Dear Dr. Eric Larour, Dr. Lambert Caron, and the anonymous second referee:

We are very pleased by your thorough, incisive, and constructive comments, which we believe have greatly improved the quality of the revised paper. Please see our responses to all of the referee comments, followed by a pointer to the modifications in the revised text. Please note that we have bolded the manuscript changes as to be able to easily identify them. Also, kindly note that to mitigate redundancy, we have referenced the modifications made to the paper in the attached revision instead of copying the modifications into this document.

Sincerely,
Giri Gopalan and co-authors

Response to referee 1:

The figures and tables are straightforward to interpret and informative, although in some cases expanded captions would be more useful if more information was reminded to the user given the length of the paper (see technical comments below).

We appreciate the suggestion to make the captions more informative. Accordingly, we have modified the specific captions referenced in the technical comments, and we have also modified all of the additional captions as well.

1. The authors briefly mention in the summary and discussion that the method is applicable to broader problems in cryosphere science. Without going into further calculation, I believe expanding on that topic in the discussion would both provide better contextualization of the problem tackled here and increase the impact of the paper. What challenges do you expect for the cryosphere science community to apply BHM approach to the non-SIA regime, e.g. for fast-discharge ice streams, or to SIA problems without analytical solution (e.g. more realistic geometry)?

We appreciate here the call to discuss the generality of this approach with respect to problems in cryosphere science. Essentially, the same BHM can be used for other cryosphere science problems by swapping out the numerical solver for the SIA with a numerical solver to a different set of dynamics. The biggest challenge with this is satisfactorily modeling the numerical errors of this solver in a general way, discussed in the following paragraph.

Changes in revision:

- Paragraph starting on line 11, page 5.

2. *The authors manage well to point out the limitations introduced by simplifications in the physical problem and choices for the statistical distribution of errors. However, even after a few readings, it remains a bit difficult for me to tell what are the limits or downsides of the BHM approach itself, particularly on the resolution of parameters and state variables (e.g. ice thickness or velocity field).*

The BHM approach is not infallible, and the biggest difficulty is in actually fitting a BHM given data (despite that coming up with a BHM may not be too difficult). In our test cases, the number of physical parameters is small (1 or 2), so model fitting is not computationally difficult. However, when the number of physical parameters becomes larger (e.g., a basal sliding field with a parameter for every spatial location), posterior computation will become inefficient, and more sophisticated approaches will be needed (for which there is a large battery of tools). Besides a large number of physical parameters, another potential difficulty in utilizing BHMs that incorporate physical dynamics via a numerical solver is that the numerical solver can also be computationally onerous, so that posterior computation is very inefficient. While this is not a hindrance in the examples studied in this paper, in general this can be problematic. We have discussed this in the following paragraph.

Changes in revision:

- Paragraph starting on line 17, page 3.

- In the context of the SIA equations, can you say something about the relationship between the number of observations and the number of parameters? That is, how does the posterior evolve in the different cases with respect to the number of observations?

The work of Brynjarsdóttir and O'Hagan (2014) gives us an indication of how the posterior for physical parameters will evolve with more observations. In their work, they show that in a simple physical system with only a single parameter, some uncertainty in the posterior distribution for the physical parameter won't go away even as more data is collected. While this is attributed to improper modeling of the discrepancy between the output of a computer simulator and actual physical process values, and we have taken care in doing this, it is plausible that a similar phenomenon would occur in the BHM; that is, confounding between an error correcting process and the posterior of physical parameters results in posterior uncertainty never going away completely. In concordance with both this comment and the second referee's comments, we have included posterior distributions for ice viscosity where the sampling period varies from once every 10 years, once every 5 years, once every year, and twice a year (as originally done to be consistent with summer and winter mass balance measurements per year). What we found is displayed in Figure 9; the general trend is that as more data is collected, the posterior becomes less biased but more diffuse. Thus, having posterior uncertainty that doesn't go away as more data is collected appears consistent with Brynjarsdóttir and O'Hagan (2014).

Changes in revision:

- Paragraph starting on line 8, page 13.
- Addition of Figure 9.

- Given the symmetry and choice of displaying only one quadrant in Figure 4, I wonder if the information (or uncertainty quantification) retrieved on the ice viscosity reflects that of 8 observations or that of 32. If one would compute a similar problem with a non-idealized glacier, how many observations would one need to obtain a similar posterior distribution for ice viscosity?

To clarify, Figure 4 displays test locations for predictions, but not the locations where data are collected. The locations where surface elevation data are collected are distributed across the glacier at 25 locations as delineated in Section 3.2; to clarify this we have included a map marking the locations of these measuring sites. We will update the manuscript to clarify this in the caption. A similar number of locations would be adequate in a non idealized glacier, so long as the locations included points of steep changes in glacial thickness (e.g., valleys and peaks), since it appears that numerical errors are largest at such locations (i.e., the dome and margin).

Changes in revision:

- Lines 17 and 18 of page 9.
- Addition of Figure 4.
- Caption of Figure 5.

- Similarly, do the authors expect the spatial distribution of the observations to play a critical role in determining the posterior given the different sensitivity of the dome, margin and interior of the glacier?

The biggest numerical errors occur where there are sharp changes in glacial thickness (i.e., peaks and valleys), such as the dome and the margin in the idealized glacier studied in Bueler (2005) and in this work. It is crucial, therefore, to ensure that such locations are sampled. Qualititatively, it is suggested that locations where there is a rapid change in glacier thickness ought to be sampled to ensure that numerical errors are adequately represented; the locations sampled in our simulation study include both the dome and locations close to the margin. The number of samples needed in general will then depend on the number of peaks and valleys in the glacier.

3. Although the true value always remains within the confidence interval, there seems to be a tendency to under-predict the ice viscosity (as seen in Table 2) and over-predict the thickness (Figure 5). Is there any reason for that or is this purely the result of randomness?

A very similar phenomenon has been documented in the work of Brynjarsdóttir and O'Hagan (2014); in their work, it is noted that good prior information must be encoded into model discrepancy (essentially what we have termed an error correcting process) and physical parameters in order to get a less biased posterior distribution for both physical parameters and predictions. A similar phenomenon can be demonstrated in our BHM. In particular, if we consider a scenario where we ignore the prior information regarding different scales of numerical errors between the interior, dome, and margin of the glacier, the bias of the posterior distribution for physical parameters is more pronounced. So while bias of the posterior of physical parameters exists in our simulation studies, the prior information we have used appears to have helped reduce this bias, consistent with the findings of Brynjarsdóttir and O'Hagan (2014). We have revised the manuscript to include an example illustrating this point in the results section.

Changes in revision:

- Paragraph starting on line 27, page 12.
- Addition of Figure 8.

4. In a non-linear PDE system, it is not guaranteed that the posterior is Gaussian or even symmetric distribution (even when propagating Gaussian errors). While the authors put a certain emphasis on the ice viscosity and basal sliding parameter, with respect to which the problem is linear, this linearity might not hold in general for every parameter or state variable one might want to keep track of. After all, a major appeal of Bayesian methods is that they require no assumption on the physics that are being solved, and are thus well suited to nonlinear problems.

As a minor point of clarification for readers, the PDE on line 11 page 6 is non-linear in H, glacial thickness, since it involves powers of H. However, I suspect the use of linear here refers to the fact that B and C_0gamma appear as constants (i.e., not functions thereof) in these equations.

With that in mind, I believe using an accurate but more general terminology would be beneficial to future users of this work:

- p10 l18-20: "the .99 posterior credibility interval was computed by taking 3 standard deviations below and above the maximum a posteriori estimate (MAP) of the posterior samples." Even though these indicators are equal for a Gaussian (or any symmetric) distribution, as a principle I would advise to refer to the mean or median instead of the maximum, as the former remain comparatively more adapted to characterize distributions even when they are not Gaussian. Perhaps the authors should also remind the reader that in a general (non-Gaussian) case, a distribution is best characterized by multiple indicators, e.g. quantiles as in Figure 5, and not just maximum and standard deviation.

The MAP was used to be consistent with the previous related work in Brinkerhoff et al. (2016), but we have used the mean instead of the MAP in the revision (the results are essentially the same). It should be noted that constructing a credibility interval in this way (mean +/- 3 sd of posterior samples) does not necessitate that the posterior distribution is Gaussian.

Changes in revision:

- Sentence starting on line 3, page 12.
- Sentences starting on line 15, page 12.
- Sentence starting on line 23, page 12.
- Change of .99 credibility interval to 3-sd credibility interval in Table 2.

- Throughout the manuscript the authors use interchangeably the phrases "3-Sigma" and ".99 Confidence" interval, as pointed out above. In a Gaussian distribution, the 3-sigma interval accounts for 0.9973 of the integral while the .99 interval represents 2.58-sigma, and clearly these are not the same. I think the authors should clarify and streamline this. It might otherwise introduce confusions and discrepancies in the exact numbers for readers that try to reproduce the results or compare them with a slightly different model setup (e.g. different geometry), especially if their method is based on numerical integration of the posterior.

Thank you for pointing out the potential confusion this can cause. To be consistent, we have updated the terminology to be '3-sd credibility interval' (again, constructed with mean +/- 3 sd of posterior samples).

Also it should be noted that credibility interval refers to an interval derived from a posterior distribution, which is distinct from a confidence interval. The latter has a particular frequentist coverage probability.

The changes in the revision are the same as above, that is:

- Sentence starting on line 3, page 12.
- Sentences starting on line 15, page 12.
- Sentence starting on line 23, page 12.
- Change of .99 credibility interval to 3-sd credibility interval in Table 2.

- I recommend the authors to display the posterior distribution of μmax, as a supplemental figure. Likewise, Figure 7 suggests non-symmetric probability distributions of the thickness originating from the error propagation, it might be beneficial to highlight the non-linearity by plotting these distributions in a similar way as Figure 6.

We have included a posterior plot of mu_max in the supplemental materials. In our humble collective opinion, individual predictive density plots do not appear to convey more information than Figure 7, so we have opted not to include these.

III. Technical comments

-Figure 5: Outside of the whiskers, small circles are displayed, but the caption doesn't indicate what they are. If they are important, the authors should improve their visibility and add explanations related to them in the caption. If these are not meaningful on the other hand, the authors should remove them.

It is typical for box and whisker plots to display outliers, defined as more than 1.5 times the interquartile range beyond the first and third quartiles; these outliers are displayed as circles. Agreeably, it is important for us to be clear about this, so we have included a note in the caption.

Changes in revision:

- Last two sentences of the caption in Figure 6.

-Figures 5, 6, 7: when referring to test cases, remind the readers the specificity of these tests, e.g. "test case B (no mass balance or basal sliding)". This would lessen the need for cross-referencing.

Thank you for this suggestion, which we have taken heed of in the revision.

Changes in revision:

- Caption of Figure 6.
- Caption of Figure 7.
- Caption of Figure 8.
- Caption of Figure 9.
- Caption of Figure 10.

-Table 2: The exponents of units are not displayed in superscript.

Thank you for spotting this, which we have corrected.

Changes in revision:

- Last row of Table 2.

-Table 3: Is the dome error not calculated the same way as the margin and the interior? If so, I did not find any explanation in the text. If not, I suggest that the authors streamline the column labels. Also, the authors should expand in the caption what RMSE stands for.

The dome error is calculated in the same way as the margin and interior, but since there is only a single dome observation, RMSE, which stands for root mean squared error, is just the absolute difference between the actual and predicted. Nonetheless, to remain consistent we have changed dome error to be RMSE. Thank you for pointing out that we ought to include what RMSE stands for, which we have revised in the manuscript.

Changes in revision:

- Header of Table 3.
- Caption of Table 3.

-Table 4: The authors should remind in the caption what the different symbols refer to.

We certainly agree and have revised the manuscript accordingly.

Changes in revision:

- Caption of Table 4.

I hope the authors will find this useful.

These comments have been extremely valuable for improving the manuscript; the referee's time and effort are appreciated.

Response to referee 2:

I believe this is an interesting, useful contribution and publishable with some revisions. Essentially, your computations assess the errors in using the numerical approximations for "f" using analytical solutions as a base line. That is, you generate "Y's" with analytical solutions but then forget about that and use numerical approximations in the BHM. "error" is then viewed as differences between Bayesian results and the analytical "truth". This is valuable work, though as you make clear, it doesn't make any assurances when the analytical model is "replaced by nature" in producing data. You also considered several cases, but I do think that your paper would be strengthened if you also studied the impact sampling plans and sample sizes (ie. What if "every other observation (in time) was removed? This is also critical in judging the impacts of your approximations used in computations (see the next paragraph).

In order to consider the impact of sampling plans, we have conducted an additional set of simulation studies where the period of observations varies: once every 10 years, once every 5 years, once every year, and twice a year (as originally conducted); however, please note that we chose two measurements per year to model how the data set from the University of Iceland was collected -- namely, a set of measurements for winter and summer mass balance.

Changes in revision:

- Paragraph starting on line 8, page 13.
- Addition of Figure 9.

My first concern is correctness of all contributions. Errors can occur when manipulating equations rather than probability distributions. I think yours turned out right, but all conditioning assumptions are not clear. Consider Appendix B1 beginning at the bottom of p. 20. The "overall model" as written at the top of p. 21 is quite brief and does not include probability assumptions. I sense that you understand the key issues based on the sentence in lines 16-17, p. 21. Namely, equations like Y = m(variables) + error are code for "the conditional distribution of Y given "variables" and the mean of "error" = 0 and some variance of "error" has conditional mean m and conditional variance equal to the variance of "error".

Thank you for pointing out some places where the probabilistic assumptions of the BHM can be made clearer. The probabilistic assumptions are specified in Section 2.2 and in the first paragraph of Appendix B. In particular, please note that in the first paragraph of Appendix B line 21 it is stated: "let epsilon_j be an independent and identically distributed MVN(0,Σ) noise term at time j", and also in Appendix B line 21 it is stated: "the corresponding observation error Zk,Z2k,...ZNk is i.i.d MV N(0,σ2 I)". The use of acronyms may make these lines unwieldy to parse, so we have revised them (that is, i.i.d is independent and identically distributed, and MVN(0,Σ) is multivariate normal with mean 0 and covariance Σ).

Additionally, we have stressed in the revision that we are conditioning on theta when computing the likelihood.

Changes in revision:

- Lines 21-23, page 24
- Lines 28-30, page 24.

The assertion that all "errors" in you models have mean zero seems to be missing, but more importantly, when you do the manipulation leading to line 14, you must have assumed both models for Yck and Y(c-1)k are conditioned on the same quantities so you can simply subtract their conditional means, etc. Further, simply taking differences of Yck and Y(c-1)k is based on their joint distribution, so cavalierly moving Y(c-1)k to the left hand side and claiming you're now looking as the distribution of Yck given Y(c-1)k and the other variables. That requires a probability computation (moving from joint to a conditional distribution) in general. Fortunately, it is common that the algebraic versions can actually be proven to be correct probabilistically for "linear manipulations", but in complicated settings, this needs to be checked (based on my quick check, I think you're OK but think you should check as well). This all relates to my suggestion that your model isn't simply lines 2-4, p. 21. What are the conditional distributions assumption (the Z's are independent etc.)?

Thank you for the call to clarify the arguments made in this section. Regarding errors having mean zero, please note that this is stated in the first paragraph of Appendix B. Regarding the remaining comments, we have taken a number of actions.

First, we have included a derivation of the complete likelihood without using any approximations before going into the approximation.

Second, we have clarified an important point you have raised, which is that the expression on line 14 page 21 cannot be used to claim that the distribution of p(Y_ck|Y_(c-1)k) is *exactly* a MVN distribution with mean $A[f(\theta, ck) - f(\theta,(c-1)k)]$ and covariance matrix $A(k\Sigma)A + 2\sigma_2 I$; this is because Y_(c-1)k and Z_(c-1)k are not independent. This expression, rather, motivates approximating p(Y_ck|Y_(c-1)k) with said MVN distribution. We have rewritten the text in this portion to be more clear about this point.

Third, we have included a simple example illustrating why this approximation is reasonable under the assumption that the output of the numerical solver is much larger in magnitude compared to the measurement error.

Changes in revision:

- Subsection B1.1 on page 25.
- Paragraph starting on line 4 on page 26.
- Supplementary note with an illustrative example regarding the approximation used.

One more related issue involves discussion of inference for X's. In a sense, you should be careful in posterior inferences about both S and X simultaneously, given q. (they are simply linear functions of each other). Again, I think you're OK but it merits your attention.

Indeed, as per the equations on the top of p. 21 (of the original submission, but page 25 of the revision), *conditioning on theta*, S is just X shifted by the output of the numerical solver. It is very important to stress that this is conditional on theta (the output of the numerical solver is fixed conditioning on theta).

I think the approximations you used on p. 21 are reasonable, but a bit more defense would be good.

Please see the aforementioned simple example in the supplemental materials regarding this issue.

Further, I'm not comfortable with the way you needed all the approximations so that you could use grid sampling to claim genuine posterior inference. I think that you could skip the approximations and did a full MCMC approach, it wouldn't be as easy as what you did but it's not that much harder. I think you should at least try some MCMC to confirm your computations and approximations.

It should be made clear that the approximation for the likelihood was not needed to compute the posterior on a grid. Many (though not all) MCMC algorithms require (log) likelihood evaluations as well, and a computationally inefficient (log) likelihood will mitigate their performance. Since we are working with only 1 to 2 physical parameters in test cases (B for the first three test cases, and B,mu_max in the last test case), computing the posterior on a grid ought to perform just as good as an MCMC approach, provided a sufficiently fine grid is chosen. Moreover, it is important to point out that the output of MCMC samples can be flawed (e.g., only one mode of a complex posterior is explored) and the samples thus may not actually reflect the posterior distribution.

This being said, we admit that we could have done a better job of checking the sensitivity of the grid sampling approach to the particular grid we had used. To check that the posterior for physical parameters is not sensitive to using a grid, we have computed the posterior on various grid widths for comparison, and have quoted summary statistics for posterior samples for comparison. The summary statistics are very close, indicating that the choice of grid width we had used did not distort or severely misrepresent the posterior distribution.

Changes in revision:

- Section B2 from line 14 onwards.

Further, what is the dependence of the value of your approximations on f. Surely you need to answer this if you plan to suggest operational use of you programs as you suggest you will do in the future.

We appreciate here the call to clarify the requirement for applying the likelihood approximation. The requirement for applying the likelihood approximation is that the values of S, and consequently of f, are much less than the measurement error -- this holds in the scenario of this paper because the values of f are on the order of one kilometer, whereas the measurement error is one meter. Please see the aforementioned simple illustrative example in the supplemental figure for justification on this point. However, the point raised about applying the code to future scenarios is certainly valid, since these conditions aren't always going to hold. We are currently in development of a way to efficiently calculate the log-likelihood in regimes where this does not hold (such as in other cryosphere problems with a poorer signal to noise ratio) and will include this functionality in the package mentioned.

Changes in revision:

- Supplementary note with an illustrative example regarding the approximation used.

Other Notes:

(1) The model for X is an explosive autoregression and hence you have built-in a limitation. A non-explosive model could be Xj = r Xj-1 + error where 0 < r < 1. If you make r a parameter and let the data tell you about r, you may be able to predict further in the future if the data suggests r can be much smaller than 1.

We agree that for more general models, it would be good to learn the parameter r directly from the data. However, for the time scales for the analysis used in this work, and based on the evidence from Figure 10, using r = 1 seemed to work out adequately. Another complication to consider is that learning an additional parameter along with the physical parameters can increase computational difficulties of posterior computation, though this is something we ought to more thoroughly investigate for future extensions of this model.

2) I think you missed emphasizing a crucial (and related) contribution of Berliner et al (2008). Namely they also treat model error through their "corrector process" and this should be mentioned.

We appreciate the call to highlight this important contribution of Berliner et al. (2008) and have accordingly updated the text. Please note that the error correcting process in that work accounts for setting basal shear stress to driving stress, a simplification. Somewhat orthogonally, the error correcting process in this paper solely accounts for numerical errors due to an imperfect numerical solver, though it is still important to consider the fact that stress terms are not physically perfect as well.

Changes in revision:

- Lines 7 and 8 on page 4.

(3) As a minor point, you should include at least one reference to Berliner, L.M. 1996. Hierarchical Bayesian time series models. In Hanson, K. and R. Silver, eds. Maximum entropy and Bayesian methods. Dordrecht, etc., Kluwer Academic Publishers, 15–22. The references by Wikle and Cressie both reference it but you should too since it urges the "data model, process model, parameter model" view.

Thank you for pointing out this additional reference, which we have added to the manuscript.

Changes in revision:

- Line 3 on page 3.

Also, since that paradigm is so key in your paper, I think you should break out the formula in line 22, p. 2 as a separate line for emphasis.

Though there is no formula on line 22, page 2, we have broken out the formula on line 20, page 2 as suggested.

To reiterate, the comments from the reviewers are greatly appreciated, and we believe they have helped us significantly improve the quality of this work.

Additional materials: We have shared R scripts written for this paper in the supplemental materials, in case that they may be helpful for the community. As such, we have included scripts to: compute the analytical solutions in test cases B-E, run the finite difference method for test cases B-E, generate the simulations based on analytical solutions in test cases B-E.

A Bayesian Hierarchical Model for Glacial Dynamics Based on the Shallow Ice Approximation and its Evaluation Using Analytical Solutions

Giri Gopalan[1], Birgir Hrafnkelsson[1], Guðfinna Aðalgeirsdóttir[2], Alexander H. Jarosch[2], and Finnur Pálsson[2]

[1]Faculty of Physical Sciences, School of Engineering and Natural Sciences; University of Iceland
[2]Institute of Earth Sciences; University of Iceland

Correspondence to: Giri Gopalan (gopalan88@gmail.com)

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.

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 (2015), 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). 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.

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 1, and the joint probability distribution is

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

The distribution $p(\theta)$ represents prior beliefs about parameters before data is 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 $p(Y, S, \theta)$ by Bayes' theorem. Estimates and assessments of uncertainty of scientific parameters and quantities can be extracted from the posterior distribution.

1.2 An Overview of 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 precise functional form of the underlying physical equations describing the physical phenomenon being modeled, or else

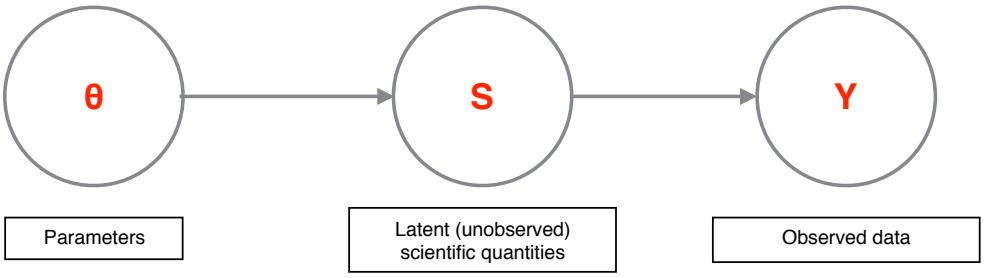

Figure 1. Schematic of a simple Bayesian hierarchical model; **here, θ represents physical parameters, S represents unobserved scientific quantities of interest, and. Y represents the observed data.**

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, 2015)** 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 2 diagrams what a prototypical physical-statistical Bayesian hierarchical model might look like.

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

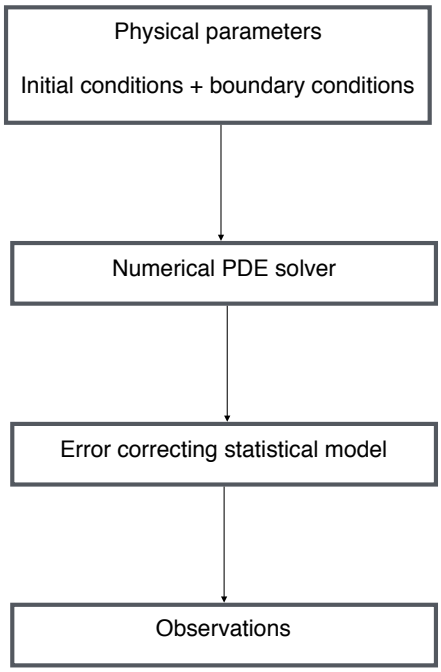

Figure 2. Schematic of a prototypical physical-statistical Bayesian hierarchical model. **At the top layer, physical parameters, initial conditions, and boundary conditions are fed into a numerical solver, and the output of this is corrected with an error correcting process; finally the actual observations are dependent on the actual physical process values.**

method can be computationally onerous, generally speaking; while this is not a detriment in the work that follows, this can be a problem 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.

2 Description of Models

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 (2017) 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).

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 1 for a summary of the model parameters utilized and a

schematic illustrating the BHM in Figure 3. 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 $S_{.,j}$ 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 (2015), 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 N(μ,τ^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** $Y_{ij} \sim \text{N}(S_{ij}, \sigma^2)$, 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 a 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 μ_{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, μ_{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: $X_{j+1} = X_j + \epsilon_{j+1}$, 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 square-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 $[1, 70] \times 10^{-24}$. $\mu_{\max}$ 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., square 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 example, 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, for example 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 is 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 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.

3 Experiments to assess the Bayesian hierarchical model

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 approximation of the true physical dynamics governing a glacier.) These analytical solutions serve as a basis for simulating data sets to validate

Parameter Name	Symbol	Description
Time index	j	A subscript which refers to discrete time points
Spatial index	i	A subscript which refers to discrete spatial points
All spatial points for a time index	$.,j$	Refers to entire spatial field at time j
ice viscosity	B	Key physical parameter driving the SIA
Basal sliding	$C_0\gamma$	Basal sliding field and key parameter driving the SIA
Max basal sliding	μ_{max}	Parameter for the basal sliding field of test case E in Bueler et al. (2005)
Physical parameters	θ	Refers to physical parameters
Measurement error	σ	Measurement error of surface elevation measurements
Error correcting covariance matrix	Σ	Covariance matrix used for the error correcting process
Error correcting parameters	$(\sigma_{\text{dome}}, \sigma_{\text{interior}}, \sigma_{\text{margin}}, \phi)$	Parameters corresponding to Σ
Mass balance field	$\dot{b}_{.,j}$	Mass balance field at time index j
Initial surface elevation	S_0	Initial surface height of the glacier

Table 1. A summary of main parameters and notation utilized.

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.

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 illustrates the locations of these measurement sites on the glacier.** Surface elevation measurements for these sites are taken

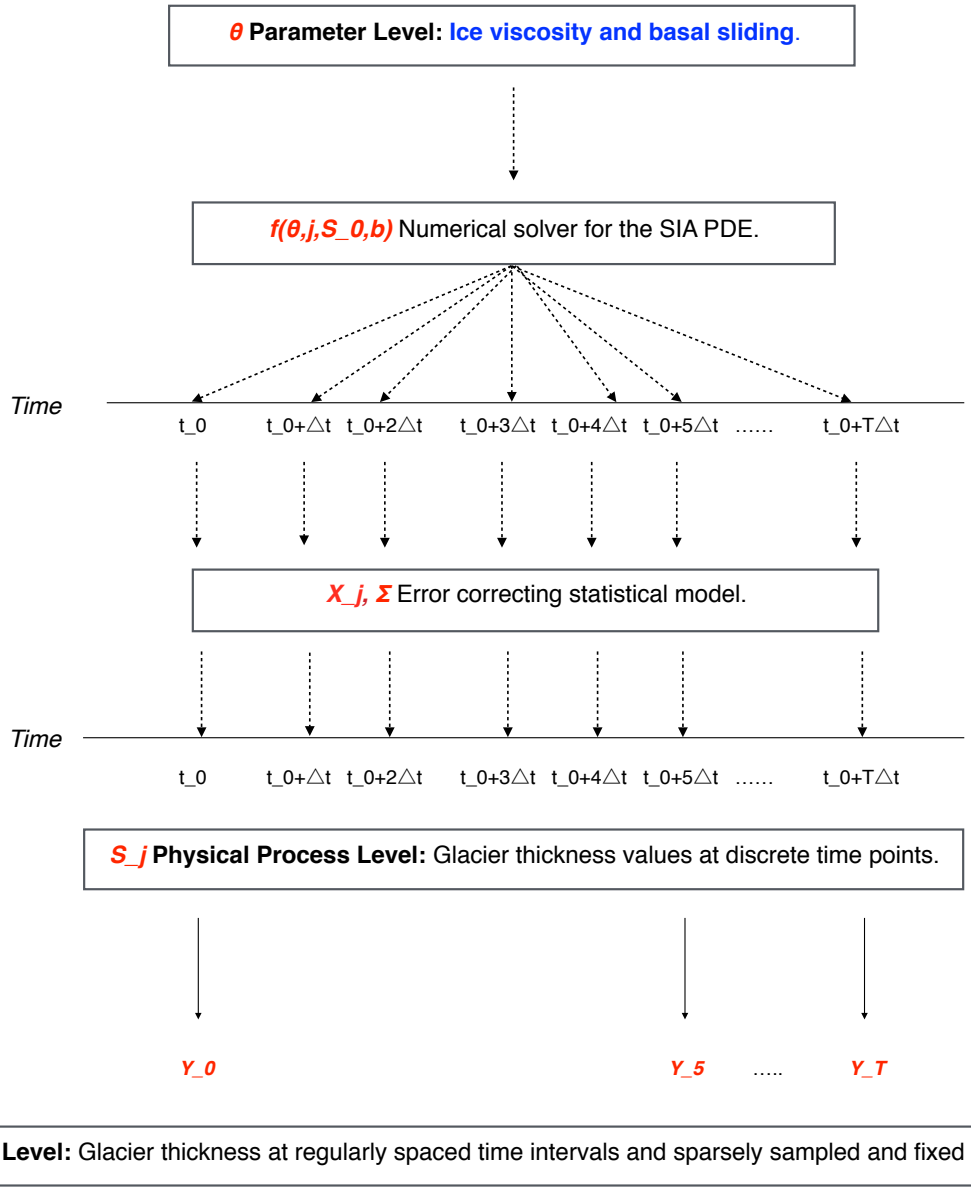

Figure 3. Schematic of the physical-statistical BHM that has been constructed based on the SIA PDE. The main parameters and variables for each module of the physical-statistical model are highlighted in red. **The main levels of a physical-statistical model shown in Figure 2 are displayed here, along with the parameters and variables describing each level.**

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).

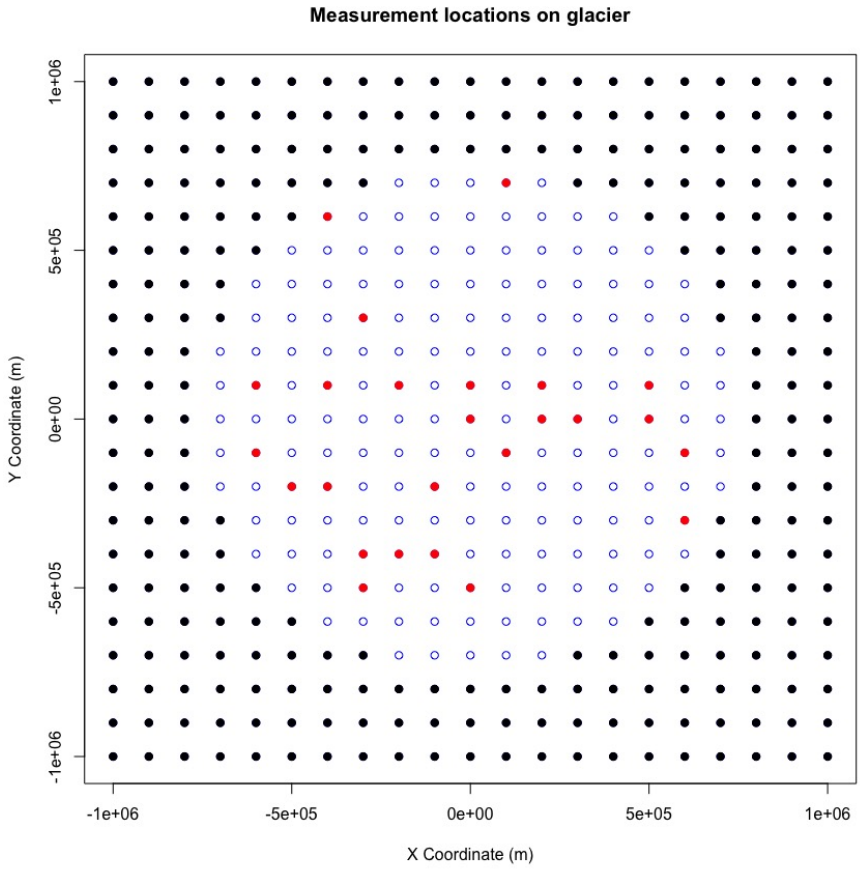

Figure 4. An illustration marking the 25 measurement sites on the glacier. This is a top level view of the glacier, where the blue points indicate the glacier, the red points indicate the measurement locations, and the black points indicate locations surrounding the glacier with no glacial thickness.

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 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 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 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 6). 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 7.

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 8. 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 9. 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 10 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, 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.

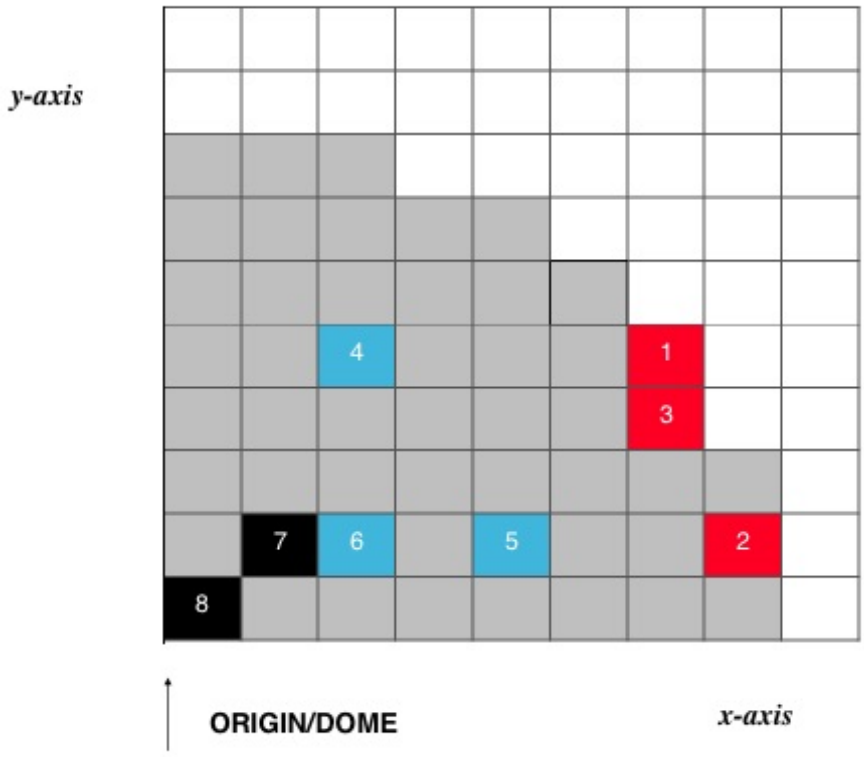

Figure 5. Grid map used to interpret the following box-plots in Figure 6. Eight randomly chosen grid points are selected **for testing predictions; these are not the same as the measurement locations**. Only one quadrant of the glacier is shown due to symmetry as is done in Figures 9,10, and 12 of Bueler et al. (2005), and the width of each cell is 10^5m. 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.

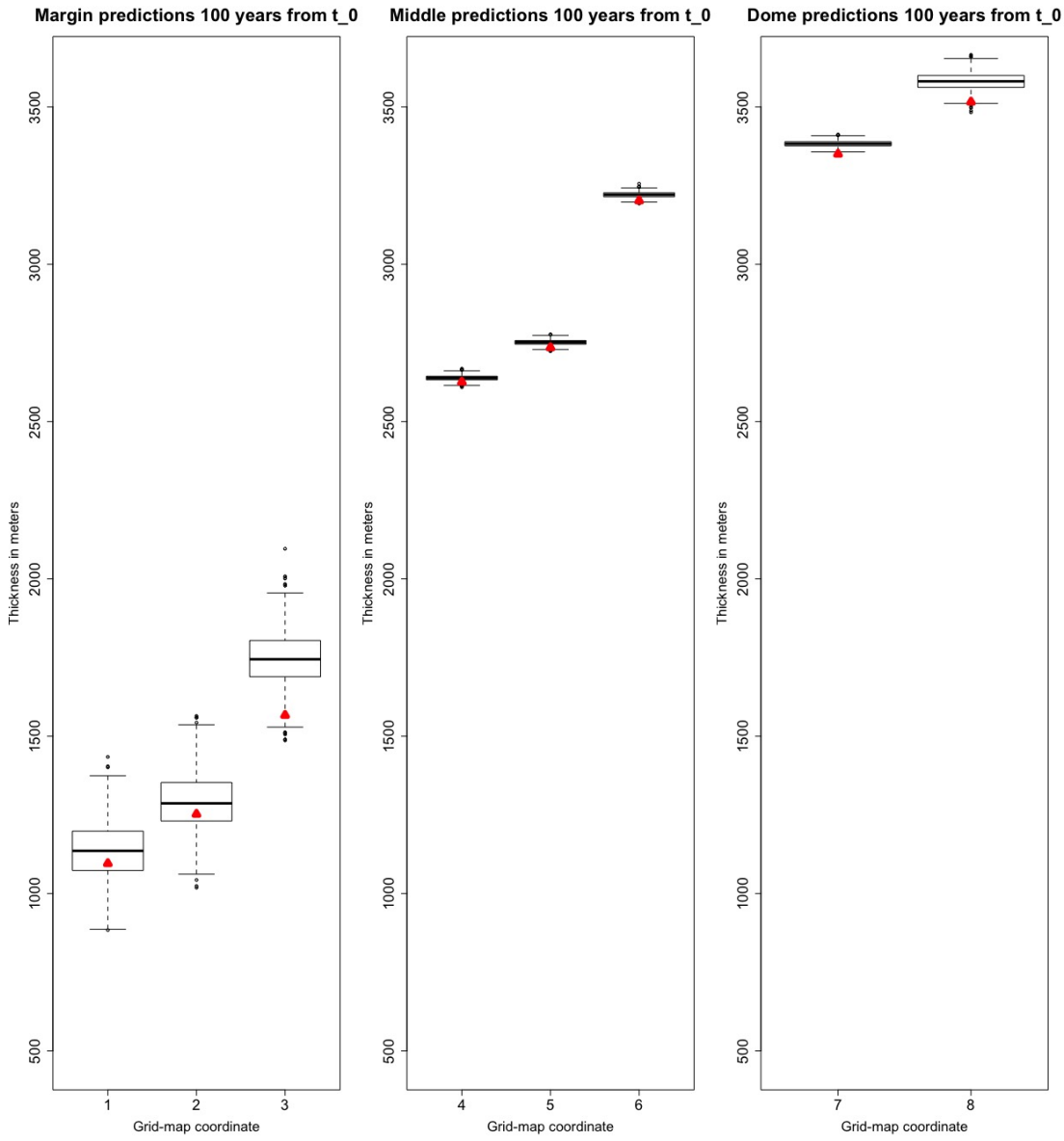

Figure 6. Thickness prediction samples 100 years from t_0 for test case B **(i.e., no mass balance field or basal sliding)**. Triangles indicate the actual thickness values from the analytical solution. The first set of plots are close to the margin (red squares of Figure 5), the second set of plots are between the dome and margin of the glacier (blue squares of Figure 5), and the final set of plots are towards the dome of the glacier (black squares of Figure 5). Refer to Figure 5 for a grid map to spatially reference each of the boxplots. 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.**

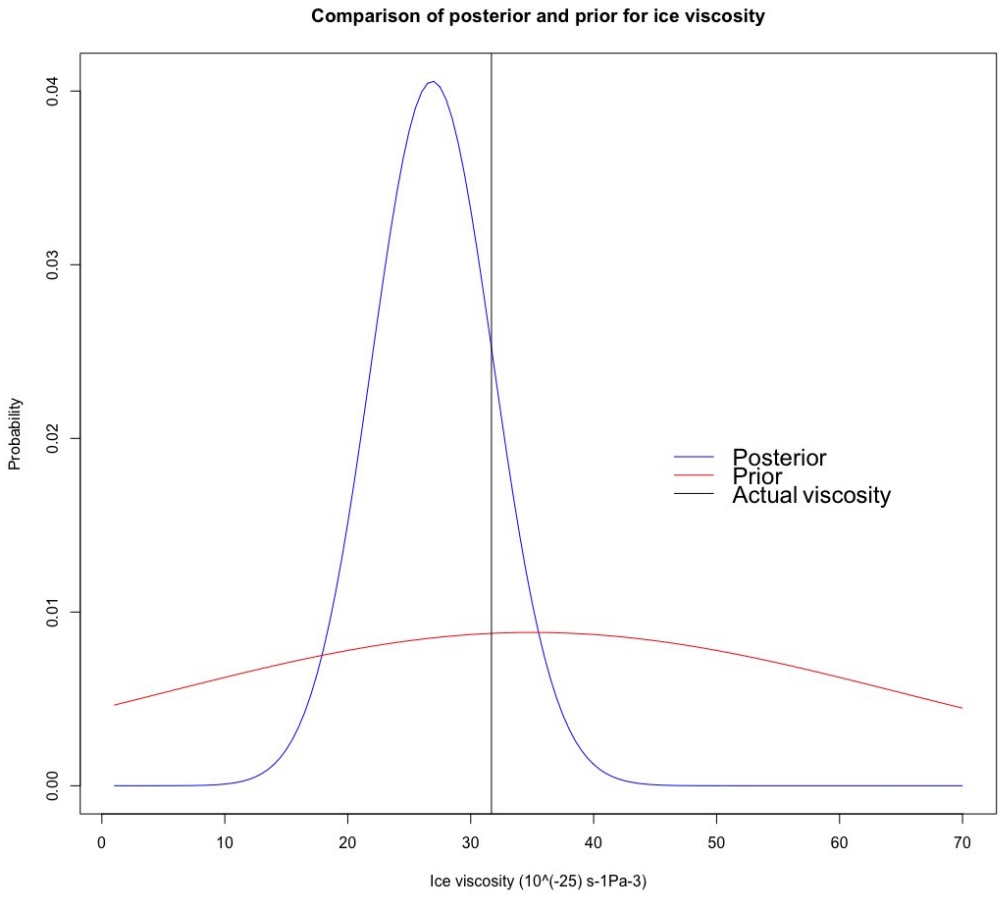

Figure 7. Comparison of posterior and prior distributions of ice viscosity for test case D **(i.e., mass balance field producing a periodic SIA solution).**

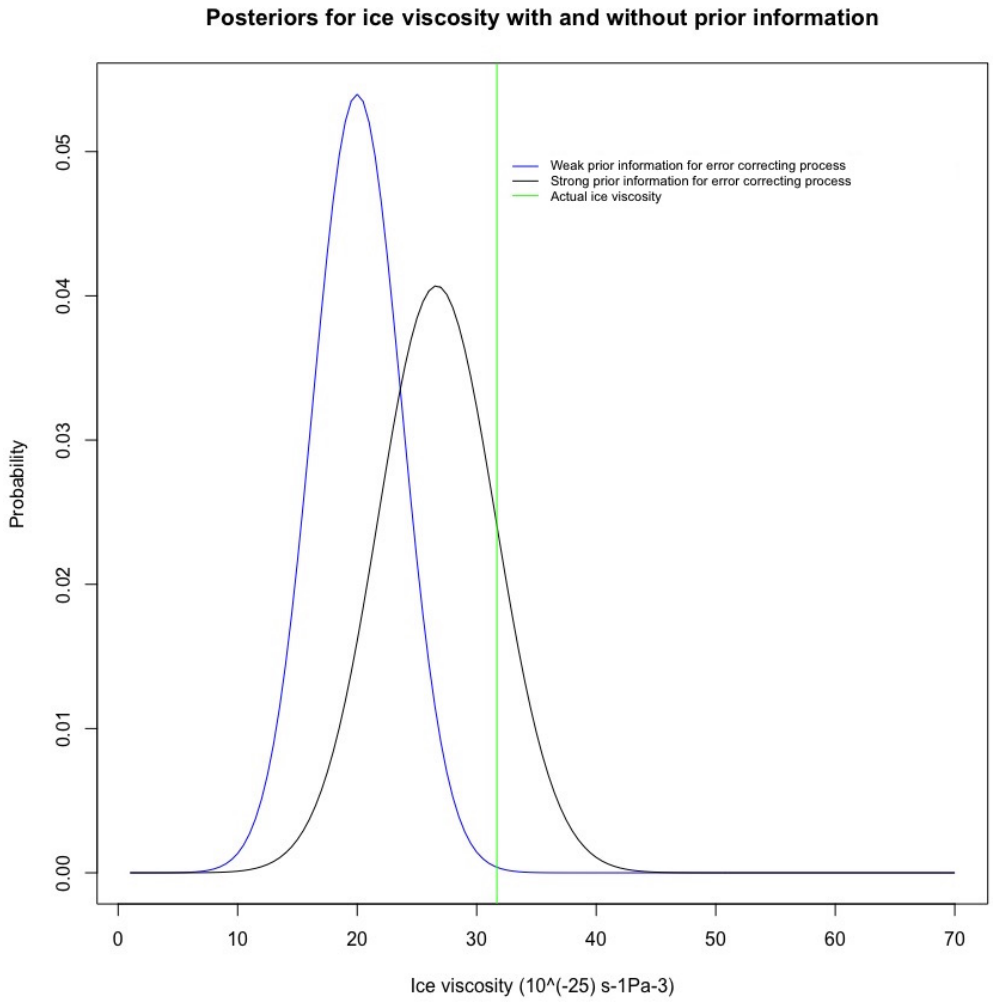

Figure 8. A comparison of posteriors under strong and weak prior information for the error correcting process in test case D (i.e., mass balance field producing a periodic SIA solution); prior information for the error correcting process leads to a less biased posterior, though with slightly more posterior uncertainty.

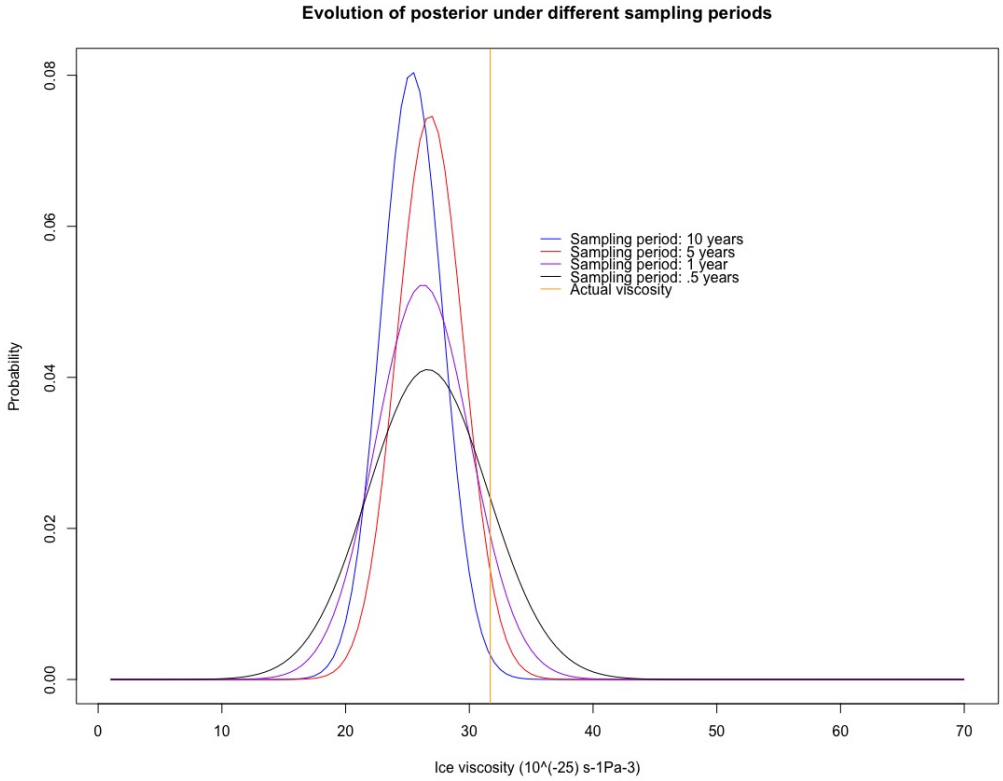

Figure 9. A comparison of posteriors in test case D (i.e., mass balance field producing a periodic SIA solution) under different sampling periods: data sampled once every 10 years, every 5 years, once a year, and twice a year. The general trend is that the posterior tends to become less biased as the 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.

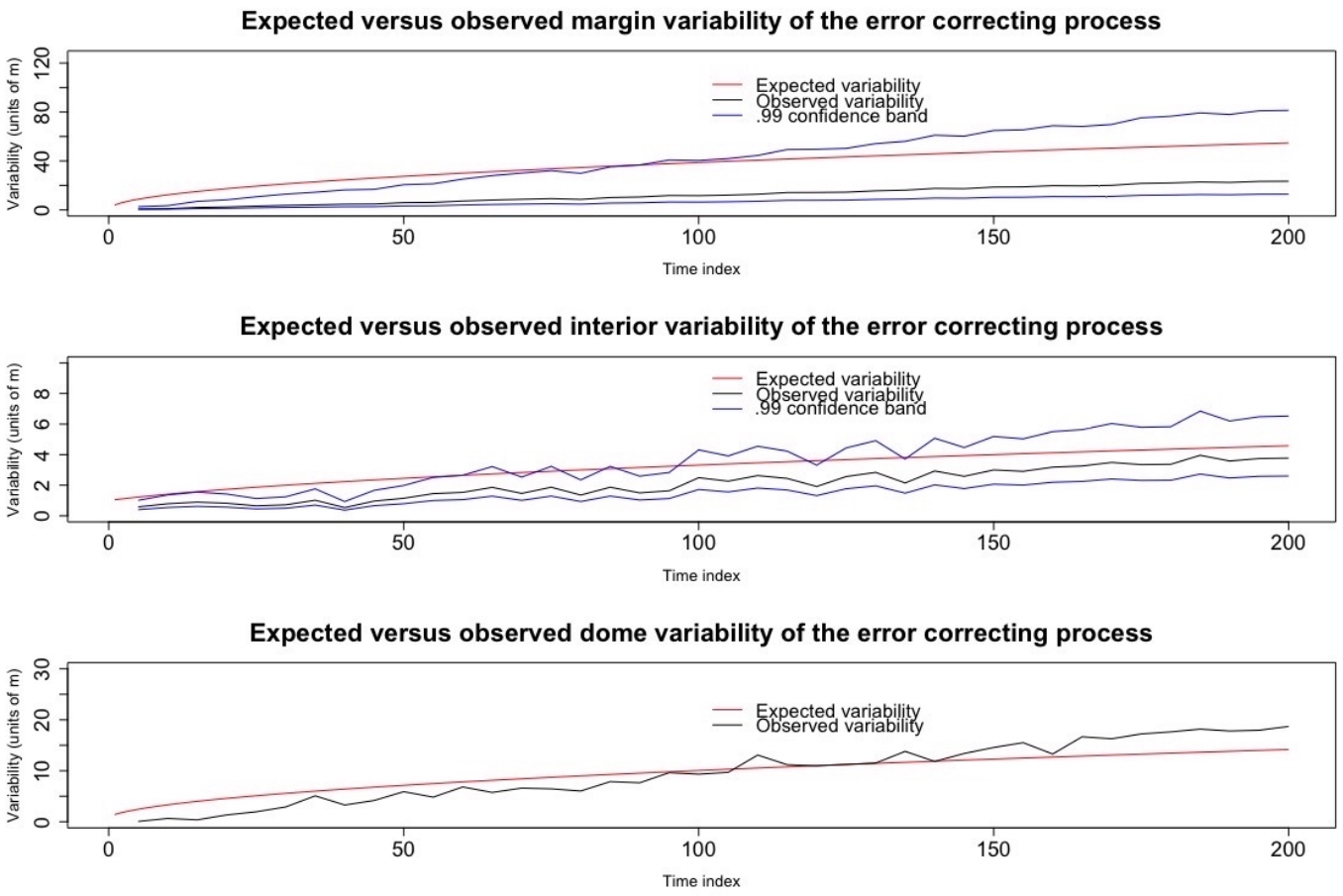

Figure 10. An illustration comparing the expected variability of the error correcting process (as per the Bayesian hierarchical model) to the observed variability of residuals at the interior, margin, and dome for **test case B (i.e., no mass balance field or basal sliding). These residuals are the differences 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 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 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. Error correcting process hyper-parameters; $\sigma^2_{\textbf{dome}}$ **is the error correcting process variance at the dome,** $\sigma^2_{\textbf{interior}}$ **is the error correcting process variance at the interior,** $\sigma^2_{\textbf{margin}}$ **is the error correcting process variance at the margin, and** ϕ **is the length scale parameter.**

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.

Author contributions. All of the glaciologists contributed equally to this work.

Acknowledgements. The Icelandic Research Fund (RANNIS) is thanked for funding this research.

Appendix A: Finite difference method for the shallow ice approximation

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 two 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) 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., $H_{xt} = H_{tx}$).

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

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

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

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

$$
\begin{aligned}
[\alpha^{n-1}]_x &= \frac{n-1}{2}(S_x^2 + S_y^2)^{\frac{n-3}{2}}(2S_xS_{xx} + 2S_yS_{yx}) \\
[\alpha^{n-1}]_y &= \frac{n-1}{2}(S_x^2 + S_y^2)^{\frac{n-3}{2}}(2S_yS_{yy} + 2S_xS_{xy})
\end{aligned}
$$

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

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

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).

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** (i.i.d) 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 is 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 i.i.d $MVN(0, \sigma^2 I)$. 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.

B1 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 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.

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

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.**

B1.1 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 $M \otimes \Sigma + \sigma^2 I$, 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).

B1.2 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:

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

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 covariance 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.

B2 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).

B3 Making spatio-temporal 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 $S_{T_{\text{end}}} = X_T + \sum_{j=T+1}^{T_{\text{end}}} \epsilon_j + f(\theta, T_{\text{end}})$. 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 (2013). 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.

References

Berliner, L. M.: Hierarchical Bayesian Time Series Models, in: Maximum Entropy and Bayesian Methods, edited by Hanson, K. M. and Silver, R. N., pp. 15–22, Springer Netherlands, Dordrecht, 1996.

Berliner, L. M.: Physical-statistical modeling in geophysics, Journal of Geophysical Research: Atmospheres, 108, n/a–n/a, https://doi.org/10.1029/2002JD002865, http://dx.doi.org/10.1029/2002JD002865, 8776, 2003.

Berliner, L. M., Jezek, K., Cressie, N., Kim, Y., Lam, C. Q., and van der Veen, C. J.: Modeling dynamic controls on ice streams: a Bayesian statistical approach, Journal of Glaciology, 54, 705–714, https://doi.org/10.3189/002214308786570917, 2008.

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

Brynjarsdóttir, J. and O'Hagan, A.: Learning about physical parameters: the importance of model discrepancy, Inverse Problems, 30, 114 007, 2014.

Bueler, E., Lingle, C. S., Kallen-Brown, J. A., Covey, D. N., and Bowman, L. N.: Exact solutions and verification of numerical models for isothermal ice sheets, Journal of Glaciology, 51, 291–306, https://doi.org/10.3189/172756505781829449, 2005.

Calderhead, B., Girolami, M., and Lawrence, N. D.: Accelerating Bayesian Inference over Nonlinear Differential Equations with Gaussian Processes, in: Proceedings of the 21st International Conference on Neural Information Processing Systems, NIPS'08, pp. 217–224, Curran Associates Inc., USA, http://dl.acm.org/citation.cfm?id=2981780.2981808, 2008.

Cressie, N. and Wikle, C. K.: Statistics for spatio-temporal data, John Wiley & Sons, 2015.

Cuffey, K. M. and Paterson, W.: The Physics of Glaciers, Academic Press, 4 edn., 2010.

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

Fowler, A. C. and Larson, D. A.: 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, http://www.jstor.org/stable/79748, 1978.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian data analysis, 3rd edition, 2013.

Glen, J.: The flow law of ice: A discussion of the assumptions made in glacier theory, their experimental foundations and consequences, IASH Publ, 47, 171–183, 1958.

Glen, J. W.: The Creep of Polycrystalline Ice, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 228, 519–538, http://www.jstor.org/stable/99642, 1955.

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

Hudson, J.: Numerical Techniques for Conservation Laws with Source Terms, Tech. rep., Engineering and Physical Science Research Council.

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

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

Isaac, T., Petra, N., Stadler, G., and Ghattas, O.: 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, https://doi.org/https://doi.org/10.1016/j.jcp.2015.04.047, http://www.sciencedirect.com/science/article/pii/S0021999115003046, 2015.

Jarosch, A. H., Schoof, C. G., and Anslow, F. S.: Restoring mass conservation to shallow ice flow models over complex terrain, The Cryosphere, 7, 229–240, https://doi.org/10.5194/tc-7-229-2013, https://www.the-cryosphere.net/7/229/2013/, 2013.

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

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

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.: Results from the EISMINT model intercomparison: the effects of thermomechanical coupling, Journal of Glaciology, 46, 227–238, 2000.

Pralong, M. R. and Gudmundsson, G. H.: Bayesian estimation of basal conditions on Rutford Ice Stream, West Antarctica, from surface data, Journal of Glaciology, 57, 315–324, https://doi.org/10.3189/002214311796406004, 2011.

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

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K.: Bayesian Computing with INLA: A Review, Annual Review of Statistics and Its Application, 4, 395–421, https://doi.org/10.1146/annurev-statistics-060116-054045, 2017.

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

van der Veen, C.: Fundamentals of Glaicer Dynamics, CRC Press, 2 edn., 2017.

Wasserman, L.: All of statistics: a concise course in statistical inference, Springer Science & Business Media, 2013.

Weertman, J.: The theory of glacier sliding, Journal of Glaciology, 5, 287–303, https://doi.org/10.1017/S0022143000029038, 1964.

Wikle, C. K.: Hierarchical Models for Uncertainty Quantification: An Overview, pp. 1–26, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-11259-6_4-1, https://doi.org/10.1007/978-3-319-11259-6_4-1, 2016.

Supplementary note: an example illustrating the approximation used for the likelihood

An example motivating this approximation

To motivate the approximation in the likelihood (section B1.2), we consider a simple, univariate example where the output of the numerical solver is 1000 for three consecutive time steps, and both the measurement error variance and numerical error variances are 1 – note that the magnitude of the numerical solver is a few orders of magnitude larger than the measurement error. That is, we have:

$$\begin{aligned} Y_1 &= 1000 + X_1 + Z_1 \\ Y_2 &= 1000 + X_1 + \epsilon_1 + Z_2 \\ Y_3 &= 1000 + X_1 + \epsilon_1 + \epsilon_2 + Z_3 \end{aligned}$$

Where Z_1, Z_2, Z_3, X_1, ϵ_1 ,and ϵ_2 are all identically and independently distributed $N(0,1)$ (normal with 0 mean and unit variance) random variables. Analytically, the conditional distribution of $p(Y_3|Y_2,Y_1)$ follows a normal distribution with mean $1000+.2(Y_1-1000)+.6(Y_2-1000)$ and variance $13/5$. In our approximation, we substitute this distribution with a $N(Y_2,3)$; to show that these distributions are indeed quite close to each other, we conduct 25 simulations and illustrate $P(Y_3|Y_2,Y_1)$ in Figure 1. These results motivate the use of the approximations in section B1.2.

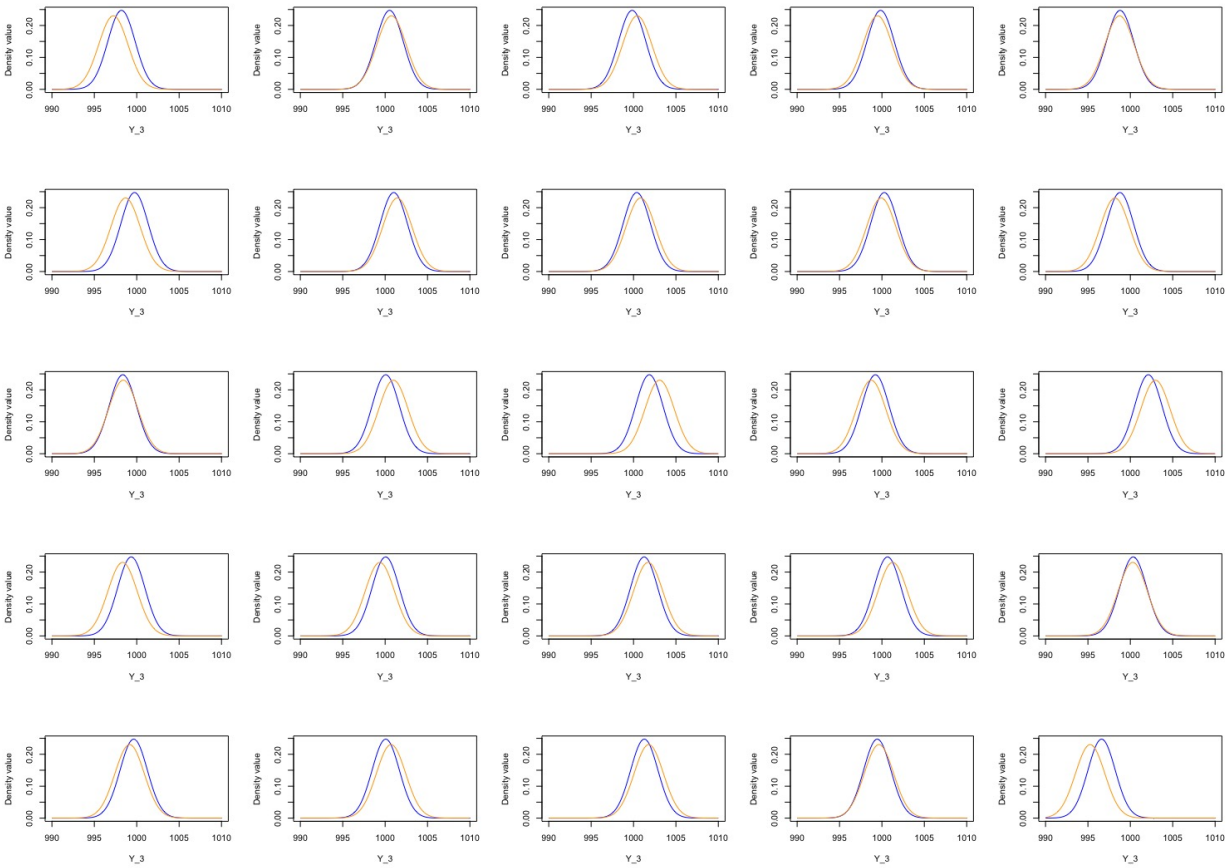

Figure 1: 25 simulations from the above model with $P(Y_3|Y_2, Y_1)$ in blue and $N(Y_2, 3)$ in orange. Visual inspection of these densities in all of the simulations shows that they are close to each other, lending evidence that it is appropriate to use our approximation in the regime where the output of the numerical solver is much larger than measurement errors.