is a map of bedrock of surface elevation, also of 100 m resolution on a grid that matches the surface elevation data.

To aid the reader in understanding these data (which form the basis of subsequent analysis), we show below a series of figures. The first figure is a map of Langjökull surface elevation at 1997, along with points specifying MB and surface elevation measurement sites. The second figure displays box-plots of the annual MB measurements across 23 measurement sites, between 1997 and 2015. The third figure is a map of Langjökull bedrock topography collected with radio-echo sounding by

the UI-IES glaciology team.

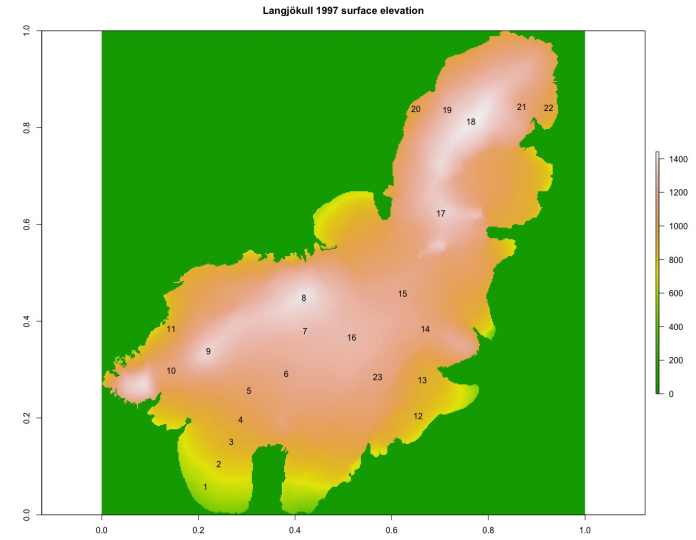

Figure 6.19. Surface elevation of Langjökull at 1997. Though the axes are scaled to be from 0-1, the length of each axis is 46.4 km. The units of elevation are in meters.

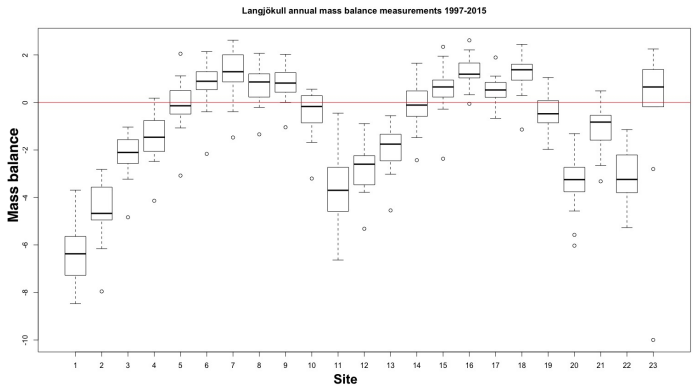

Figure 6.20. Range of annual MB measurements for each of the 23 sites in the previous figure.

The Langjökull surface elevation at 1997 is of a 100 m resolution, with minimum 436.3 m above sea level (ASL) and maximum 1442 m ASL. The grid that contains the surface elevation data and bedrock measurements is of dimensions 43800 m by 46400 m. Langjökull is about 50 km and 20 km wide, which explains the name “long glacier".

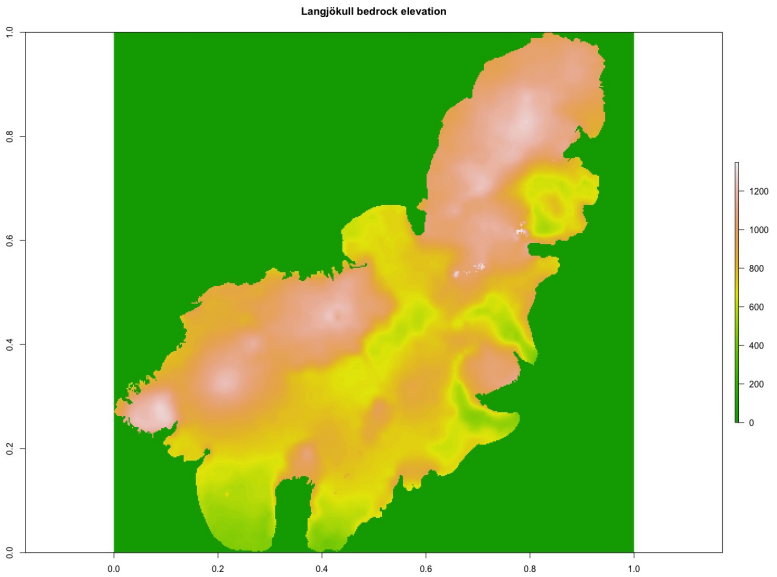

Figure 6.21. Langjökull bedrock elevation in meters. The bedrock elevation was obtained by means of a radio-echo sounding device developed at the University of Iceland.

6.3 Mass balance predictions

MB (i.e., rate of change of mass with respect to time) is a critical quantity that is necessary for glacial-dynamical equations, such as the SIA PDE as is written in Section 4.2.1. For the purposes of the current work, it is needed for all spatial points on the 100 m resolution grid and all years between 1997-2015, as input to the PDE solver we use in the Bayesian hierarchical model of Papers 1 and 2. Pálsson et al. (2012) use a manual interpolation method (assisted with spatial kriging) to make MB predictions away from the sites at which MB measurements are taken (the sites that are depicted in Figure 6.19). Instead, we use a spatial statistical model to make MB predictions on the 100 meter resolution grid described in the previous section, for the years 1997-2015. In particular, we use a linear model with latitude, longitude, and elevation as fixed effects, and a spatially varying random effect that follows a Matérn covariance kernel.

The use of a linear model for MB is motivated by the glaciology literature. For instance, Aðalgeirsdóttir et al. (2006) model precipitation (which is directly linked to MB) with a linear relationship in x-coordinate, y-coordinate, and z-coordinate (i.e., elevation) (equation 4). Additionally, Figure 6 of Pálsson et al. (2012) shows a nearly linear relationship be-

tween MB and elevation, with the exception of higher elevations. One possible reason for this finding for winter balance is that higher elevations tend to exhibit snow drift.

For a fixed year, site i, the winter MB is:

$$MB_{wi} = \beta_{0w} + \beta_{1w}x_i + \beta_{2w}y_i + \beta_{3w}z_i + U_w(s_i) + \varepsilon_w(s_i). \quad (17)$$

Similarly, for a fixed year and at site i, the summer MB is:

$$MB_{si} = \beta_{0s} + \beta_{1s}x_i + \beta_{2s}y_i + \beta_{3s}z_i + U_s(s_i) + \varepsilon_s(s_i). \quad (18)$$

Note that s_i is the spatial coordinate of the i-th location, and it consists of x_i, y_i, and z_i, which are the longitude (or x-coordinate), latitude (or y-coordinate), and elevation coordinates of the i-th location, respectively. Moreover, β_{0s} and β_{0w} are summer and winter intercepts for (and likewise for the remaining β parameters). ε_s and ε_w are unstructured zero mean normal error terms. The processes U_s and U_w have mean 0, and the covariance is determined by the Matérn covariance function:

$$C(s_a, s_b) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{8\nu}||s_a - s_b||/\rho) K_\nu (\sqrt{8\nu}||s_a - s_b||/\rho).$$

Here, σ is the marginal standard deviation, ρ is the spatial range parameter, ν is the smoothness parameter, K_ν is the modified Bessel function of second kind with order ν, and Γ is the gamma function (Bakka et al., 2018). A process with Matérn covariance is the solution to the following stochastic partial differential equation (SPDE):

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau U(s)) = W(s)$$

where $s \in \mathbb{R}^d$, $\tau > 0$, Δ is the Laplacian, $\kappa > 0$, $\alpha > d/2$, and W is a Gaussian white noise process.

As shown in more detail in Lindgren et al. (2011) the approach of Lindgren et al. (2011) approximates the solution to this SPDE as a linear combination of basis functions, which are derived from a finite element method (FEM). For instance, the finite element mesh for our analysis is displayed in Figure 6.22. Using the FEM approach results in a sparse precision matrix for the coefficients in the linear combination, and sparsity can be used to make computations more efficient. The computational

benefits of sparsity when working with Gaussian Markov random fields (GMRFs) are described in Lindgren et al. (2011) and further used in Paper 2, and essentially rely upon bandwidth-limited matrix algebra routines like those found in Golub and Van Loan (2012).

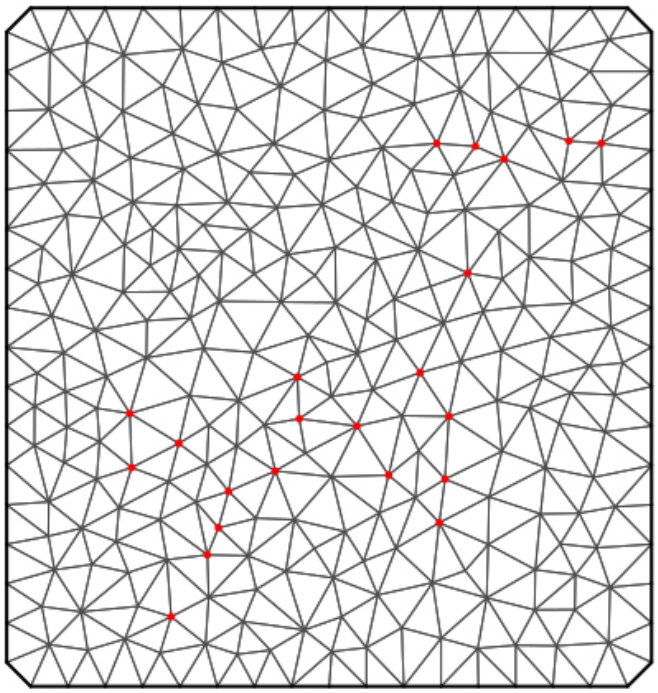

Figure 6.22. The R-INLA mesh we used for the SPDE model, with red points indicating MB measurement sites. We tried slightly finer and coarser meshes, but found negligible difference in MB predictions.

The R-INLA software (Rue et al., 2009) uses a fully Bayesian approach. The scale and range parameters have penalized-complexity (PC) priors (Simpson et al., 2017). These priors are based on parameterizations in terms of the KL-divergence to a base-model; the prior is essentially an exponential distribution in this parameterization where the exponential distribution hyper-parameter is selected based on controlling a tail-area

probability (i.e., user-defined scaling). In R-INLA, the PC-priors can be instantiated with the inla.spde2.pcmatern function, where ρ_0 is set to the grid-width of 100, and the tail area probability p_ρ is set to .01 (a fairly small value). Based on Pálsson et al. (2012), σ_0 is set to 30 cm with p_σ set to .05. To clarify, the p_ρ and ρ_0 values specify the left tail-area probability for the prior on the spatial range parameter, ρ (i.e., $P(\rho < \rho_0) = p_\rho$). Similarly, p_σ and σ_0 specify the right tail-area for the prior on σ, (i.e., $P(\sigma > \sigma_0) = p_\sigma$). This is in accordance with the user-defined scaling principle for PC priors, described more in Simpson et al. (2017).

By default, priors for the fixed effect parameters are normally distributed with mean 0 and precision .001, and the intercept has a normal prior with mean 0 and precision 0 (i.e., flat). We began by fitting the models with the default settings, but noticed a large variability in the posteriors on a year to year basis. For example, see Figure 6.23, which shows the year to year variability in the posterior for the latitude fixed effect parameter, for the summer MB model.

Moreover, regularizing linear model coefficients tends to be good for prediction (e.g., ridge regression (Hoerl and Kennard, 1970)). Therefore, we decided to change the fixed effect parameter variances and make them smaller than the 1000 default setting. To help motivate the choice of precisions for each of the fixed effect parameters (i.e., elevation, latitude, and longitude), we have fit several linear models for both winter and summer MB (i.e., with the `lm` function in `R`) using all of the years worth of data together, as opposed to separate models for each year:

1. $MB \sim Long + Lat + Elevation$,
2. $MB \sim Year + Long + Lat + Elevation$,
3. $MB \sim Year * Long + Lat + Elevation$,
4. $MB \sim Long + Year * Lat + Elevation$,
5. $MB \sim Long + Lat + Year * Elevation$,
6. $MB \sim Year * Long + Year * Lat + Elevation$,
7. $MB \sim Year * Long + Lat + Year * Elevation$,
8. $MB \sim Long + Year * Lat + Year * Elevation$,

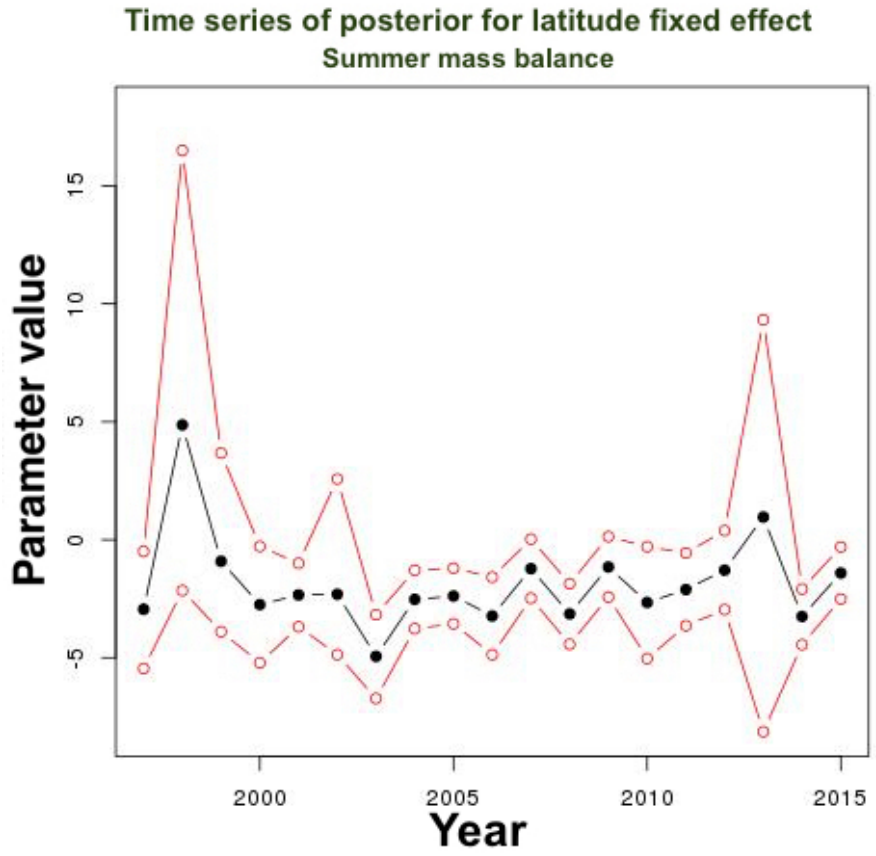

Figure 6.23. Posterior for the latitude fixed effect parameter, through time, when the default prior setting in R-INLA is used. The red points are the 2.5th and 97.5th percentiles, whereas the black points are the medians. This is for summer MB.

9. $MB \sim Year * Long + Year * Lat + Year * Elevation.$

The notation used above is essentially syntax used for linear modeling in R; i.e., the * refers to an interaction between variables. The fit of each model is compared with the resultant Akaike information criterion (AIC) (Akaike, 1974), and the results are displayed below.

Model Number	Summer Mass Balance AIC	Winter Mass Balance AIC
1	1127.6260	725.7081
2	914.6192	532.9997
3	921.8139	540.4028
4	925.0915	550.2093
5	**907.4420**	524.8316
6	936.5179	551.0461
7	915.9444	524.0919
8	909.0320	524.5416
9	909.6316	**523.7783**

Table 6.8. A comparison of linear model fits using AIC. The purpose of this analysis is to determine prior precisions for fixed effect parameters.

The idea behind this approach is to see if an interaction between year and each of latitude, longitude, or elevation is needed. If there is such an interaction, then the prior variance for the corresponding fixed effect parameter should be chosen to be larger than the prior variances for the other fixed effect parameters i.e., as to induce greater yearly variation. For the winter MB, the best fitting model in terms of AIC was Year*Long + Year*Lat + Year*Elevation, implying equal prior variances for all the fixed effect parameters. However, the best fitting model (again, in terms of AIC) for the summer MB data was Long + Lat + Year*Elevation, implying smaller prior variances for long and lat than elevation. Based on this result, we used a precision of .1 for elevation and 1 for both longitude and latitude, as to allow for greater prior variance in elevation than longitude and latitude for summer MB. We evaluated the efficacy of this hyper-parameter setting by performing leave-one-site-out cross validation on a randomly selected year, 2007. This combination of parameters produced the smallest root mean square error (RMSE) of 0.53 m/year. In comparison the RMSE for all precisions set to .001 (default), .1, 1, and 1000 were 0.54, 0.54, 0.59, and 0.61 m/year, respectively. For illustrative purposes, Table 6.9 yields the leave-one-site-out cross validation during 2007, summer MB.

Site	Actual MB	MB Prediction	Squared Residual	SD
1	-8.02	-6.01	4.04	0.17
2	-5.52	-5.71	0.04	0.19
3	-3.98	-4.28	0.09	0.13
4	-3.38	-3.67	0.09	0.12
5	-2.87	-2.91	0.002	0.11
6	-2.35	-2.24	0.01	0.12
7	-1.46	-1.56	0.01	0.13
8	-1.29	-0.71	0.33	0.16
9	-1.22	-1.03	0.04	0.21
10	-2.38	-2.83	0.21	0.17
11	-4.87	-4.46	0.17	0.22
12	-4.29	-4.46	0.03	0.19
13	-3.72	-4.24	0.28	0.17
14	-2.14	-2.31	0.03	0.13
15	-1.91	-2.38	0.22	0.10
16	-1.60	-1.78	0.04	0.12
17	-1.45	-1.31	0.02	0.13
18	-1.19	-0.79	0.16	0.17
19	-2.04	-1.69	0.12	0.17
20	-3.61	-3.84	0.05	0.22
21	-2.36	-2.58	0.049	0.19
22	-4.25	-3.75	0.25	0.24
23	-2.73	-2.21	0.27	0.14
			RMSE = .53	

Table 6.9. 2007 summer MB leave-one-site-out prediction results. MB units are meters (water equivalent) per year. The resultant RMSE is 0.53 m per year.

Additionally, the posterior (2.5 percentile, 50 percentile, and 97.5 percentile) for standard deviation for 2007 summer MB was (0.011, 0.082, 0.429), all in units of meters per year (water equivalent). Likewise, the posterior (2.5 percentile, 50 percentile, and 97.5 percentile) for the range parameter for 2007 summer MB was (91.09, 506.2, 3313), in units of meters. For perspective, 506.2 m is about 5 grid widths long.

Using the models in Equations (17) and (18) implemented in R-INLA,

we made predictions of summer and winter MB for all of the years from 1997 to 2015, inclusive. To aid in the speed of computation, we set up R Studio Server on an Amazon Web Services (AWS) instance (m4.2xlarge), which has 8 virtual CPUs. For illustration, we show the output of summer MB, winter MB, and net MB for the years 1997 and 2015 in Figures 6.24-6.27. Additionally, we tested to see how the number of MB measurement sites affects the accuracy of MB predictions. To do so, we randomly subsampled 5 of 25 sites (a fifth), 12 of 25 sites (about a half), and 20 of 25 sites (four fifths) for a training set, and predicted MB on the left out sites (using RMSE as the measure of predictive accuracy). Box-plots illustrating the RMSE of each of the random subsamples, for both winter and summer MB, are illustrated in Figure 6.28. The median of RMSEs decreases as more sites are subsampled, though much more so from 5 to 12 samples than 12 to 20 samples. Also, while the minimum RMSE for both summer and winter MB is achieved with 20 training sites, there is a larger variance in RMSE for 20 training sites, partially because of a smaller number of left out sites.

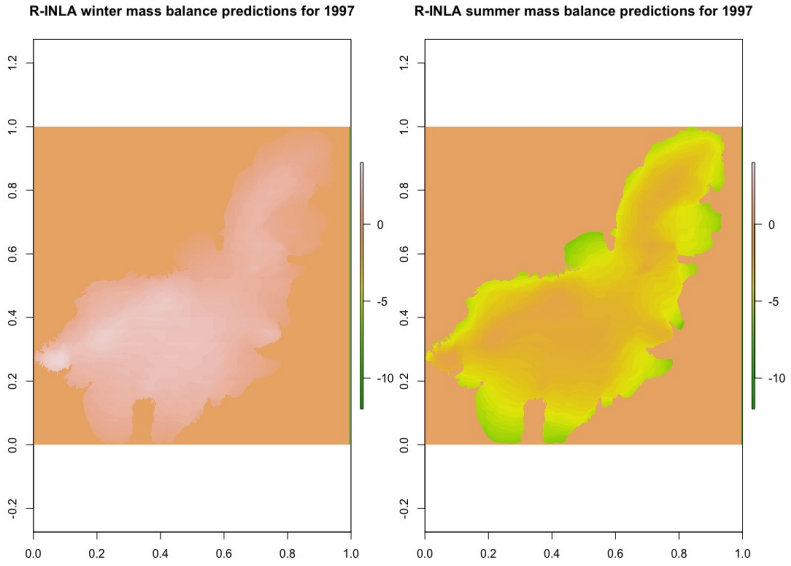

Figure 6.24. On the left are the winter MB predictions across Langjökull during 1997. On the right are summer MB predictions across Langjökull during 1997. As to be expected, MB values tend to be negative during the summer due to melting.

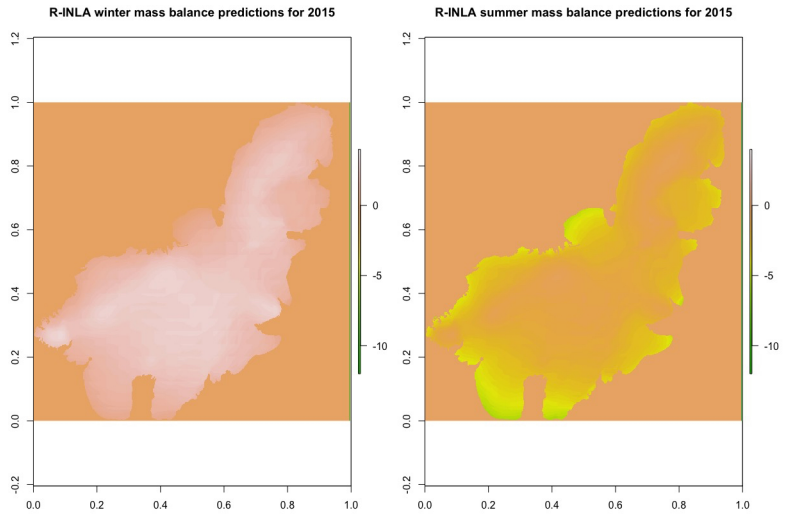

Figure 6.25. On the left are the winter MB predictions across Langjökull during 2015. On the right are summer MB predictions across Langjökull during 2015. MB values tend to be negative during the summer due to melting, though less melting is observed than during 1997.

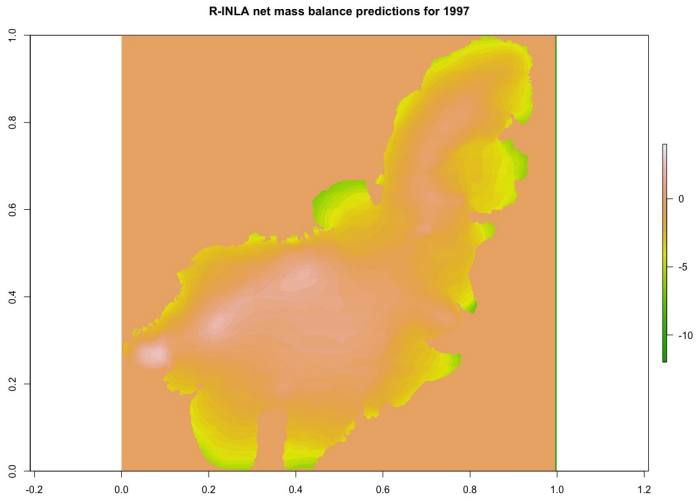

Figure 6.26. Net MB predictions (sum of summer and winter) during 1997.

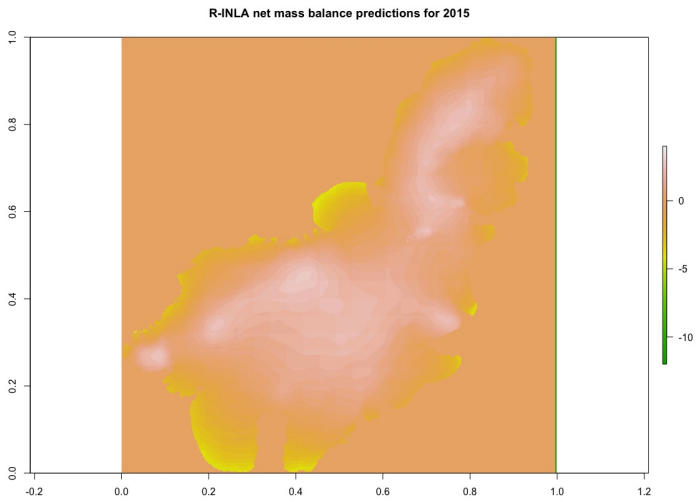

Figure 6.27. Net MB predictions (sum of summer and winter) during 2015, which appear to be less negative (i.e., less melt) than 1997.

6.4 Results of inference and prediction

For the final part of the analysis of Langjökull data, we used the Bayesian hierarchical model from Paper 1 and Paper 2 to infer ice viscosity and make surface elevation predictions at the 22 sites for which there are measurements for all years from 1997 to 2015. Some important details in implementing the model and methodology from the previous papers are addressed in the following paragraphs.

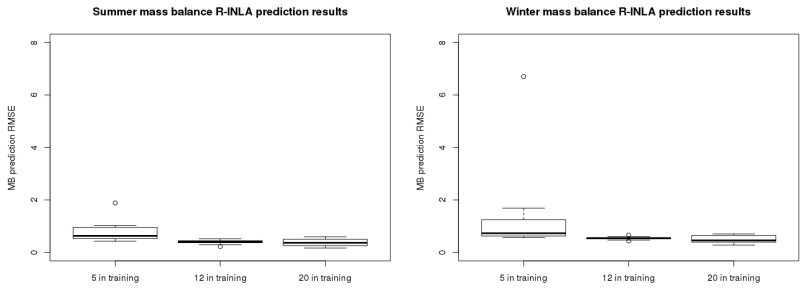

Figure 6.28. Box-plots of MB predictions for random training sets of size 5, 12, and 20. Generally, RMSE decreases as more sites are sampled, though the reduction appears larger going from 5 to 12 sites as opposed to 12 to 20 sites.

We used years 1997 to 2014 for fitting the Σ matrix, the covariance matrix for the random walk error term that is defined in Section 4.2.2 of Paper 1, Section 5.3 of Paper 2, and is discussed more in Section 5.4.4 of Paper 2. The year 2015 was used to test the prediction of surface elevation, while model fitting (i.e., training) used only data from 1997-2014. Hence, this was a leave-last-year-out scenario. The surface elevation measurements were assumed to be taken a year apart (though in reality there may be a week or so of variation), starting at the end of April of each year from 1997 to 2015, inclusive. We considered only the first 22 sites since these had no data gaps for all of the years between 1997 to 2015.

The computer simulator used was based upon the numerical solver for the SIA PDE from Paper 1, where Δ_x and Δ_y were taken to be the grid width of the digital elevation map, 100 m. Additionally, Δ_t was taken to be .1 years, as in the previous papers. The initial surface elevation input into the computer simulator was the digital elevation map of 100 m resolution, measured in late April of 1997 as discussed in Section 6.2. We also input the MB predictions generated from R-INLA, discussed in the previous section. The MB at a particular location was assumed to be constant throughout the year, and is equivalent to the sum of the predicted summer and winter MB values across the 100 m resolution grid. Because the initial conditions are at the end of April 1997, the current year's summer MB and the next year's winter MB are added together in determining the net MB. Additionally, we needed to account for the fact that the site locations are not exactly aligned to a grid. Since the grid

was of a high resolution, we simply used a weighted average of the four nearest grid points, where the weighting is proportional to the inverse of squared distance; this method is also implemented in Geirsson et al. (2015).

As described in more detail in Paper 2, a crucial component in the model is the matrix Σ of the error-correcting process. To fit the matrix Σ we used the same methodology from Section 5.4.4 of Paper 2. Additionally, to run the numerical SIA PDE solver from Paper 1, we needed to start with initial estimates of the viscosity and basal sliding parameters. The initial ice viscosity parameter was from Minchew et al. (2015), $2.4 \times 10^{-24} s^{-1} Pa^{-3}$. Using this value for B, and the horizontal velocities as mapped in Minchew et al. (2015), and the horizontal velocity equation (Equation 24) from Flowers et al. (2005), an estimate of $-9.51 \times 10^{-12} Pa^{-1} ms^{-1}$ for $C_0\gamma$ (basal sliding parameter) was used.

The prior mode for the truncated-normal prior for ice viscosity was set to be $2.4 \times 10^{-24} s^{-1} Pa^{-3}$, to be consistent with Minchew et al. (2015). Otherwise, the same parameter values from Paper 1 were used, and the posterior was computed in the same manner as in Paper 1. The resultant estimate for ice viscosity was $5.05 \times 10^{-24} s^{-1} Pa^{-3}$ – nearly all of the posterior mass concentrated on this value. This is in contrast to the simulation studies of Paper 1, in which there was much more posterior uncertainty. Nonetheless, the value of $5.05 \times 10^{-24} s^{-1} Pa^{-3}$ is within the range from Cuffey and Paterson (2010). Using the methodology of Appendix B of Paper 1, we then made posterior predictions of surface elevation at the 22 sites in 2015. The resultant predictions are shown below, along with marginal standard deviation values based on the error-correcting process variance. As can be seen, 20 of 22 of the prediction intervals contain the observed surface elevation value, yielding a coverage of 90.9 percent. Moreover, the average standard deviation of the predictions is 5.74 m, which is small in comparison to distances on a glacial scale, though comparable to the measurement errors of 1 m.

Site	Observed Elevation	Predicted Elevation	Standard Deviation
1	588.7	592.1	7.09
2	742.1	743.3	10.6
3	878.9	878.9	5.86
4	984.3	984.6	4.74
5	1102.6	1105.9	4.53
6	1197.0	1195.6	3.74
7	1282.8	1282.5	3.97
8	1406.9	1410.1	5.74
9	1376.1	1375.1	5.55
10	1127.7	1130.3	4.98
11	992.1	992.7	3.41
12	826.6	828.8	7.04
13	880.2	880.5	7.69
14	1152.4	1152.8	5.59
15	1182.5	1185.9	5.78
16	1260.6	1259.9	6.03
17	1358.1	1344.3	3.59
18	1434.7	1427.5	3.19
19	1278.1	1282.7	5.90
20	1091.5	1096.4	9.90
21	1148.5	1151.6	6.43
22	951.3	957.0	4.88

Table 6.10. Predictions and associated standard deviations for surface elevation at Langjökull for late April, 2015, based on training the Bayesian hierarchical model on data from 1997 to 2014. The observed elevations are included, for comparison.

Site	Lower Bound	Upper Bound	Coverage
1	577.9	606.3	1
2	722.1	764.4	1
3	867.2	890.6	1
4	975.1	994.1	1
5	1096.9	1115.0	1
6	1188.1	1203.1	1
7	1274.6	1290.5	1
8	1398.6	1421.6	1
9	1363.9	1386.2	1
10	1120.4	1140.3	1
11	985.8	999.5	1
12	814.7	842.9	1
13	865.1	895.9	1
14	1141.6	1163.9	1
15	1174.4	1197.5	1
16	1247.8	1271.9	1
17	1337.1	1351.5	0
18	1421.2	1433.9	0
19	1270.9	1294.5	1
20	1076.6	1116.2	1
21	1138.8	1164.5	1
22	947.2	966.7	1

Table 6.11. Prediction intervals for surface elevations at Langjökull for late April, 2015, based on training the Bayesian hierarchical model on data from 1997 to 2014. The lower bound is formed by taking the prediction and subtracting 2 standard deviations (see previous table). Likewise the upper bound is formed by taking the prediction and adding 2 standard deviations. 20 of 22 sites cover, and the interval lengths are on the order of 10 meters.

6.5 Future directions and conclusion

There are a few ways to extend this work. The first is to use the Bayesian hierarchical model to infer the spatially-varying basal sliding field, which was assumed to be constant in the previous section. Another possibility is to use the model to infer bedrock topography. These are more complex problems that require inferring a parameter for every grid point – 110,704 in total. Additionally, we could apply the same model and methodology to other glaciers or ice sheets, in Iceland and in other parts of the world, such as Greenland. Additionally, we have not used the standard deviation estimates, returned by R-INLA, for MB predictions, but rather only the posterior predictive means. The standard deviations returned by R-INLA could be used for a prior over MB values.

In conclusion, this paper has applied the Bayesian hierarchical model and methodology from Paper 1 and Paper 2 to Langjökull MB and surface elevation data that was collected by the UI-IES. MB predictions across a 100 m grid were derived using a linear model with elevation, x-coordinate, and y-coordinate as fixed effects, and a random spatially varying effect with spatial correlation determined by a Matérn covariance kernel; the model was fit with the R-INLA software. The resultant posterior for ice viscosity concentrates sharply around the value of 5.05×10^{-24}, units of $s^{-1}Pa^{-3}$, a plausible value that is in the range given by Cuffey and Paterson (2010). Nonetheless, the sharp concentration for the posterior of B is an unexpected result, based on the simulation studies of Paper 1. Moreover, 20 of 22 prediction intervals capture the observed surface elevation values at 22 sites at 2015; the predictions are generated using surface elevation data from 1997-2014 as well as the MB predictions derived from R-INLA. As a complement to the simulation studies of Paper 1 that give predictions for 100 years, these prediction results suggest utility for yearly forecasting of surface elevation for glacier elevation measurement sites; nonetheless, it must be noted that the quality of these surface elevation predictions relies upon accurate MB predictions.

7 R code

This section includes the main pieces of R code used during the work of this thesis. All of the code was written in the R programming language, version 3.1. While this code has been tested and used throughout the thesis, it is possible that some of it can be made faster (e.g., by replacing for loops). Additionally, some of the code has been split to take two lines instead of one, as to be able to fit onto a page without running off the right margin.

7.1 R packages

The packages we had used are as follows:

- mvtnorm_1.0-8 (Genz et al., 2019)
- randomForest_4.6-12 (Liaw and Wiener, 2002)
- raster_2.8-4 (Hijmans, 2018)
- INLA_18.07.12 (Håvard et al., 2009)

Additionally, base R was used as well.

7.2 Physical constants

```
#PART 1: physical constants
############################
# x-width
del_x <- 10^5
#y-width
del_y <- 10^5
```

```
#t step
del_t <- .1

#max range to go out to
RANGE <- 10^6

#flat bedrock
R <- matrix(rep(0,M*N),ncol=N)
#physical constants
C_0 <- 0
gamma <- 0
rho <- 910
g <- 9.80665
g_const <- 9.80665
n <- 3
B <- 10^(-16)/(31556926)
mu_max <- 2.5*10^(-11)
C_1 <- (31556926)*(2*B)*((rho*g)^n)/(n+2)
R_0 <- 750*10^3
H_0 <- 3600
lambda <- 0
f <- 0
t_0 <- 5000
N <- dim(H_mat_full)[2]
H <- H_mat_full
R <- matrix(rep(0,M*N),nrow=M)
```

7.3 Analytical solutions

These analytical solutions are used for the simulations of Paper 1 and Paper 2.

7.3.1 Bueler et al. (2005) test B

```
ALPHA <- (2-(n+1)*lambda)/(5*n+3)
BETA <- (1+(2*n+1)*lambda)/(5*n+3)

t_0 <- BETA/C_1*((2*n+1)/((1-f)*(n+1)))^n*
```

```
R_0^(n+1)/(H_0^(2*n+1))
#height function in polar
H_true<- function(r,t)
{
  H_0*(t/t_0)^(-ALPHA)*
  (1-((t/t_0)^(-BETA)*r/R_0)^((n+1)/n))^(n/(2*n+1))
}
```

7.3.2 Bueler et al. (2005) test C

```
lambda <- 5
ALPHA <- (2-(n+1)*lambda)/(5*n+3)
BETA <- (1+(2*n+1)*lambda)/(5*n+3)
t_0 <- BETA/C_1*((2*n+1)/((1-f)*(n+1)))^n*
R_0^(n+1)/(H_0^(2*n+1))
#height function in polar
H_true<- function(r,t)
{
  H_0*(t/t_0)^(-ALPHA)*
  (1-((t/t_0)^(-BETA)*r/R_0)^((n+1)/n))^(n/(2*n+1))
}
MB <- function(r,t)
{
  return(5*H_true(r,t)/t)
}
```

7.3.3 Bueler et al. (2005) test D

```
t_0 <- 5000
R_0 <- 30*10^3
f <- 0
L <- 750*10^3
T_p <- 5000
C_p <- 200
C <- (1-f)^n*C_1*H_0^(2*n+2)/(2*(1-1/n)*L)^n

X <- function(r)
{
  return(4*r/(3*L)-1/3+(1-r/L)^(4/3)-(r/L)^(4/3))
}
```

```
X_deriv <- function(r)
{
  -4*((r/L)^(1/3)+(1-r/L)^(1/3)-1)/(3*L)
}
g_deriv <- function(r)
{
  -pi*sin(pi*(r-.6*L)/(.3*L))/(.6*L)
}
g_d_deriv <- function(r)
{
  return(-pi^2*cos(pi*(r-.6*L)/(.3*L))/(.18*L^2))
}
X_d_deriv <- function(r)
{
  -4*((r/L)^(-2/3)-(1-r/L)^(-2/3))/(9*L^2)
}
H_ss <- function(r)
{
  3*H_0*(-5*X(r)^(-13/8)/8*(X_deriv(r))^2+X(r)^
  (-5/8)*X_d_deriv(r))/(8*(2/3)^(3/8))
}
H_p_d_deriv <- function(r,t)
{
  H_ss(r)+C_p*sin(2*pi*t/T_p)*g_d_deriv(r)
}
H_s_deriv <- function(r)
{
  3*H_0*X(r)^(-5/8)*X_deriv(r)/(8*(2/3)^(3/8))
}
H_p_deriv <- function(r,t)
{
  return(H_s_deriv(r)+C_p*sin(2*pi*t/T_p)*
  g_deriv(r))
}
g <- function(r)
{
  return(cos(pi*(r-.6*L)*(.6*L)^(-1))^2)
}
```

```
M_s <- function(r)
{
  s <- r/L
  return(C*(s^(1/n)+(1-s)^(1/n)-1)^(n-1)*
  (2*s^(1/n)+(1-s)^(1/n -1)*(1-2*s)-1)/(L*s))
}
H_s <- function(r)
{
  s <- r/L
  return(H_0/(1-1/n)^(n/(2*n+2))*
  ((1+1/n)*s-1/n+(1-s)^(1+1/n)-s^(1+1/n))^
  (n/(2*n+2)))
}
P <- function(r,t)
{
  val <- 0
  if(r < .9*L && r > .3*L)
  {
    val <- C_p*sin(2*pi*t/T_p)*
    cos(pi*(r-0.6*L)/(0.6*L))^2
  }
  return(val)
}
H_p <- function(r,t)
{
  return(H_s(r)+P(r,t))
}
M_c <- function(r,t)
{
  val <- 0
  if(r < .9*L && r > .3*L)
  {
  T_1 <- 2*pi*C_p*g(r)*cos(2*pi*t/T_p)/T_p
  T_2 <- -M_s(r)
  T_3 <- -C_1*H_p(r,t)^4*H_p_deriv(r,t)^2*
  (H_p(r,t)*H_p_deriv(r,t)/r+5*H_p_deriv(r,t)^2+
  3*H_p(r,t)*H_p_d_deriv(r,t))
  val <- T_1+T_2+T_3
```

```
  }
  return(val)
}
#######
H_true <- function(r,t)
{
  return(H_s(r)+P(r,t))
}
H_init <- function(r)
{
  return(H_s(r)+P(r,t_0))
}
MB <- function(r,t)
{
  return(M_s(r)+M_c(r,t))
}
```

7.3.4 Bueler et al. (2005) test E

```
R_0 <- 30*10^3
H_0 <- 1000
gamma <- (31556926)*2*B*(rho*g)^n/(n+2)
t_0 <- 5000
n <- 3
f <- 0
L <- 750*10^3
M_0 <- .3
r_1 <- 200*10^3
r_2 <- 700*10^3
theta_1 <- 10
theta_2 <- 80

C_v <- (2^(n-1)*M_0/gamma)^(1/(2*n+2))
w <- function(r)
{
  L^(1+1/n)-r^(1+1/n)
}
H_v<- function(r)
{
```

```
  C_v*w(r)^(n/(2*n+2))
}
H_v_deriv <- function(r)
{
  -.5*C_v*r^(1/n)*w(r)^((-n-2)/(2*n+2))
}
mu <- function(r,theta)
{
  if(r_1 <r && r < r_2 &&
   theta_1 < theta && theta < theta_2)
  {
    return(mu_max*4*(r-r_1)*(r_2-r)*4*
    (theta-theta_1)*(theta_2-theta)/
    ((r_2-r_1)^2*(theta_2-theta_1)^2))
  }
  return(0)
}
mu_deriv <- function(r,theta)
{
  if(r_1 <r && r < r_2 &&
   theta_1 < theta && theta < theta_2)
  {
    return(mu_max*4*(theta-theta_1)*(theta_2-theta)*
    (theta_2-theta_1)^(-2)*
    4*(r_1+r_2-2*r)*(r_2-r_1)^(-2))
  }
  return(0)
}
H_v_d_deriv <- function(r)
{
  -C_v*w(r)^((-3*n-4)/(2*n+2))*
  (r^((1-n)/n)*w(r)+(n+2)*r^(2/n)/2)/(2*n)
}
M_b <- function(r,theta)
{
  return(-rho*g*(H_v(r)^2*H_v_deriv(r)*
  (r^(-1)*mu(r,theta)+mu_deriv(r,theta))+
  mu(r,theta)*H_v(r)*(2*H_v_deriv(r)^2+
```

```
  H_v(r)*H_v_d_deriv(r))))
}
MB <- function(r,theta)
{
  M_0+M_b(r,theta)
}
b <- matrix(rep(0,dim(D)))
H_true <- function(r,t)
{
  return(H_v(r))
}
```

7.4 Log-likelihood evaluation (with approximation)

```
log_llh_B_approx <- function(B_prop)
{
  log_llh_score <- 0
  H_cur <- matrix(u_0,nrow=M)
  for(index in 1:Time)
  {
    H_cur <- matrix(forward_compute(
    H_cur,H_cur+R,R,B_prop,0,t_0+del_t*
    (index-1)),nrow=M)
    H_vec <- as.vector(H_cur)
    if(index %in% time_samples)
    {
      ind <- which(index == time_samples)
      if(index == 5)
      {
        log_llh_score <- log_llh_score+
        dmvnorm(Y[ind,],mean=H_vec[space_samples],
        sigma=time_period*Sig_phi+
        diag(length(space_samples))*measurement_var,
        log=TRUE)
      }
      else
```

```
        {
          log_llh_score <- log_llh_score+
          dmvnorm(Y[ind,],
          mean=H_vec[space_samples]-
          H_vec_prev[space_samples]+Y[ind-1,],
          sigma=time_period*Sig_phi+
          2*diag(length(space_samples))*measurement_var,
          log=TRUE)
        }
        H_vec_prev <- H_vec
      }
    }
    return(log_llh_score)
}
```

7.5 Log-likelihood evaluation (exact)

```
kronecker_mat <- matrix(rep(0,length(time_samples)*
length(time_samples)),nrow=length(time_samples))
for(i in 1:length(time_samples))
{
  for(j in 1:length(time_samples))
  {
    kronecker_mat[i,j] <- time_period*min(i,j)
  }
}
full_Sig <- kronecker(kronecker_mat, Sig_phi)

log_llh_B_exact <- function(B_prop)
{
  log_llh_score <- 0
  H_cur <- matrix(u_0,nrow=M)
  output_vector <- c()
  for(index in 1:Time)
  {
    H_cur <- matrix(forward_compute(H_cur,
    H_cur+R,R,B_prop,0,t_0+del_t*(index-1)),nrow=M)
    H_vec <- as.vector(H_cur)
```

```
      if(index %in% time_samples)
      {
        output_vector <- c(output_vector,
        H_vec[space_samples])
      }
    }
    log_llh_score <- dmvnorm(as.vector(t(Y)),
    mean=output_vector,
    sigma=full_Sig+measurement_var*
    diag(dim(full_Sig)[1]),log=TRUE)
    return(log_llh_score)
}
```

7.6 Emulator code

7.6.1 Code to train the emulator

```
theta_trial <- seq(10*10^(-25),70*10^(-25),
2.5*10^(-25))
X_train <- matrix(rep(0,length(theta_trial)))
training_data <- array(rep(0,length(time_samples)*
length(u_0)*length(theta_trial)),
dim=c(length(time_samples),
length(u_0),length(theta_trial)))

for(iter in 1:length(theta_trial))
{
  print(iter)
  theta_prop <- theta_trial[iter]
  X_train[iter,] <- theta_prop
  H_cur <- matrix(u_0,nrow=M)
  for(index in 1:Time)
  {
    H_cur <-
    matrix(forward_compute(
    H_cur,H_cur+R,R,theta_prop,0,t_0+
    del_t*(index-1)),
    ncol=M)
    if(index %in% time_samples)
```

```
    {
      training_data[which(index == time_samples)
      ,,iter] <- as.vector(H_cur)
    }
  }
}

SVD_left_sing_vector <-
array(rep(0,length(time_samples)*
length(u_0)*length(theta_trial)),
dim=c(length(time_samples),
length(u_0),length(theta_trial)))
SVD_sing_val <-
matrix(rep(0,length(theta_trial)*
length(time_samples)),
nrow=length(time_samples))
SVD_right_sing_vector <-
array(rep(0,length(time_samples)*length(theta_trial)*
length(theta_trial)),
dim=c(length(time_samples),length(theta_trial),
length(theta_trial)))

for(i in 1:length(time_samples))
{
  print(i)
  Y_SVD <- svd(training_data[i,,])
  SVD_left_sing_vector[i,,] <- Y_SVD$u
  SVD_sing_val[i,] <- Y_SVD$d
  SVD_right_sing_vector[i,,] <- Y_SVD$v
}
theta <- matrix(theta_trial,ncol=1)
test_vec <- matrix(rep(0,length(theta_trial)),ncol=1)

#train random forests
RForests <-vector(mode="list",
length=length(time_samples))
for(index in 1:length(time_samples))
{
```

```
  RForests[[index]] <- vector(mode="list",
  length=length(theta_trial))
}
for(index in 1:length(time_samples))
{
  for(i in 1:dim(SVD_right_sing_vector[index,,])[1])
  {
    RForests[[index]][[i]]<-
    randomForest(theta,SVD_right_sing_vector[index,,i])
  }
}
```

7.6.2 Code for log-likelihood with an emulator

```
log_llh_theta_emulator_2 <- function(theta_prop)
{
  log_llh_score <- 0
  output_vector <- c()
  index <- 1
  for(index in 1:length(time_samples))
  {
    for(i in 1:dim(SVD_right_sing_vector[index,,])[1])
    {
      test_vec[i] <-
      predict(RForests[[index]][[i]],theta_prop)
    }
    H_vec <- SVD_left_sing_vector[index,,]
    output_vector <-
    c(output_vector,H_vec[space_samples])
  }
  log_llh_score <- dmvnorm(as.vector(t(Y)),
  mean=output_vector,
  sigma=full_Sig+measurement_var*diag(dim(full_Sig)[1]),
  log=TRUE)
  return(log_llh_score)
}
```

7.7 R-INLA code

```
#see Bolin and Lindstrom (2017) R tutorial:
#https://www.stat.washington.edu/peter/591/INLA.html
#of which the following code is based on
cur_yr <- data.frame(read.csv(files[cur_file]))
names_dat <- names(cur_yr)
names_dat[4] <- 'elevation'
names(cur_yr) <- names_dat

#extract mass balance data
#(toggle for summer or winter balance)
Y <-cur_yr$bw
elevation <- cur_yr$elevation

setwd('/home/ec2-user/R_Data')
#calculate the border
load('x_index.RData')
load('y_index.RData')
load('bedrock_matrix.RData')
load(file='surface_matrix.RData')
x_vals <- x_index
y_vals <- y_index
top_border <-
matrix(rep(0,length(x_vals)*2),ncol=2)
for(i in 1:length(x_vals))
{
  top_border[i,] <-
  c(x_vals[i],y_vals[1])
}
bottom_border <-
matrix(rep(0,length(x_vals)*2),ncol=2)
for(i in 1:length(x_vals))
{
  bottom_border[i,] <-
  c(x_vals[i],y_vals[length(y_vals)])
}
left_border <-
matrix(rep(0,(length(y_vals)-2)*2),ncol=2)
for(i in 2:(length(y_vals)-1))
```

```
{
  print(i)
  left_border[i-1,] <-
   c(x_vals[1],y_vals[i])
}
right_border <-
matrix(rep(0,(length(y_vals)-2)*2),ncol=2)
for(i in 2:(length(y_vals)-1))
{
  right_border[i-1,] <-
  c(x_vals[length(x_vals)],y_vals[i])
}
PRborder <- rbind(
top_border,bottom_border,left_border,right_border)
#scale appropriately
locations <- as.matrix(cur_yr[,11:12])
#scale_locations for INLA
scaled_locations <- locations
shift_x <- min(PRborder[,1])
shift_y <- min(PRborder[,2])
#scale_factor <- max(PRborder)
scale_factor <- 50000

scaled_locations[,1] <- scaled_locations[,1]-shift_x#
scaled_locations[,2] <- scaled_locations[,2]-shift_y#
PRborder[,1] <- PRborder[,1]-shift_x
PRborder[,2] <- PRborder[,2]-shift_y
PRborder <- PRborder/scale_factor
coords <- scaled_locations/scale_factor

#create a mesh for SPDE method.
m1 <-
 inla.mesh.2d(coords,loc.domain=PRborder,max.edge=.1)

#observation matrix A
A <- inla.spde.make.A(prmesh,loc=coords)
#PC priors
spde <- inla.spde2.pcmatern(prmesh,alpha=2,
```

```
prior.range = c(.002,.01),prior.sigma = c(.3,.05))

#inla.stack function
mesh.index <- inla.spde.make.index(name='field',
n.spde=spde$n.spde)
stk.dat <-
inla.stack(data=list(y=Y),A=list(A,1),tag="est",
effects=list(c(mesh.index,list(Intercept=1)),
list(long=inla.group(coords[,1]),
lat=inla.group(coords[,2]),
elevation=inla.group(elevation))))

#linear model with SPDE random effect
f.s <- y ~ -1 + Intercept + long + lat +
elevation + f(field, model = spde)
#INLA fit with non-default fixed effects
r.s <- inla(f.s,family="t",
data=inla.stack.data(stk.dat),
verbose=TRUE,
control.predictor=
list(A=inla.stack.A(stk.dat),compute=TRUE),
control.fixed=
list(prec=list(elevation=.1,long=.1,lat=.1)))
```

References

Akaike, H. (1974), "A new look at the statistical model identification," in *Selected Papers of Hirotugu Akaike*, Springer, 215–222.

Aðalgeirsdóttir, G. (2003), "Flow dynamics of Vatnajökull ice cap, Iceland," *Mitteilungen der Versuchsanstalt fur Wasserbau, Hydrologie und Glaziologie an der Eidgenossischen Technischen Hochschule Zurich*.

Aðalgeirsdóttir, G., Gudmundsson, G. H., and Björnsson, H. (2000), "The response of a glacier to a surface disturbance: a case study on Vatnajökull ice cap, Iceland," *Annals of Glaciology*, 31, 104–110.

Aðalgeirsdóttir, G., Jóhannesson, T., Björnsson, H., Pálsson, F., and Sigurðsson, O. (2006), "Response of Hofsjökull and southern Vatnajökull, Iceland, to climate change," *Journal of Geophysical Research: Earth Surface*, 111.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018), "Spatial modeling with R-INLA: A review," *Wiley Interdisciplinary Reviews: Computational Statistics*, 10, e1443.

Banerjee, S., Carlin, B., and Gelfand, A. (2003), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press, URL https://books.google.com/books?id=A_R4AgAAQBAJ.

Baum, L. E. and Petrie, T. (1966), "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *Annals of Mathematical Statistics*, 37, 1554–1563, URL https://doi.org/10.1214/aoms/1177699147.

Berger, J. O., Bernardo, J. M., and Sun, D. (2009), "The formal definition of reference priors," *Annals of Statistics*, 37, 905–938, URL https://doi.org/10.1214/07-AOS587.

Berliner, L. M. (1996), "Hierarchical Bayesian Time Series Models," in Hanson, K. M. and Silver, R. N. (editors), *Maximum Entropy and Bayesian Methods*, Dordrecht: Springer Netherlands.

— (2003), "Physical-statistical modeling in geophysics," *Journal of Geophysical Research: Atmospheres*, 108, n/a–n/a, URL http://dx.doi.org/10.1029/2002JD002865. 8776.

Berliner, L. M., Jezek, K., Cressie, N., Kim, Y., Lam, C. Q., and van der Veen, C. J. (2008), "Modeling dynamic controls on ice streams: a Bayesian statistical approach," *Journal of Glaciology*, 54, 705–714.

Berrocal, V., Gelfand, A., and Holland, D. (2014), "Assessing exceedance of ozone standards: a space-time downscaler for fourth highest ozone concentrations," *Environmetrics*, 25, 279–291.

Björnsson, H. (1986), "Surface and Bedrock Topography of Ice Caps in Iceland, Mapped by Radio Echo-Sounding," *Annals of Glaciology*, 8, 11–18.

Björnsson, H. and Pálsson, F. (2008), "Icelandic glaciers," *Jökull*, 58, 365–386.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017), "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, 112, 859–877, URL https://doi.org/10.1080/01621459.2017.1285773.

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45, 5–32, URL https://doi.org/10.1023/A:1010933404324.

Brinkerhoff, D. J., Aschwanden, A., and Truffer, M. (2016), "Bayesian Inference of Subglacial Topography Using Mass Conservation," *Frontiers in Earth Science*, 4, 8, URL http://journal.frontiersin.org/article/10.3389/feart.2016.00008.

Brynjarsdóttir, J. and O'Hagan, A. (2014), "Learning about physical parameters: the importance of model discrepancy," *Inverse Problems*, 30, 114007, URL http://stacks.iop.org/0266-5611/30/i=11/a=114007.

Bueler, E., Lingle, C. S., Kallen-Brown, J. A., Covey, D. N., and Bowman, L. N. (2005), "Exact solutions and verification of numerical models for isothermal ice sheets," *Journal of Glaciology*, 51, 291–306.

Calderhead, B., Girolami, M., and Lawrence, N. D. (2008), "Accelerating Bayesian Inference over Nonlinear Differential Equations with Gaussian Processes," in *Proceedings of the 21st International Conference on Neural Information Processing Systems*, NIPS'08, USA: Curran Associates Inc., URL http://dl.acm.org/citation.cfm?id=2981780.2981808.

Castillo, I. and Nickl, R. (2014), "On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures," *Annals of Statistics*, 42, 1941–1969, URL https://doi.org/10.1214/14-AOS1246.

Chkrebtii, O. A., Campbell, D. A., Calderhead, B., Girolami, M. A., et al. (2016), "Bayesian Solution Uncertainty Quantification for Differential Equations," *Bayesian Analysis*, 11, 1239–1267.

Conrad, P. R., Girolami, M., Särkkä, S., Stuart, A., and Zygalakis, K. (2017), "Statistical analysis of differential equations: introducing probability measures on numerical solutions," *Statistics and Computing*, 27, 1065–1082.

Cressie, N. and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, John Wiley & Sons.

Cuffey, K. M. and Paterson, W. (2010), *The Physics of Glaciers*, Academic Press, 4 edition.

Diaconis, P. and Ylvisaker, D. (1979), "Conjugate Priors for Exponential Families," *Annals of Statistics*, 7, 269–281, URL https://doi.org/10.1214/aos/1176344611.

Flowers, G. E., Marshall, S. J., Björnsson, H., and Clarke, G. K. (2005), "Sensitivity of Vatnajökull ice cap hydrology and dynamics to climate warming over the next 2 centuries," *Journal of Geophysical Research: Earth Surface*, 110.

Fowler, A. C. and Larson, D. A. (1978), "On the Flow of Polythermal Glaciers. I. Model and Preliminary Analysis," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 363, 217–242, URL http://www.jstor.org/stable/79748.

Friedman, J., Hastie, T., and Tibshirani, R. (2001), *The Elements of Statistical Learning*, volume 1, Springer series in statistics New York, NY, USA:.

Geirsson, Ó. P., Hrafnkelsson, B., and Simpson, D. (2015), "Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid," *Environmetrics*, 26, 339–353, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2343.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), "Bayesian Data Analysis, 3rd edition," .

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2019), *mvtnorm: Multivariate Normal and t Distributions*, URL https://CRAN.R-project.org/package=mvtnorm. R package version 1.0-10.

Glen, J. (1958), "THE FLOW LAW OF ICE A discussion of the assumptions made in glacier theory, their experimental foundations and consequences," *International Association Hydrological Sciences Publications*, 47, 171–183.

Glen, J. W. (1955), "The creep of polycrystalline ice," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 228, 519–538, URL http://www.jstor.org/stable/99642.

Golub, G. H. and Van Loan, C. F. (2012), *Matrix Computations*, volume 3, Johns Hopkins University Press.

Gopalan, G. and Bornn, L. (2015), "FastGP: An R Package for Gaussian Processes," *arXiv preprint arXiv:1507.06055*.

Gopalan, G., Hrafnkelsson, B., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F. (2018), "A Bayesian hierarchical model for glacial dynamics based on the shallow ice approximation and its evaluation using analytical solutions," *The Cryosphere*, 12, 2229–2248.

Gopalan, G., Hrafnkelsson, B., Wikle, C. K., Rue, H., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F. (2019), "A Hierarchical Spatio-Temporal Statistical Model Motivated by Glaciology," *arXiv e-prints*, arXiv:1811.08472.

Gu, M., Wang, X., and Berger, J. O. (2018), "Robust Gaussian stochastic

process emulation," *Annals of Statistics*, 46, 3038–3066, URL https://doi.org/10.1214/17-AOS1648.

Guan, Y., Haran, M., and Pollard, D. (2016), "Inferring Ice Thickness from a Glacier Dynamics Model and Multiple Surface Datasets," *ArXiv e-prints*.

Gupta, A. and Kumar, V. (1994), "A scalable parallel algorithm for sparse Cholesky factorization," in *Proceedings of the 1994 ACM/IEEE Conference on Supercomputing*, Supercomputing '94, Los Alamitos, CA, USA: IEEE Computer Society Press, URL http://dl.acm.org/citation.cfm?id=602770.602898.

Håvard, R., Sara, M., and Nicolas, C. (2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 319–392, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), "Computer Model Calibration Using High-Dimensional Output," *Journal of the American Statistical Association*, 103, 570–583.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM Journal on Scientific Computing*, 26, 448–466.

Hijmans, R. J. (2018), *raster: Geographic Data Analysis and Modeling*, URL https://CRAN.R-project.org/package=raster. R package version 2.8-4.

Hoerl, A. E. and Kennard, R. W. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12, 55–67, URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634.

Hooten, M. B., Leeds, W. B., Fiechter, J., and Wikle, C. K. (2011), "Assessing First-Order Emulator Inference for Physical Parameters in Nonlinear Mechanistic Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 16, 475–494, URL https://doi.org/10.1007/s13253-011-0073-7.

Hudson, J. (1998), "Numerical Techniques for Conservation Laws with Source Terms," Technical report, Engineering and Physical Science Research Council.

Hutter, K. (1982), "A mathematical model of polythermal glaciers and ice sheets," *Geophysical & Astrophysical Fluid Dynamics*, 21, 201–224, URL https://doi.org/10.1080/03091928208209013.

— (1983), *Theoretical Glaciology: Material Science of Ice and the Mechanics of Glaciers and Ice Sheets*, Mathematical Approaches to Geophysics, Springer, URL https://books.google.com/books?id=75kqTGNKV9wC.

Isaac, T., Petra, N., Stadler, G., and Ghattas, O. (2015), "Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet," *Journal of Computational Physics*, 296, 348 – 368, URL http://www.sciencedirect.com/science/article/pii/S0021999115003046.

Jarosch, A. H., Schoof, C. G., and Anslow, F. S. (2013), "Restoring mass conservation to shallow ice flow models over complex terrain," *The Cryosphere*, 7, 229–240, URL https://www.the-cryosphere.net/7/229/2013/.

Jeffreys, H. (1946), "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London Series A*, 186, 453–461.

Kennedy, M. C. and O'Hagan, A. (2001), "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 425–464.

Kusnierczyk, W. (2012), *rbenchmark: Benchmarking routine for R*, URL https://CRAN.R-project.org/package=rbenchmark. R package version 1.0.0.

Lehmann, E. and Casella, G. (2003), *Theory of Point Estimation*, Springer Texts in Statistics, Springer New York, URL https://books.google.com/books?id=0q-Bt0Ar-sgC.

Liaw, A. and Wiener, M. (2002), "Classification and Regression by randomForest," *R News*, 2, 18–22, URL https://CRAN.R-project.org/doc/Rnews/.

Lindgren, F., Rue, H., and Lindström, J. (2011), "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423–498.

Liu, F. and West, M. (2009), "A dynamic modelling strategy for Bayesian computer model emulation," *Bayesian Analysis*, 4, 393–411, URL

https://doi.org/10.1214/09-BA415.

Madsen, H. (2007), *Time Series Analysis*, Chapman and Hall/CRC.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012), "Approximate Bayesian computational methods," *Statistics and Computing*, 1–14.

Minchew, B., Simons, M., Hensley, S., Björnsson, H., and Pálsson, F. (2015), "Early melt season velocity fields of Langjökull and Hofsjökull, central Iceland," *Journal of Glaciology*, 61, 253–266.

Murray, I., Adams, R. P., and MacKay, D. J. (2010), "Elliptical slice sampling," *Journal of Machine Learning Research W&CP*, 9, 541–548.

Neal, R. (2011), "MCMC Using Hamiltonian Dynamics," *Handbook of Markov Chain Monte Carlo*, 113–162.

Neal, R. M. (2003), "Slice sampling," *Annals of Statistics*, 705–741.

Owhadi, H. and Scovel, C. (2017), "Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis," *ArXiv e-prints*.

Pagendam, D., Kuhnert, P., Leeds, W., Wikle, C., Bartley, R., and Peterson, E. (2014), "Assimilating catchment processes with monitoring data to estimate sediment loads to the Great Barrier Reef," *Environmetrics*, 25, 214–229.

Pálsson, F., Guðmundsson, S., Björnsson, H., Berthier, E., Magnússon, E., Guðmundsson, S., and Haraldsson, H. H. (2012), "Mass and volume changes of Langjökull ice cap, Iceland, 1890 to 2009, deduced from old maps, satellite images and in situ mass balance measurements," *Jökull*.

Payne, A. J., Huybrechts, P., Abe-Ouchi, A., Calov, R., Fastook, J. L., Greve, R., Marshall, S. J., Marsiat, I., Ritz, C., Tarasov, L., and Thomassen, M. P. A. (2000), "Results from the EISMINT model intercomparison: the effects of thermomechanical coupling," *Journal of Glaciology*, 46, 227–238.

Pralong, M. R. and Gudmundsson, G. H. (2011), "Bayesian estimation of basal conditions on Rutford Ice Stream, West Antarctica, from surface data," *Journal of Glaciology*, 57, 315–324.

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org.

Rasmussen, C. and Williams, C. (2006), *Gaussian Processes for Machine*

Learning, Adaptative computation and machine learning series, University Press Group Limited, URL https://books.google.com/books?id=vWtwQgAACAAJ.

Robert, C. (2007), *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer Texts in Statistics, Springer New York, URL https://books.google.com/books?id=NQ5KAAAAQBAJ.

Rue, H. (2001), "Fast sampling of Gaussian Markov random fields," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 325–338.

Rue, H. and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, CRC press.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017), "Bayesian Computing with INLA: A Review," *Annual Review of Statistics and Its Application*, 4, 395–421.

Salter, J. M., Williamson, D. B., Scinocca, J., and Kharin, V. (2019), "Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases," *Journal of the American Statistical Association*, 0, 1–24, URL https://doi.org/10.1080/01621459.2018.1514306.

Sargsyan, K. (2016), *Surrogate Models for Uncertainty Propagation and Sensitivity Analysis*, Cham: Springer International Publishing, 1–26, URL https://doi.org/10.1007/978-3-319-11259-6_22-1.

Shen, X. and Wasserman, L. (2001), "Rates of convergence of posterior distributions," *Annals of Statistics*, 29, 687–714, URL https://doi.org/10.1214/aos/1009210686.

Sigurdarson, A. N. and Hrafnkelsson, B. (2016), "Bayesian prediction of monthly precipitation on a fine grid using covariates based on a regional meteorological model," *Environmetrics*, 27, 27–41, URL https://ideas.repec.org/a/wly/envmet/v27y2016i1p27-41.html.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017), "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors," *Statistical Science*, 32, 1–28, URL https://doi.org/10.1214/16-STS576.

Solin, A. and Särkkä, S. (2014), "Explicit Link Between Periodic Covariance Functions and State Space Models," in Kaski, S. and Corander, J. (editors), *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, Reykjavik, Iceland: PMLR, URL http://proceedings.mlr.press/v33/solin14.html.

Stan Development Team (2018), "RStan: the R interface to Stan," URL http://mc-stan.org/. R package version 2.17.3.

Sverrisson, Marteinn and Jóhannesson, Ævar and Björnsson, Helgi (1980), "Instruments and Methods: Radio-Echo Equipment for Depth Sounding of Temperate Glaciers," *Journal of Glaciology*.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288, URL http://www.jstor.org/stable/2346178.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22, 1701–1728, URL https://doi.org/10.1214/aos/1176325750.

van der Vaart, A. (2000), *Asymptotic Statistics*, Asymptotic Statistics, Cambridge University Press, URL https://books.google.com/books?id=UEuQEM5RjWgC.

van der Veen, C. (2013), *Fundamentals of Glacier Dynamics*, CRC Press, 2 edition.

Wasserman, L. (2010), *All of Statistics: A Concise Course in Statistical Inference*, Springer.

Weertman, J. (1964), "The Theory of Glacier Sliding," *Journal of Glaciology*, 5, 287–303.

Whittle, P. (1954), "ON STATIONARY PROCESSES IN THE PLANE," *Biometrika*, 434–449.

— (1963), "Stochastic processes in several dimensions," *Bulletin of the International Statistical Institute*, 40, 974–994.

Wikle, C. K. (2016), *Hierarchical Models for Uncertainty Quantification: An Overview*, Springer International Publishing, 1–26.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), "Hierarchical Bayesian space-time models," *Environmental and Ecological Statistics*, 5, 117–154, URL https://doi.org/10.1023/A:1009662704779.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), "Spatiotemporal Hierarchical Bayesian Modeling Tropical Ocean Surface Winds," *Journal of the American Statistical Association*, 96, 382–397.

Zammit-Mangion, A., Rougier, J., Bamber, J., and Schön, N. (2014), "Resolving the Antarctic contribution to sea-level rise: a hierarchical modelling framework," *Environmetrics*, 25, 245–264.