Response to Anonymous reviewers for the manuscript: A framework for time-dependent Ice Sheet Uncertainty Quantification, applied to three West Antarctic ice streams by Beatriz Recinos et al.

Dear Editor and Anonymous reviewers,

Thank you for taking the time to review and improve our manuscript. We believe that we have addressed all points raised by reviewers and hereby submit the revised version together with a point by point reply to each of the reviewers comments. For minor comments and corrections please referred to changes highlighted in blue, in the diff.pdf file attached to this reply.

Below reviewer's comments are given in italics and our answer in normal font.

Reply to Anonymous Referee # 1

RC: My main comment is that the conclusion indicates that the regularisation weights suggested by the L-Curve analysis seem to lead to priors that are too confident, suppressing the propagation of the uncertainty from the velocity data-sets used for the calibration. However, I found that the method for the L-Curve is not very well described as there is 4 parameters to calibrate, and it is not to clear if they are chosen independently?, and there is a high level of user-judgement in the selection of these parameters; Comparing the values given in section 4.1 to those used in Table 1, it appears that the main differences are on the δ parameters for which the results are not shown. I am also wondering part of the issue cannot come from wrong priors as they are particluarly poorly constrained and here, the prior for the friction parameter α is 0, so that pure sliding everywhere? So maybe the conclusion could be revisited a little to not put too much attention on the L-Curve?.

AR: Both L-curves γ_α and γ_β in sect. 4.1 are computed independently from each other - i.e. varying one parameter over several orders of magnitude while the other parameter remains fixed, following control-method applications found in the ice-sheet modelling literature - e.g Jay-Allemand et al. (2011); Gillet-Chaulet et al. (2012); Seddik et al. (2017); Barnes et al. (2021). We agree with reviewer's #1 comment; in our manuscript and in the literature there is a high level of user judgement and a large variability in the application of the L-curve criterion among ice-sheet modelling studies. Our aim in this section is not to refine or improve upon the application of the L-curve methodology in glaciological inversions but simply to apply it to a standard defined by the glacial literature. We discuss our approach in terms of (i) user-judgement; (ii) calibrating multiple parameters; (iii) consideration of our "variance cost" parameters δ_{alpha} and δ_{beta}; and (iv) choice of prior mean below.

The L-curve criterion used in our methods and in previous studies (listed above) is based on Hansen (1992, 2001), where L-curves are defined as a log-log of the norm of a regularised solution versus corresponding residual norm. The correct regularisation term is chosen by locating the "corner" of the L-curve, which represents an optimal trade-off between fit to data and smoothness of solution. Hansen (1992, 2001) identify the "corner" as the point on the L-curve with maximum curvature. However, as pointed out by these papers, accurate calculation of curvature as a function of the regularisation parameter is challenging and requires dense sampling of the L-curve which could be computationally expensive. Preferably, the finding of the "corner" value is often chosen heuristically. For example, some studies choose the point where the cost function is at its minimum (Gillet-Chaulet et al., 2012) whereas other studies choose points just before the curvature of their L-curve analysis (Seddik et al., 2017). Others
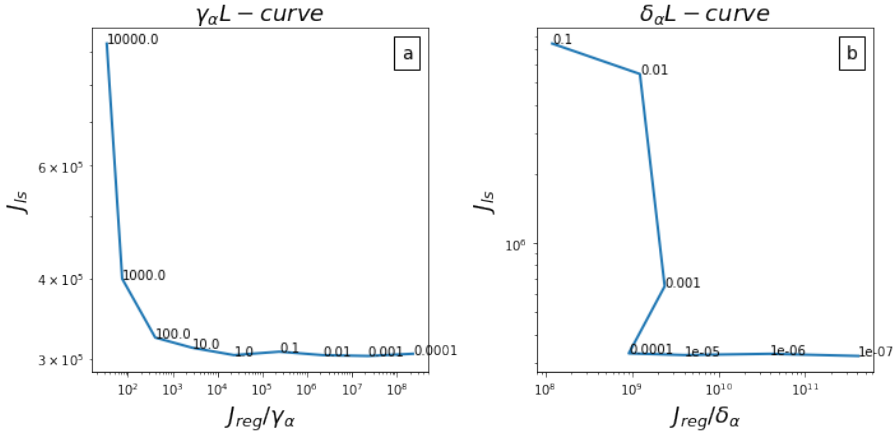
Figure 1: *L*-curve analysis for **a:** γ_α and **b:** δ_α.

emphasise the goal of finding the "best trade-off", aiming to pick parameters whose values lie near the corner of the L, where neither J_{ls} or J_{reg} take high values (Barnes et al., 2021). Therefore we made our choices in accordance with this loosely defined criterion and decided on $\gamma_\alpha = 100.0$, based on a visual assessment of the corner in a log-log scale plot. In theory, any point between 1000 - 1.0 could be a potential value for γ_α (see Fig. 1), though we know from our physically informed prior that it should be in the order of 1. However, this is not evident if we follow the method of finding the "corner" of the *L*-curve as suggested in the literature.

The classic *L*-curve approach is prescribed for a single regularisation parameter, whereas large-scale glaciological data assimilation often involves multiple parameters, leaving some ambiguity of how to proceed. While some studies consider two-dimensional "*L*-surfaces" in order to jointly optimise parameters (e.g., Fürst et al., 2015; Goldberg et al., 2019), such an approach in our case would have required a prohibitively expensive sampling of four-dimensional space and complicated visualisations. Rather, we followed the approach of Barnes et al. (2021), in which one parameter is varied while others are held constant.

For these reasons and in line with common practices among the ice sheet modelling community, we do not make significant changes in sect. 4.1 or our conclusions but clarify the methods that we follow in constructing the *L*-curves of sect. 4.1 (see L390-403 of attached diff report), as our main point is to highlight that *L*-curve analysis leaves room for user interpretation and is not based on existing prior knowledge and physical concepts that define each control parameter. The variance σ_{c0}^2 and auto-covariance length scale l_{c0} of each control parameter still provides a more physically informed prior and allows for the study of the interaction between all the poorly constraint parameters.

Additionally to the experiments shown in Sect. 4.1, we constructed a *L*-curve **only for** δ_α, as in Sect. 5.1 we show that only a strong prior on α will influence the posterior uncertainty of VAF. We have added this analysis here and to appendix of the manuscript (see Appendix A2, L770-772 of the attached diff report). Fig. 1b suggests a value of $\delta_\alpha = 1 \times 10^{-4}$ (we use $\delta_{\alpha,\beta} = 1 \times 10^{-5}$ in the analysis of Sect. 4.1). We know from the δ_α values shown in Table 1 that any $\delta_\alpha \geqslant 1 \times 10^{-7}$ will result in too strong of a prior. Therefore it is unlikely that any value for δ_α (or δ_β) based on an *L*-curve analysis would be appropriate for uncertainty quantification. As this does not impact our results or overall conclusions, we do not investigate this further but show the *L*-curve for δ_α in the appendix.

Regarding the prior mean of $\alpha_0 = 0$, we chose zero as α can be positive or negative therefore,

a zero mean seems appropriate. Furthermore, any guess not based on sub-glacial exploration would implicitly involve physics of the model.

RC: L124 'constant surface mass balance"; Is it constant and uniform; or is there spatial variability?
AR: Constant means uniform and constant in time. There is no spatial or temporal variability. We have clarified that in the text (see L137 of the attached diff report).

RC: EQ6. Q_T here is defined as the VAF while is it use as the difference of VAF from t=0 in the manuscript. What is the meaning of the "+" symbol?
We have modified this equation to state that Q_T is equal to VAF and time T, and modified the captions and y-axis of Figures 5a and 10d to reflect that we plot trajectories of change in VAF ($Q_T - Q_0$) over time. The "+" refers to the positive part of the volume that contributes to global mean sea level. We have added this explanation, see L156 and Figures 5 and 10 of the attached diff report.

RC: L142 $Hf = max(0, -R(\rho_w/\rho_i))$
AR: Corrected, see L148.

RC: L196 if the prior is strong, γ is "large" not "small"? (in agreement to line 342-check for consistency everywhere)
AR: Corrected, see L155 of the attached diff report. This was the only inconsistency found.

RC: Sec 4.1 would be interesting to discuss the smoothing parameters in terms of variance and correlation length scales (Eq. 13-14) as it appears that the parameters used here lead to a very small variance compared to the values used in Table 1.
AR: The variance σ_{c0}^2 and auto-covariance length scale l_{c0} has been added to L-curve derived priors in Table A1, see page 47 of the attached diff report.

RC: L389 J^c should be J_{mis}^c? Check for consistency everywhere. I don't understand why it does not change with the number of observations as according to Eq.8 it should depend on the number of observations?
AR: Indeed there was a mistake in the caption and y-axis of Fig. 7, it should have been J_{mis}^c. This has been corrected in Figure 7 in the text (See Fig 7, L445, L457, L461 of the attached diff report). In this experiment and in the results of J_{mis}^c plotted in Figure 7, we do several inversions (with the same priors); at each inversion we retained different % of data points (or observations). For every % of data points retrained, we invert for both α and β and evaluate J_{mis}^c. We do this to find out, if by dropping observations our ability to reproduce the observed velocity decreases. A point of confusion may be the data used to evaluate J_{mis}^c – this was distinct from the inversion constraints, and consistent across the experiment, which we now make clear (see L445 and L457 or diff report).

RC: L477 "as the basal stress does not scale with effective stress in the interior". I don't understand the argument here.
AR: We apologise for not being more clear here. The Cornford sliding law has the approximate form

$$\tau_{\mathbf{b}} \approx \alpha^2 u^{-2/3}\mathbf{u} \tag{1}$$

in the interior i.e. where N is large. It can be seen that this differs from the Weertman-Budd sliding law by a factor of $N^{1/3}$, and thus we expect that the inverse solution for α with a Cornford sliding law should be larger here by approximately this factor. Any prior covariance, therefore, should reflect this change in scale, and hence $\sigma_{\alpha(0)}$ was made larger to reflect this. We added the Cornford sliding law prior configuration to Table 1 and now state (see L545-550 of the attached diff report):

"The prior distribution used is similar to the highlighted parameters in Table 1, but with a modified $\sigma^2_{\alpha(0)}$ – since in the interior the basal stress is independent of the effective stress (Cornford et al., 2020), and thus we expect variations of α to have a different scale (see last configuration in Table 1)."

RC: L486 "is due to insensitivity of basal stress to α when the ice is near floatation". The Weertman-Budd relation Eq. 1 is also insensitive to α near flotation as it depends on N; main difference is that Eq.3 tends to a Coulomb regime, independent of α, for high velocity and low effective pressure. However using eq. 12 for N tends to restrict this domain to the close vicinity of the grounding line (Joughin et al., 2019)

AR: We are unsure which equation is referred to by eq. 12 (as this is relating to prior distributions) so we might be misunderstanding, but we feel that we are mostly in agreement with the reviewer. Since Eq. 3 tends to a Coulomb regime near the grounding line, this mutes dependence on α, which is why, after rescaling of colorbars, there is high uncertainty near the Smith grounding line in Fig 10b (Cornford) but not 10a (Weertman-Budd). In the Weertman-Budd sliding law, there is still quadratic independence of basal stress on α; effective stress (N) becomes small, but at the same time ice speed is large, and the result is that α is better constrained (relative to other regions of the domain) than in the Cornford case – indeed, this is what Figs. 10 (a,b) show.

Reply to Anonymous Referee # 2

RC: The 3 sentences (l52-56) are not enough to introduce basic concepts of Bayesian inference to the community, and especially to connect to the ice sheet model present study. To elaborate, please define clearly here what you mean by prior/posterior/covariance, link it directly to glaciological quantities, and give some intuition on the method. Also, it could be better motivated. If I understand, L54-55, propagation of errors between uncertain control parameters, and VAF could be obtained by proceeding to a massive amount of model realization, which is prohibitively expensive due to the costs of Stokes solving, right? This is what motivates you to take another approach? If yes, I suggest to re-structure your paragraph starting from this motivation statement, and then elaborating (substantially) on Bayesian approach, and what this means in the context of your problem.

AR: We agree with the reviewer's comment and have expanded the explanation of the Bayesian inference framework in that paragraph. We have now defined what low/high dimension means (L47-48 and L50-51 of the attached diff report) and each basic concept of Bayesian inference and relate each concept to variables in our study (L52-68 of the attached diff report).

RC: Despite several passes, Section 2.4.1 and 2.5 remains unclear to me, probably because I have no prior experience in Bayesian approaches, and I have not looked at the references. Here, I would expect to at least get a rough idea from these sections without having to go to references. E.g. where do the finite element matrices use to define Γ_{prior} come from? What is the role of the operator (11) in the story? Justifications and explanations would be very welcome to explain all equations given in 2.4.1. As this is central in the paper, this part must be self-explained (i.e. referencing if not enough). Similarly, Eq. (16) and (17) are highly important, but under-explained, please elaborate, give some intuition, and connect to what this means in the context of your glaciological problem. Several sentences could be founded an other articles on using Bayesian approach for a completely different problem. Therefore, there is room to better connect the approach and the application.

AR: We apologise for any lack of clarity in these sections. Indeed our aim was to place fo-

cus in this paper on the experimental results, and avoid a lengthy coverage of the underlying mathematical and numerical framework, which was the reason we only stated key results from previous papers. But we agree that it may still be confusing to readers not familiar with the literature. We have now made extensive changes to these sections in order to provide clarity, and to attempt to introduce concepts in an order that is not too abrupt and have a better progression. We have also added text to give better intuition for the expressions provided. Changes/additions are as follows:

- We do not introduce concepts of variance and autocovariance of the prior until after the discussion of bayesian inversion, as we feel this might have contributed to the difficulty of 2.4.1. The previous section 2.4.1 is now gone, and we discuss the deterministic form of the regularisation cost in 2.4, mentioning only that the name "prior" is due to its Bayesian interpretation. We now introduce a separate subsection (2.6) in which we discuss the statistical properties of the covariance matrix and how they relate to regularisation parameters – we do this in a separate section because of the relevance of these properties to our investigation, and to avoid introducing statistical concepts ahead of discussing our Bayesian interpretation of the cost function.

- We now introduce Bayes' theorem, and how it relates to our cost function, at the beginning of 2.5 (L235-252 of the attached diff report). While this adds additional text and an equation, we feel it will be helpful for those less familiar with Bayesian concepts, and that it is introduced in an informative way.

- We now give more text explaining the meaning and implications of eq 15 (formerly eq 16), and give an example of how these implications might manifest in the context of ice-sheet inversions (L256-263 of the attached diff report).

- We added more text explaining the importance of Eq 16 (formerly 17) (L267-269 of the attached diff report), and directly reference its sources as it is a nontrivial result.

Eq 12 (previously eq 11) fits into the story as follows: If the linear helmholtz equation $\gamma\nabla^2 y - \delta y = F$ is solved for y via finite elements, it would result in a linear system of equations $\mathbf{L}\overline{y} = \overline{f}$, where $\overline{y}$ are the nodal values of the solution and $\mathbf{L}$ is the **stiffness** matrix. We now refer to $\mathbf{L}$ as the "stiffness" matrix, which we feel is sufficient (see L207 of the attached diff report).

RC: Following my last point, several times in the paper, one refers to "priors" or "posteriors" in a generic way, without specifying the meaning (regularization strength). E.g. a number of sentences are general statements with Bayesian vocabulary and unspecific to the ice flow problem considered here, and this contributes to making the paper abstract for non-specialists. Efforts are required to make the paper further "educational", and the choice of words really matters in that respect..

AR: We added a few clarifications of prior and posterior definitions and relate those definitions to the control parameters that we study. However, due to the length of our manuscript we limit these explanations to the introduction, methods and conclusions. See L52-86, Sections 2.4-2.6 and L734-735 of the attached diff report.

RC: The large amounts of data points in TS-inferred velocity max be redundant, with implications for error propagation of VAF: This is not surprising. RS products may be very dense as the efficiency of feature-tracking algorithms has improved. Therefore, trying to fit densely covered (poss. noisy) observation fields is probably more difficult than if we were selecting only a sparse version of the data, with the result of relaxing / giving room to the optimization. This is an interesting outcome, however, I find it a bit distracting to find this technical point coming

back several times in other sensitivity experiments. Why not simply taking 1.6% of the data in all the paper explaining this choice somewhere. I don't feel this is sufficiently important finding to be part of the abstract.

AR: Agreed. This has been deleted form the abstract and only mentioned in Sect 4.3 and conclusions.

RC: I'm not sure to understand when you say that the regularization is too strong (L506-512): Would you have expected the VAF error (propagated) due to regularization higher than the one induced by different observed velocity products? What are the implications, and your recommendations for regularizing in future studies?

AR: An overly informative prior or a strong prior (strong regularisation) means that the information contained in the prior distribution of a control parameter (e.g. prior point-wise STD of the ice stiffness parameter) dominates the information contained in the ice velocity observations being analysed, and hence our error propagation framework underestimate the variability in SLR projections. We find the variability in SLR projections when we compute VAF trajectories using different satellite velocity products to run our time-dependent ice sheet model. By choosing different satellite products our model leads to different estimates of VAF after 40 years. We use this difference in model output to quantify the variance that projections of VAF are expected to have after 40 years and identify prior strengths (regularisation strengths) that can reproduce that variability. We find prior strengths which are weak enough that the variability seeing in the satellite velocity observations can be propagated to VAF projections. Our reasoning is now made more explicit in the discussion (see L580-587 of the attached diff report.)

Therefore, instead of using L-curve inform prior's, we recommend to use the variance and length scale arising from a physical interpretation of the prior to define the regularisation parameters or prior covariance; as these definitions (see Sect. 2.6) will inform the ice sheet model with a more realistic spatial variability regarding the basal sliding and ice stiffness parameters. Moreover, this way of computing the prior covariance will allow our simulations to study the interaction among all regularisation parameters, which is not possible via L-curve analysis (see reply to Reviewer #1). This explanation has been added to the Conclusions, see L737-740 of the attached diff report.

RC: Have you tried to include ice thickness as part of the control parameters? or is this not justified in the special case of ice streams?

AR: We take the ice thickness distribution from BedMachine (V.2.0 Morlighem et al., 2020) and assumed that the errors in this dataset play no role in our calibration uncertainty, which is a common approach in current assimilations of ice-sheet velocities. Early approaches (e.g. Macayeal et al, 1995) considered bed and surface elevation as a control parameter (though not ice stiffness) with the view that the incredibly coarse DEMs available at the time would incorrectly attribute velocity variations to basal stress, but there have been significant advances in altimetry since. We now add to the end of Section 2.4 (**Cost Function L228-233** of the attached diff report):

"Previous assimilations of satellite velocities also considered ice bed and surface elevations as control parameters (MacAyeal et al., 1995), because available elevation products did not capture the small-scale features driving variations in velocity. We consider this to be less of an issue with the elevation products currently available, though future studies with our framework could consider topographic uncertainty and how it covaries with uncertainty of other parameters."

RC: L73: "overly informative prior" is an example of Bayesian wording for which I have no intuition. Please try to in other words, or better connect with glaciological words
AR: See added explanation on L85-86 of the attached diff report and our reply to previous items.

RC: Can you explain what you mean by "low-dimension", "high-dimensional" or "low-rank" in several places in the text (l47, 226)
AR: We use the terms low and high dimension to refer to the number of elements in a finite element mesh, or the number of unknowns in an inverse problem. In other words, how many dimensions are needed to represent the hidden parameter field $C(\alpha, \beta)$ in our domain.
The sliding parameter α and ice stiffness parameter β in these other studies are taken as scalar and global (same at every point in the mesh), in our study these parameters scale with the size of our mesh. – i.e. we calibrate each parameter for every element or point in our mesh (see Fig. 1). We added this explenation to the introduction L47-48, L50-51 of the attached diff report.
A low-rank approximation refers to the approximation of the Hessian matrix using a rank of r where r is smaller than the number of rows and columns of the matrix. We only compute 10^{4th} out of 10^{5th} eigenvectors and eigenvalues (r) of the Hessian matrix thus we do not construct the full matrix but an approximation of this matrix, which is usually referred to as a "low rank" approximation in mathematics. We acknowledge that not everyone is aware of this terminology depending on their field. However, we feel that these are basic mathematical concepts available in algebra books as well as easier searches on-line. Therefore, we do not add an explanation of the term "low-rank" to our manuscript, but we modified our manuscript by relating these concepts to parameters and unknowns in our model. See L47-66, L267-269 of the attached diff report.

RC: The norm $\| \cdot \|_{\Gamma_{obs}}$ is not defined here, I understood later than Γ_{obs} are STD weighting the field in the norm computation, but I don't think this is clearly said, or defined at this point.
AR: We have defined the $\| \cdot \|_{\Gamma_{obs}}$, see Line 189. However we respectfully disagree with the comments regarding the definition of Γ_{obs}, which was defined already in Section 2.4, L191-192 of the attached diff report.

RC: The method section is not well structured (e.g. the section "notation", 1 subsection 2.4.1).
AR: We agree that it is odd to have only one subsection within the main section. We have reorganised Section 2.4 (as well as 2.5) in response to your comments above, and 2.4.1 no longer exists.

RC: L502 "different estimates" : please quantify it in percentage.
We do not feel this is appropriate for the discussion, but now give this percentage where the output is first introduced in the Result section (see Section 4.2 of the attached diff report).

RC: In general, one refrains from starting sentences with mathematical symbols and the paper introduces an impressive number of symbols without any reason, also it would greatly help the reader not to refer to symbols (as it requires the reader to memorize it), but instead to its meaning.
AR: We respectfully disagree. As our manuscript is already long, using symbols allows us to say more with less wording. We also provide a notation section to help the reader with the symbols used in the manuscript, and relate symbols to concepts through out the introduction and methods section of the manuscript. We only use symbols in the results and discussion sections in order to summarised our findings.

References

Barnes, J. M., Dias dos Santos, T., Goldberg, D., Gudmundsson, G. H., Morlighem, M., and De Rydt, J. (2021). The transferability of adjoint inversion products between different ice flow models. *The Cryosphere*, 15(4):1975–2000.

Cornford, S. L., Seroussi, H., Asay-Davis, X. S., Gudmundsson, G. H., Arthern, R., Borstad, C., Christmann, J., Dias dos Santos, T., Feldmann, J., Goldberg, D., et al. (2020). Results of the third Marine Ice Sheet Model Intercomparison Project (MISMIP+). *The Cryosphere*, (7):2283–2301.

Fürst, J., Durand, G., Gillet-Chaulet, F., Merino, N., Tavard, L., Mouginot, J., Gourmelen, N., and Gagliardini, O. (2015). Assimilation of antarctic velocity observations provides evidence for uncharted pinning points. *The Cryosphere*, 9(4):1427–1443.

Gillet-Chaulet, F., Gagliardini, O., Seddik, H., Nodet, M., Durand, G., Ritz, C., Zwinger, T., Greve, R., and Vaughan, D. G. (2012). Greenland ice sheet contribution to sea-level rise from a new-generation ice-sheet model. *The Cryosphere*, 6(6):1561–1576.

Goldberg, D. N., Gourmelen, N., Kimura, S., Millan, R., and Snow, K. (2019). How accurately should we model ice shelf melt rates? *Geophysical Research Letters*, 46(1):189–199.

Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34(4):561–580.

Hansen, P. C. (2001). *The L-Curve and Its Use in the Numerical Treatment of Inverse Problems*, volume 4, pages 119–142.

Jay-Allemand, M., Gillet-Chaulet, F., Gagliardini, O., and Nodet, M. (2011). Investigating changes in basal conditions of variegated glacier prior to and during its 1982–1983 surge. *The Cryosphere*, 5(3):659–672.

MacAyeal, D. R., Bindschadler, R. A., and Scambos, T. A. (1995). Basal friction of ice stream e, west antarctica. *Journal of Glaciology*, 41(138):247–262.

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., Goel, V., Greenbaum, J. S., Gudmundsson, H., Guo, J., Helm, V., Hofstede, C., Howat, I., Humbert, A., Jokat, W., Karlsson, N. B., Lee, W. S., Matsuoka, K., Millan, R., Mouginot, J., Paden, J., Pattyn, F., Roberts, J., Rosier, S., Ruppel, A., Seroussi, H., Smith, E. C., Steinhage, D., Sun, B., Broeke, M. R. v. d., Ommen, T. D. v., Wessem, M. v., and Young, D. A. (2020). Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the antarctic ice sheet. *Nature Geoscience*, 13(2):132–137.

Seddik, H., Greve, R., Zwinger, T., and Sugiyama, S. (2017). Regional modeling of the shirase drainage basin, east antarctica: full stokes vs. shallow ice dynamics. *The Cryosphere*, 11(5):2213–2229.

A framework for time-dependent Ice Sheet Uncertainty Quantification, applied to three West Antarctic ice streams

Beatriz Recinos[1], Daniel Goldberg[1], James R. Maddison[2], and Joe Todd[1]

[1]School of GeoSciences, The University of Edinburgh, UK.
[2]School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh, UK.

Correspondence: B. Recinos (beatriz.recinos@ed.ac.uk)

Abstract. Ice sheet models are the main tool to generate forecasts of ice sheet mass loss; a significant contributor to sea-level rise, thus knowing the likelihood of such projections is of critical societal importance. However, to capture the complete range of possible projections of mass loss, ice sheet models need efficient methods to quantify the forecast uncertainty. Uncertainties originate from the model structure, from the climate and ocean forcing used to run the model and from model calibration. Here we quantify the latter, applying an error propagation framework to a realistic setting in West Antarctica. As in many other ice-sheet modelling studies we use a control method to calibrate grid-scale flow parameters (parameters describing the basal drag and ice stiffness) with remotely-sensed observations. Yet our framework augments the control method with a Hessian-based Bayesian approach that estimates the posterior covariance of the inverted parameters. This enables us to quantify the impact of the calibration uncertainty on forecasts of sea-level rise contribution or volume above flotation (VAF), due to the choice of different regularisation strengths (prior strengths), sliding laws and velocity inputs. We find that by choosing different satellite ice velocity products our model leads to different estimates of VAF after 40 years. We use this difference in model output to quantify the variance that projections of VAF are expected to have after 40 years and identify prior strengths that can reproduce that variability. We demonstrate that if we use prior strengths suggested by L-curve analysis, as is typically done in ice-sheet calibration studies, our uncertainty quantification is not able to reproduce that same variability. The regularisation suggested by the L-curves is too strong and thus propagating the observational error through to VAF uncertainties under this choice of prior leads to errors that are smaller than those suggested by our 2-member "sample" of observed velocity fields. ~~Additionally, our experiments suggest that large amounts of data points may be redundant, with implications for the error propagation of VAF.~~

1 Introduction

Ice sheet models are important tools not only for generating knowledge, but also for operational forecasts. In this way, they are analogous to weather models and oceanographic models and have emerged as the de facto standard for generating projections of ice sheet contribution to sea-level rise. However, quantifying the uncertainty in forecasts produced by these models ~~remain~~ remains one of the most challenging goals of scientific inquiry (Aschwanden et al., 2021). Here, we seek to characterise the uncertainty in model projections of marine ice sheet loss which arises from calibration with data.

The paradigm of ice-sheet projection is the calibration of the model parameters with observations (via control methods e.g.
Macayeal, 1992) followed by running of the calibrated model forward in time forced by future ocean and climate scenarios.
The process is uncertain due to (i) model/structural uncertainty (i.e. uncertainty in the formulation of the model and its ability to
represent the physics of the system), (ii) uncertainty in external forcing (e.g. ocean melting of ice shelves), and (iii) calibration
uncertainty (i.e. the uncertainty in calibrated parameters, sometimes referred to as parametric uncertainty). In this study we
use control methods and a Bayesian inference approach to characterise (iii). The Bayesian framework computes posterior
information given the assumed model and external forcing. We do not attempt to quantify (i) and (ii) but we discuss how these
uncertainties can be quantified and incorporated into our error propagation framework.

The use of control methods ("inverse methods") in ice-sheet modelling dates back to Macayeal (1992). Since then, their use
in estimating basal and internal conditions (hidden properties) of glaciers and ice sheets from measured surface velocities has
become widespread (e.g. Sergienko et al., 2008; Morlighem et al., 2010; Cornford et al., 2015; Hill et al., 2021, to name a few).
This is mostly due to the ability of these methods to perform large-scale inversions via the minimisation of a cost function,
thus allowing a better representation of basal and rheological conditions to which the ice flow is sensitive (Barnes et al., 2021).
However, these data assimilation techniques ~~are not~~ may not be well posed (Petra et al., 2014) and a unique solution is never
guaranteed, regardless of the control method used (Barnes et al., 2021). Control methods have regularisation terms which need
to be chosen in order to impose smoothness on the inverted parameters (Koziol et al., 2021). In many studies, the strength
of the regularisation is determined heuristically through L-curve analysis (Gillet-Chaulet et al., 2012; Barnes et al., 2021).
Additionally, control methods do not provide calibration uncertainty. They can be interpreted as methods that return only the
mode of a posterior probability density function (PDF) of the inverted model parameters, which does not fully characterise
calibration uncertainty (Koziol et al., 2021) nor does it propagate the observational uncertainty onto projections of sea-level
rise.

Previous works attempt to quantify uncertainty by considering the forcing uncertainty (Tsai et al., 2017; Robel et al., 2019;
Levermann et al., 2020) or structural uncertainty (Hill et al., 2021). Others consider calibration uncertainty (Isaac et al., 2015;
DeConto and Pollard, 2016; Brinkerhoff et al., 2021; Brinkerhoff, 2022) but use low-dimensional parameter sets (i.e. smaller
than ~20) to describe the ice rheology and basal friction. Here we carry out the first assessment of calibration uncertainty using
a time-dependent marine ice-sheet model (FEniCS_ice, Koziol et al., 2021) in which the calibration of the ice dynamic
parameters scale with the dimension of the numerical grid – i.e. we calibrate each parameter for every element in our mesh of
approximately 100,000 unknowns (see Fig. ~~1~~1).

~~The uncertainty associated with ice-sheet model calibration in this sense can be quantified through~~ We deal with the problem
of estimating the uncertainty in the calibrated parameters (or in the solution of our infinite-dimensional inverse problem)
with the framework of Bayesian inference (Tarantola, 2005; Stuart, 2010), in which prior knowledge is "updated" with ob-
servational evidence (Koziol et al., 2021). ~~The solution of the Bayesian inference problem in our framework hence takes
the form of a very high-dimensional posterior PDF. The calculation of covariances from integration using this PDF, due
to the complexity of the Stokes partial differential equations and our domain, is not possible~~ Given satellite ice velocity
observations (and their uncertainty), a forward model that maps parameters to observations (e.g. FEniCS_ice), and a

prior probability density on model parameters that encodes any prior knowledge or assumptions regarding the parameters

60 (e.g. prior covariance of the ice stiffness parameter in Glen's ice flow law, Glen and Perutz, 1955; Pattyn, 2010), we find the *posterior probability density* of the parameters conditioned on the observational data. This posterior probability density function (PDF) is defined as the Bayesian solution of our ice sheet inverse problem (Petra et al., 2014). A standard approach to characterise this posterior PDF is based on sampling via state-of-the-art Markov chain Monte Carlo (MCMC) methods~~and prohibited~~. Yet the use of conventional MCMC approaches becomes intractable and prohibitive by computational expense ~~(Isaac et al., 2015; Koziol et al., 202~~

65 ~~.~~ for large-scale ice sheet inverse problems where we would need a very large number of realisations of the time-dependent nonlinear Stokes equation (Isaac et al., 2015; Koziol et al., 2021).

However, it can be shown that under certain assumptions, the posterior covariance ~~of the~~ (a property of the joint posterior PDF) of the inverted parameters can be characterised by the inverse of the Hessian (the matrix of second derivatives) of the cost function with respect to the inverted parameters (Thacker, 1989; Kalmikov and Heimbach, 2014; Petra et al., 2014; Isaac et al.,

70 2015). Our framework augments the control method by using this Hessian-based Bayesian approach that not only inverts for the ice dynamic parameters such that model velocities match observations, but characterises the posterior covariance of each inverted parameter (also referred to as control parameters in this study).

We perform a joint inversion for ~~the~~ a basal sliding coefficient and ~~the~~ a rheological parameter for describing ice stiffness. Beginning with a cost function definition which allows for velocity data to be imposed at arbitrary locations (i.e. a point-cloud),

75 we generate a low-rank update approximation to the posterior covariance of the control parameters via the use of the Hessian of the cost function, and find the sensitivities of a time-evolving Quantity of Interest (QoI) to the control parameters. We then project the covariance on to the resulting linear sensitivity to estimate the growth of the QoI uncertainty over time; here our QoI is the sea-level rise contribution or volume above flotation (VAF).

We apply for the first time this error propagation framework to a realistic setting (three ice streams in West Antarctica) and

80 present several model experiments that explore the impact on the uncertainty in forecasts of VAF due to the choice of different strengths of priors (regularisation strength), sliding laws and velocity inputs. We find that significant differences in satellite ice velocity products (particularly at the ice margins) can lead to different projected estimates of sea-level rise contributions or VAF trajectories. We also find that the choice of regularisation strength or regularisation parameters, suggested by L-curve analysis – a common means of estimating such parameters – may lead to an overly informative prior~~and hence underestimate~~

85 ~~the variability in such projections~~. Here the prior information is sufficiently strong that we gain a false low posterior error estimate.

We investigate the effect that data density (density of observed velocity data points) has on the resulting inference. This diagnostic suggests that large amount of data points may be redundant when inverting for the control parameters, with potential implications for observational velocity error models and how they inform the uncertainty in our projections. Additionally, we test

90 our inversion results against the numerical framework of a different ice sheet model ~~(i.e. STREAMICE module of MITgcm Goldberg and H~~ (i.e. the STREAMICE module of MITgcm Goldberg and Heimbach, 2013), in order to qualitatively inspect model structural uncertainty and forcing uncertainty.

2 Methods

The mathematical framework of `FEniCS_ice` is explained in detail in Koziol et al. (2021). In this section we summarise the model physics, the data assimilation techniques used for the calibration of two key ice dynamic parameters, and explain how we quantify calibration uncertainty in projections of sea-level rise contributions or volume above flotation (VAF).

2.1 Physics

`FEniCS_ice` solves the Stokes equation by implementing the well-known Shallow Shelf Approximation (SSA; MacAyeal, 1989; Schoof, 2006; Shapero et al., 2021; Hill et al., 2021). The ice velocity $\mathbf{u}$ is vertically integrated and has two components: internal deformation and basal sliding (see Sect. 3 from Koziol et al., 2021, for details). The model uses data assimilation methods to optimise these velocity components based on observations by estimating two "hidden" properties of the ice; i) the basal friction coefficient (α) in the sliding law, and ii) the rheological parameter for describing ice stiffness (β) in Glen's flow law (both properties are referred to as control parameters in this study). In this section we define the control parameters and the time-dependent SSA whereas the details of the inverse methodology are explained in Sect. 2.4.

2.1.1 Ice rheology and basal sliding

We define the ice viscosity ν, which depends on ~~the~~ ε_e, the second invariant of the strain-rate tensor ~~ε_e~~ as

$$\nu = \frac{1}{2} B \varepsilon_e^{\frac{1-n}{2n}}.$$

B is generally referred to as the "stiffness" of the ice and is thought to depend on ice temperature (Pattyn, 2010). Here we define the control parameter β as the square root of that stiffness where $\beta = \sqrt{B} = \sqrt{A^{-1/n}}$. A in this definition is the rate factor commonly known as the ice creep parameter in Glen's ice flow law (Glen and Perutz, 1955) and n is the exponent of Glen's flow law with the widely accepted value of 3 (Cuffey and Paterson, 2010).

Basal sliding is considered the dominant component of surface velocities in fast-flowing ice streams (Hill et al., 2021), making the time-dependent part of the ice sheet model sensitive to the choice of sliding law (Brondex et al., 2019; Barnes and Gudmundsson, 2022) thus, we consider two different sliding laws. The first is the Weertman–Budd sliding law (Weertman, 1957; Budd et al., 1979; Budd and Jenssen, 1987) defined here as

$$\tau_{\mathbf{b}} = \alpha^2 N^{1/3} u^{-2/3} \mathbf{u} \tag{1}$$

where $\tau_{\mathbf{b}}$ is basal stress, α is the scalar, spatially varying sliding coefficient, u is ice speed, and N is the effective pressure. Here N is defined as

$$N = \rho_i g H + \min(0, R) \rho_w g \tag{2}$$

where ρ_i and ρ_w are ice and ocean densities, g is the magnitude of the gravitational acceleration, H is the ice thickness and R is the bed elevation (Koziol et al., 2021). Furthermore basal stress is nonzero only where ice is grounded, i.e. where

$\rho_i g H + \rho_w R > 0$. The second sliding law considered is often referred to as the Cornford sliding law (Asay-Davis et al., 2016; Cornford et al., 2020) and is defined as

$$\tau_{\mathbf{b}} = \frac{\mu \alpha^2 N u^{\frac{1-m}{m}}}{[\alpha^{2m} u + (\mu N)^m]^{1/m}} \mathbf{u} \tag{3}$$

where $\mu = \frac{1}{2}$ and $m = 3$. A key property of both sliding laws is that as the grounding line is approached, effective pressure becomes small, leading to a smoother transition across the grounding line in terms of the basal drag from floating to grounded conditions.

While additional sliding laws have been proposed and are now implemented within a number of existing ice flow models (Hill et al., 2021), in this study we use the Weertman–Budd sliding law for most of our experiments, as it is one of the most commonly used. However, we trial our error model framework using the Cornford sliding law (Asay-Davis et al., 2016) and compare the results of both sliding laws in Sect. 5.3.

2.1.2 Time-dependent ice sheet model

The resulting calibrated fields of α and β (see Sect. 2.4 for details regarding the parameters calibration) are then input into our forward-in-time simulations where the continuity equation is solved:

$$H_t + \nabla \cdot (H\mathbf{u}) = b. \tag{4}$$

b represents localised changes in mass at the surface and/or the base of the ice sheet, i.e. accumulation due to snow-fall or basal melting of the ice-shelf by the ocean. We assume a constant and uniform surface mass balance field in time and space (i.e. surface mass balance of 0.38 mm of sea level equivalent based in Arthern et al., 2006) and implement a simple depth-dependent parameterisation of ocean melt rate m, which gives the melt rate as a function of ice-shelf draft only. Such parameterisations have been used previously to examine the response of marine ice sheets to ice-shelf melting (e.g., Favier et al., 2014; Seroussi et al., 2017; Lilien et al., 2019; Robel et al., 2019). The form we use is

$$m(z_b) = \frac{M_{max}}{2}\left(1 + \tanh\left[2\left(\frac{z_b - z_{th}}{z_{th}}\right)\right]\right) \tag{5}$$

where z_b is ice-shelf depth, M_{max} is the maximum melt rate and z_{th} represents the depth of the ocean thermocline. m is nonzero only where ice-shelf thickness H is below floatationflotation, and is also set to zero where thickness H is below 10 m. We use such a parameterisation because our aim is to study glaciers which are strongly forced by modified Circumpolar Deep Water (CDW), which is present on certain parts of the Antarctic continental shelf as a warm deep layer overlain by cold surface-modified waters (e.g., Jacobs et al., 2011; Dutrieux et al., 2014; Jenkins, 2016; Jenkins et al., 2018). The form of (5) is chosen because the melt profile transitions from low melt rates above the thermocline depth z_{th} to strong melting at depth, and saturates at M_{max} rather than growing without bound. Defining the parameterisation in this way rather than a piecewise-linear function helps maintain differentiability which aids the later application of algorithmic differentiation (Section 2.5). We discuss our particular choice of M_{max} and z_{th} below in Sect. 3.

The continuity equation is solved with the purpose of finding the loss of ice volume above flotation (VAF), the volume of ice that can contribute to sea level at a certain time T (e.g. $T = 40$ years) which is defined as

$$Q_T^{VAF} = \int_\Omega \left(H(H - H_f) - H_f \right)^+ dA,, \tag{6}$$

where H_f is the ~~floatation~~ flotation *thickness* defined by ~~$-R\left(\frac{\rho_w}{\rho_i}\right)$, and~~ $\max\left(0, -R\left(\frac{\rho_w}{\rho_i}\right)\right)$, Ω is the computational domain (see Sect. 2.2 and Sect. 3 for details), and the $+$ refers to the positive part of the bracketed quantity. Note that we have simplified the ice sheet surface mass balance and the basal melting of the ice shelf thus calculations of future VAF loss estimates presented here do not constitute realistic projections. However, equations (4) and (6) are convenient to calculate such projections and sufficiently nontrivial and nonlinear that the effect of uncertainty arising from the calibration of α and β with observations can be seen.

2.2 Discretisation

We solve the Shallow-Shelf Approximation (SSA) momentum balance as well as (4) using the FEniCS finite-element software library (Alnæs et al., 2015). We discretise velocity $\mathbf{u}$, bathymetry R, and drag and stiffness parameters α and β using first-order continuous Lagrange elements on a triangular finite element mesh. Thickness is defined to be constant within elements (a DG(0) discretisation). The only non-standard aspect of the formulation is in the weak definition of the driving stress $\rho g H \nabla z_{surface}$, which is written as $\mathcal{F} + \mathcal{W}\nabla R$ (see Sect. 3 in Koziol et al., 2021, for $\mathcal{F}$ and $\mathcal{W}$ definitions and the SSA formulation). This formulation is equivalent to the more standard form of driving stress when H is discretised using continuous finite elements, but with this form H can be discretised using zero-order discrete-galerkin (DG(0)) elements as well. The continuity equation (4) is solved using a simple first-order upwind scheme, which is found to be more stable when using a DG(0) thickness function. Details of the mesh generation are explained in Sect. 3 when we discuss the study area.

2.3 Notation

To facilitate readability of this and subsequent sections we adopt formatting conventions for different mathematical objects. Coefficient vectors corresponding to finite-element functions appear as $\overline{c}$; other vectors and vector-valued functions as $\breve{d} \in \mathbb{R}^q$; and matrices as $\mathbf{E}$.

2.4 Cost function J^c

To calibrate the basal sliding coefficient (α) and the rheological parameter for describing ice stiffness (β) we apply data assimilation techniques ~~usually~~ typically used in glaciology (Morlighem et al., 2010; Joughin et al., 2010; Cornford et al., 2015), where the aim is to find the parameter sets which gives the best fit to ice velocity observations. Our approach augments such data assimilation techniques by using a Hessian-based Bayesian approach to characterise uncertainty of α and β. In Sect. 2.5 we describe how we propagate the errors that result from this calibration into projections of VAF. Here we describe how

we invert for the control parameters via the minimisation of a scalar cost function which takes the general form

$$J^c = J^c_{mis} + J^c_{reg}. \tag{7}$$

J^c_{mis}, the misfit cost, is half the square-integral of the misfit between the surface velocity of the ice model and remotely-sensed observations, normalised by the observational ~~standard deviation~~variance. These terms are discretised to implement the control method (as described in Sect. 2.2). The misfit cost ~~is~~ can be written as

$$J^c_{mis} = \frac{1}{2}\|\breve{u}_{obs} - \breve{u}\|^2_{\mathbf{\Gamma}^{-1}_{\mathbf{obs}}}. \tag{8}$$

Here $\breve{u}_{obs}$ is the observed velocity given as cloud point data (location and velocity value)~~and~~; $\breve{u}$ is the velocity estimated via the SSA approximation, interpolated at $\breve{u}_{obs}$ coordinates. ~~As~~; and the norm $\|\cdot\|_{\mathbf{\Gamma}^{-1}_{\mathbf{obs}}}$ is defined by

$$\|\breve{x}\|_{\mathbf{\Gamma}^{-1}_{\mathbf{obs}}} = \sqrt{\breve{x}^T \mathbf{\Gamma}^{-1}_{\mathbf{obs}} \breve{x}}, \tag{9}$$

where $\mathbf{\Gamma}_{\mathbf{obs}}$ is the observational covariance. As non-diagonal error covariance is not given for the considered observational datasets, $\mathbf{\Gamma}_{\mathbf{obs}}$ is a diagonal matrix containing the ~~standard deviation (STD)~~ variance of the observations – note that this neglects observational covariance (see Sect. 3.1.3 and Sect. 6.4 where we discuss observational error covariance).

~~J^c_{reg}~~J^c_{reg}, the regularisation cost, is imposed to prevent instabilities, and is typically chosen as a Tikhonov operator which penalises the square-integral of the gradient of the parameter field (e.g., Morlighem et al., 2010; Cornford et al., 2015). It is defined as

$$J^c_{reg} = \frac{1}{2}\|\bar{c} - \bar{c}_0\|^2_{\mathbf{\Gamma}^{-1}_{\mathbf{prior}}}, \tag{10}$$

where $\bar{c}$ is our hidden field, which depends on both control parameters $\bar{c} = (\alpha, \beta)$. $\bar{c}_0$ is the prior mean and the symmetric positive definite $\mathbf{\Gamma}_{\mathbf{prior}}$ is the prior covariance matrix of the control parameters~~, the construction of which is described below.~~. The terminology "prior" is used because, even though Eq. 10 can be interpreted as a regularisation cost in the context of a deterministic control method inversion, it can be interpreted in terms of a prior PDF in a Bayesian context as discussed in Sections 2.5 and 2.6.

2.4.1 ~~Prior distribution of parameters~~

~~The prior covariance between α and β is set to zero and, following from Koziol et al. (2021), we define a prior covariance for~~ $\mathbf{\Gamma}_{\mathbf{prior}}$ is block-diagonal, with blocks corresponding to each of α and β. Following from Koziol et al. (2021), each block is defined as

$$\mathbf{\Gamma}_{\mathbf{prior}} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L}^{-1}, \tag{11}$$

where $\mathbf{M}$ is the finite element *mass matrix*, and $\mathbf{L}$ is the ~~matrix arising from the~~ *stiffness matrix* that arises from a finite-element discretisation of the differential operator

$$\mathcal{L}(\cdot) \equiv \gamma\nabla^2(\cdot) - \delta(\cdot), \tag{12}$$

where depending on the parameter in question (sliding α or ice stiffness β coefficient) γ is either γ_α or γ_β, and δ is either δ_α or δ_β. J^c_{reg} determines the degree of smoothness of the inverted parameters (determined by $\gamma_{\alpha,\beta}$) and deviation from the prior mean (determined by $\delta_{\alpha,\beta}$). α_0, the prior mean of α, is zero. The prior mean of β is given by

$$\beta_0 = \left(I - \frac{\gamma}{\delta} \nabla^2 \right)^{-1} \beta_{bgd}, \tag{13}$$

where β_{bgd} is the initial guess, described in Section 3. ~~Our form differs from the square-gradient regularisation sometimes used in control methods (e.g., Morlighem et al., 2010; Cornford et al., 2015), but avoids infinite variance as the mesh is refined (Bui-Thanh et al., 2013).~~

~~An important aspect of implementing a prior is to choose the strength with which it is imposed. Here γ and δ in (12) determine the strength of our prior. If the prior is strong (e.g. γ is very small) it will force the posterior to match the prior; dragging the posterior PDF away from the control parameter values which the data suggests are more likely (see Sect. 2.5 for the definition of the posterior PDF of the control parameters). A weak prior imposes only weak restrictions on the plausible range of parameters — potentially beyond the physically plausible range expected a priori. A weak prior can also present practical difficulties, as the posterior maximiser can contain undesirable and nonphysical features that lead to non-stable solutions of the time-dependent SSA. As γ and δ are nonintuitive, we make use of the following expressions for a characteristic pointwise variance $\sigma^2_{c(0)}$ and auto-covariance length scale l_0 of each control parameter (see Sect. 2.2 in Lindgren et al., 2011, for details) :~~

$$\sigma^2_{c(0)} = \frac{1}{(4\pi\gamma\delta)}$$

$$l_{c(0)} = \sqrt{\frac{\gamma}{\delta}}$$

~~In in Sect. 5.1, we show the impact on the VAF projection uncertainty due to the choice of different prior strengths~~ Previous assimilations of satellite velocities also considered ice bed and surface elevations as control parameters (MacAyeal et al., 1995), because available elevation products at the time did not capture small-scale features that could drive variations in velocity. We consider this to be less of an issue with current elevation products (e.g. BedMachine Antarctica v2.0 Morlighem et al., 2020), though future studies with our framework could consider topographic uncertainty and how it covaries with the uncertainty of other parameters.

2.5 Error propagation framework

~~Finally, our~~ Our goal is to find the **posterior probability density function** (PDF) of the control parameters ($\overline{c}$) given the observational data ($\breve{u}_{obs}$); $p(\overline{c}|\breve{u}_{obs})$, and propagate forward the associated uncertainty in ~~time-dependent VAF (denoted here as just~~ projections of VAF (Q_T ~~for short~~). The error propagation framework used here follows from Isaac et al. (2015) and similar studies (Bui-Thanh et al., 2013; Petra et al., 2014), and has been described in detail by Koziol et al. (2021) – here the key elements are summarised.

240   The cost function (7) can be interpreted in a Bayesian sense, ~~with J^c_{reg} being the negative logarithm of the prior density, and J^c_{mis} being the negative logarithm of the likelihood density (each~~ . The misfit term of the cost function, J^c_{mis}, is commonly used in ice-sheet data assimilation, but also is (up to a normalisation term) ~~. J^c is then~~ the negative logarithm of ~~the posterior~~ a multivariate Gaussian PDF with mean $\breve{u}$ and covariance matrix $\mathbf{\Gamma}_{obs}$, and J^c_{reg} has a similar property. This means that Eq. 7 is actually an expression of *Bayes' theorem* (Stuart, 2010):

$$p(\bar{c}|\breve{u}_{obs}) = \frac{p(\breve{u}_{obs}|\bar{c})p(\bar{c})}{p(\breve{u}_{obs})}. \tag{14}$$

The relationship states that the PDF of the ~~control~~ inverted parameters conditioned on the data ~~. The mode of the posterior PDF (i.e. the~~ is determined by both; the *likelihood* of observing the data conditioned on the modelled velocity values, and the *prior* distribution of sliding and stiffness parameters – which is not conditioned on data. The connection with Eq. 7, can be seen by taking the negative logarithm of both sides and ignoring $p(\breve{u}_{obs})$ which is essentially a normalisation constant.

250   Thus, minimising the cost function is equivalent to finding the maximum (or *mode*) of the posterior – often referred to as the maximum a posteriori, or MAP, estimate ~~, assumed here to be unique) is given by the value of the parameters at which J^c obtains its minimum.~~

In the case of a linear model, the posterior inverse covariance, denoted $\mathbf{\Gamma}^{-1}_{\mathbf{post}}$, is given by the Hessian matrix (here referred to as the 'Hessian') of J^c evaluated at the MAP point. In the general case the Hessian defines a Gaussian approximation

255   for the posterior PDF (as the second order approximation for its negative logarithm at the MAP point) and thus defines an approximation for the inverse posterior covariance. Thus, even if the posterior is non-Gaussian, we can learn about its shape in the vicinity of the MAP point, giving more information than if we simply minimised J^c.

If we have estimates of VAF at a given time (6), which depend linearly on the control parameters, and if the posterior ($\mathbf{\Gamma}_{\mathbf{post}}$) is Gaussian, then the posterior variance of VAF at a time T, is given by

260    $$\sigma^2(Q_T) = \left(\frac{\partial Q_T}{\partial \bar{c}}\right)^T \mathbf{\Gamma}_{post}\mathbf{\Gamma}_{\mathbf{post}} \left(\frac{\partial Q_T}{\partial \bar{c}}\right). \tag{15}$$

~~with derivatives evaluated at the minimiser of J^c.~~ Essentially, the posterior parameter uncertainty is projected onto the VAF projection. If for example sliding coefficients in a certain region have high uncertainty due to error-prone data but have little influence on VAF, this will not contribute greatly to VAF uncertainty. In the case that $\mathbf{\Gamma}_{\mathbf{post}}$ is not Gaussian, or estimates of VAF depend non-linearly on the control parameters, (15) yields ~~to~~ an approximation of that posterior variance $\sigma^2(Q_T)$. We

265   discuss in Sect. 6 the limitations of these assumptions.

We use the time-dependent adjoint capabilities of `FEniCS_ice` to find the sensitivities of VAF to the control parameters ($\frac{\partial Q_T}{\partial \bar{c}}$), for discrete values of T over 40 years. The ~~inverse Hessian $\mathbf{\Gamma}_{\mathbf{post}}$ is itself~~Hessian itself, which is a good approximation for $\mathbf{\Gamma}^{-1}_{\mathbf{post}}$, is in general a large, dense matrix which is difficult to invert – and so $\mathbf{\Gamma}_{\mathbf{post}}$ is approximated using a low-rank update to the prior covariance matrix,

270    $$\mathbf{\Gamma}_{post} \approx \mathbf{\Gamma}_{prior} - \mathbf{C}_r \mathbf{\Lambda}_r \left(\mathbf{I}_r + \mathbf{\Lambda}_r\right)^{-1} \mathbf{C}_r^T, \tag{16}$$

where $\mathbf{\Lambda}_r$ and $\mathbf{C}_r$, respectively, represent the leading r eigenvalues and eigenvectors of the *prior-preconditioned misfit Hessian*:

$$\tilde{\mathbf{H}}_{mis} = \mathbf{\Gamma}_{prior}\left(\frac{\partial^2 J_{mis}^c}{\partial \bar{c}^2}\right). \tag{17}$$

Notably, this decomposition has the quality that the leading eigenvectors (those with the largest eigenvalues) are those *most informed* by the data. The leading eigenvectors define the components of the control parameters for which the observations change the estimated posterior uncertainty, relative to the prior uncertainty, by the largest factor. Thus the retained eigenvectors of the Hessian, inform in which space of our mesh the model inversion gained the most information from the observations and the prior (see Fig. 9 and Sect. 2.6 and 5.1 for details).

Computationally the key ingredients to compute $\sigma^2(Q_T)$ are the ability to find a minimiser of J^c, the ability to compute the derivatives of VAF with respect to the control parameters ($\frac{\partial Q_T}{\partial \bar{c}}$), and the ability to compute Hessian information. The minimisation of J^c can be accelerated using gradient-based methods, if J^c can itself be differentiated with respect to $\bar{c}$. Here the required first and second derivative information is obtained using `tlm_adjoint` (Maddison et al., 2019), with L-BFGS (Zhu et al., 1997; Morales and Nocedal, 2011) used to perform the minimisation of J^c and SLEPc (Hernandez et al., 2005, 2007) used to calculate the eigendecomposition. Important points to make are that the eigenproblem requires only the *action* of the misfit Hessian, which would be computationally infeasible to form in full. Additionally, the Hessian takes account of the full nonlinearity of the ice-sheet model, in contrast with the Gauss-Newton approximation to the Hessian (Shapero et al., 2021). In Koziol et al. (2021) a comparison was made between the two in the context of an idealised problem, and results were minimal. For more details on the error propagation framework, see Koziol et al. (2021).

For all the experiments presented in this study, we calculate up to 10^4 (out of 10^5) eigenvalues and eigenvectors to ensure the convergence of $\sigma^2(Q_T)$ against the number of eigenvalues (see results in Sect. 5). The uncertainty of VAF at discrete times $\sigma^2(Q_T)$ is then found using (15), which can then be linearly interpolated to find a "trajectory" of uncertainty.

2.6 Prior distribution of parameters

As mentioned above, the regularisation cost J_{reg}^c can be interpreted in the Bayesian sense in terms of a prior PDF. This prior expresses knowledge of our parameter fields before any data constraints are applied (Arthern, 2015). We model the prior as Gaussian, meaning it is completely defined by its covariance $\mathbf{\Gamma}_{\mathbf{prior}}$ and its mean. Although $\mathbf{\Gamma}_{\mathbf{prior}}$ is determined by the scalars γ and δ (Section 2.4), their meanings are not intuitive. In some of our experiments we therefore make use of the following expressions for a characteristic pointwise variance $\sigma_{c(0)}^2$ and auto-covariance length scale $l_{c(0)}$ of each control parameter (see Sect. 2.2 in Lindgren et al., 2011, for details):

$$\sigma_{c(0)}^2 = \frac{1}{(4\pi\gamma\delta)} \tag{18}$$

$$l_{c(0)} = \sqrt{\frac{\gamma}{\delta}}. \tag{19}$$

For example, if $l_{\alpha(0)}$ and $\sigma^2_{\alpha(0)}$ (the auto-covariance scale and pointwise variance of α, respectively) are both large, then samples of α from the prior are likely to deviate strongly from 0 but show little variation over short length scales. Meanwhile, if $l_{\beta(0)}$ and $\sigma^2_{\beta(0)}$ are both small, then samples of β are likely to vary at small scales but with small deviations from β_0. We consider prior distributions of α and β to be independent.

305   We note that our form of J^c_{reg} differs from the square-gradient regularisation sometimes used in control methods (e.g., Morlighem et al., 2 – but it is used because the associated prior distribution avoids infinite pointwise variance as the mesh is refined (Bui-Thanh et al., 2013) . In Sect. 5.1, we show the impact on the VAF projection uncertainty due to the choice of prior properties.

3   Study area, model domain and data sources

Our study area, shown in Fig. 1, covers part of the Amundsen Sea Embayment (ASE) in West Antarctica and includes three ice
310   streams: Pope, Smith and Kohler Glaciers (PSK), as well as, the Dotson and Crosson ice shelves. PSK glaciers have exhibited some of the highest retreat rates in Antarctica throughout the satellite observing record, with their grounding lines receding over 30 km in recent decades (Scheuchl et al., 2016; Goldberg and Holland, 2022). Their catchment can potentially contribute up to 6 cm to the global mean sea level (Morlighem et al., 2020), double the global mean sea level contribution of the inventory of Earth's mountain glaciers (when excluding the Antarctic and Greenland periphery, Hock et al., 2023). A complete collapse
315   of the ice shelves in this area would likely lead to accelerated mass loss from adjacent ice streams, including Thwaites Glacier (Goldberg and Holland, 2022). Previous modelling studies have shown that past and future retreat of these glaciers is strongly tied to ocean-forced melting, but that the method of calibration may affect projected rates of ice loss as well (Lilien et al., 2019; Goldberg and Holland, 2022). As such, and due to the vast quantity of data available for this region, we choose this area to test our model error framework in a realistic setting.

320   The domain is set up by generating an unstructured finite element mesh using time-averaged strain rates computed from satellite velocity observations (MEaSUREs v1.0 1996 - 2012, Rignot et al., 2014). Additionally, BedMachine Antarctica v2.0 (Morlighem et al., 2020) is used to provide geometry field information and the raster mask from which we define our boundary conditions; ice/ocean (calving) and ice/ice (edge of domain) boundaries in Fig. 1. The mesh generation occurs in two phases, first by generating an initial uniform-resolution mesh of 1000 elements with the mesh generator Gmsh (v.4.8.4 Geuzaine and
325   Remacle, 2009) and second by refining that mesh with the calculated strain metric in the MMG software (v5.5.2 Dobrzynski, 2012). This generates a finer triangular mesh in the areas of the domain where high resolution is needed (e.g. close the calving front and in areas where velocities are higher in Fig. 2). The mesh resolution is highly heterogeneous and depends on the observed strain rates, with a minimum resolution of approx. 200 m and a final mesh size of 102,852 elements. BedMachine Antarctica v2.0 (Morlighem et al., 2020) is also used to define the model's bed, ice thickness and surface elevation fields.
330   The initial guess for β, β_{bgd}, is generated from the temperature dataset of Pattyn (2010). Based on coupled ice sheet-ocean modelling for the region (Goldberg and Holland, 2022), spatially uniform melt parameters of $M_{max} = 30$ m yr^{-1} and $z_{th} =$ 600 m were chosen.

3.1 Velocity input data sources

In the last two decades, ice velocity mapping at the continental scale (Rignot et al., 2011; Gardner et al., 2019, 2018) has
allowed major advances in the study of polar regions by providing complete observations of the complex flow pattern of ice
sheets and glaciers (Mouginot et al., 2017). Much emphasis has been put on the fast processing of large data volumes and
products with complete spatial coverage. However the metadata of such measurements is often highly simplified regarding the
measurements precision and uncertainty (Altena et al., 2022). Moreover, the methods used to estimate errors in the observed
velocities tend to often produce errors that are unrealistically small (see Fig. 2 or Gardner et al., 2019). A quantification of the
error estimation or dispersion (standard deviation) for each individual velocity measurement can be important for the inversion
of unknown ice dynamic parameters (e.g. the basal friction coefficient α). Errors in the velocity data can propagate into derived
results in a complex way, making model outcomes very sensitive to velocity noise and outliers (Altena et al., 2022). Therefore,
we use two satellite velocity products to carry out inversion experiments and calibration uncertainty propagation; MEaSUREs
InSAR-Based Antarctica Ice Velocity Map (MEaSUREs v2.0. Rignot et al., 2017; Mouginot et al., 2017) and ITS_LIVE
surface velocities (Gardner et al., 2019, 2018). To avoid large data gaps in the observations we focus on data acquired between
2013 and 2014 (see Fig 2). MEaSUREs provides surface velocities from July 2013 to July 2014 and ITS_LIVE from January to
December 2014, thus we investigated the effect of the 6 month offset between both products, which turned out to be negligible
(see Fig. A1 of Appendix A). In this section, we describe the acquisition sensors and standard deviation (STD) of each data-set,
as this is relevant to understanding the differences between each product, our experimental design and our results (see Sect. 4.2
and 5.2).

3.1.1 MEaSUREs v2.0

The grid spacing of this data set is 450 m. According to the product metadata (Rignot et al., 2017; Mouginot et al., 2017),
the 2013–2014 year is a result of the data gathered by several instruments: RADARSAT-2 (CSA, 2012-2016), Sentinel-1
(Copernicus/ESA/EU, 2014-2016) and Landsat-8 (2013-2016). Landsat-8 is an optical sensor and it has mapped most of the
ice sheet interior and the Antarctic coast, whereas RADARSAT-2 and Sentinel-1, are C-band synthetic aperture radar (SAR)
instruments and have mostly captured velocities in the coast. Mouginot et al. (2017) notice that along the Antarctic coast,
large differences ($\geq 50 \, \mathrm{m.yr^{-1}}$) between Landsat-8 and SAR based velocities are found, which can be due to stronger weather,
ionospheric noise, and ongoing velocity changes.

3.1.2 ITS_LIVE

The grid spacing of this data set is 240 m. Surface velocities are derived only from optical sensor imagery (Landsat 4, 5,
7, and 8) using the auto-RIFT feature tracking processing chain described in Gardner et al. (2018). Data scarcity and/or low
radiometric quality are significant limiting factors for many regions in the earlier product years. However annual coverage is
nearly complete for the years following the Landsat 8 launch in 2013 (Gardner et al., 2019).

3.1.3 Observational Error model

365 The construction of $\mathbf{\Gamma}_{obs}$ based on reported errors deserves attention. Neither velocity product reports information on spatial error covariance, so $\mathbf{\Gamma}_{obs}$ is diagonal for both products. We interpret the likelihood PDF, $p(\breve{u}|\bar{c})$, as the density associated with the likelihood for a *single outcome* of an observation, as opposed to the distribution of the *average* outcome over an ensemble of observations. Essentially, we consider the standard deviation of observations, as opposed to the standard error of a sample mean.

370 The MEaSUREs product reports both error and standard deviation (STD), and we use the latter to construct $\mathbf{\Gamma}_{obs}$. The ITS_LIVE product does not report standard deviation, but gives the number (count) of measurements for each data point, and expresses error variance as an inverse weighted sum of individual measurement variances (Gardner et al., 2019). We therefore express standard deviation of each velocity component as

$$\text{STD}_{ITS} = count^{\frac{1}{2}} \times err_{ITS}. \tag{20}$$

375 Note this formula assumes uniform variance over all measurements contributing to a data point, which is not likely to be true.

In Koziol et al. (2021) it is shown for an idealised problem that the diagonality of $\mathbf{\Gamma}_{obs}$ leads to ever-decreasing posterior uncertainty as data density is increased. This is only an issue if errors correlate over the scale of separation of data points, but assessing error covariance is beyond the scope of this study. Still, this deficiency guides our investigation of the impacts of data density, described below in Section 4.

4 Experimental design and rationale

380 All inversion methods contain regularisation parameters which must be chosen (Barnes et al., 2021) and ~~L~~L-curve analysis (e.g. Fürst et al., 2015; Jay-Allemand et al., 2011; Gillet-Chaulet et al., 2012; Barnes et al., 2021) is a commonly used technique to make an informative guess regarding the value of these parameters – although there are alternative approaches (see Sect. 2.6 or Waddington et al., 2007; Habermann et al., 2013). Another common aspect of inversions in ice sheet modelling, due to

385 data availability, is to use only one type of remotely-sensed ice velocity product for the calibration of the control parameters. In this section we study how these ongoing practices can impact the forecast of VAF and its uncertainty (see Sect. 4.1 and 4.2). Additionally, we assess the effect that data density (i.e. decreasing the number of observations) has on the inference (see Sect. 4.3). The experiments described in this section lay the groundwork for the model configurations used in Sect. 5.

4.1 *L*-curve analysis on the control parameters

390 The *L*-~~curves are used~~ -curve criterion often used in the ice-sheet modelling literature (Jay-Allemand et al., 2011; Gillet-Chaulet et al., 2012) is based on Hansen (1992, 2001), and is used to visualise the trade-off between the magnitude of the regularisation term (how much the control parameters should vary) and the quality of the fit (how well can we reproduce observations)~~and~~. *L*-curves are generally created by plotting the regularisation terms against the misfit in a log-log scale, the right regularisation term is chosen by locating the "corner" of the *L*-curve. However, there ~~are several ways to construct an *L*-curve as it depends on the cost~~

395 ~~function definition. For consistency with control-method applications in the ice-sheet modelling literature (e.g. Barnes et al., 2021), we~~ is a large variability in the application of this criterion among studies (e.g. compare Seddik et al., 2017; Barnes et al., 2021) and finding the "corner" or "best trade-off" is often chosen heuristically. Here we aim to pick parameters (γ and δ) whose values lie near the corner of the L, where neither J^c or J^c_{reg} take high values, following Barnes et al. (2021) approach.

We generate L-curves by varying the smoothing parameters γ and δ, rather than the variance and length scale arising from a

400 physical interpretation of the prior (Sect. 2.6). Both L-curves in figures 3e and f are computed independently from each other – i.e. varying one parameter over several orders of magnitude while the other three parameters remain fixed (L-curves model configurations are shown in Table A1). To shorten this analysis we show only L-curves for γ in Fig. 3, L-curves for δ can be found in Appendix A2. The L-curves presented in figures 3e and f are created by using ITS_LIVE surface velocities to find the misfit J^c_{mis} (7) and by plotting the regularisation terms against the cost function value J^c, as the regularisation terms γ_α and

405 γ_β (12) vary over several orders of magnitude (10^4 to 10^{-4}).

In order to understand the effect that the strength of the regularisation (or prior strength) has on the control parameters, we show α and β spatial distributions computed using the extreme values of the L-curves (see figures 3a to d). If the prior strength is strong (i.e. a large γ_α) the inverted parameter field (in this case the sliding coefficient) is relatively smooth (see Fig. 3b) and J^c generally small. The L-curve for γ_α in Fig. 3e suggests a $\gamma_\alpha = 100.0$ as a reasonable trade-off between the cost function

410 value and the ~~regularization~~ regularisation term. For γ_β this value is one order of magnitude smaller ($\gamma_\beta = 10.0$, see Fig. 3f). For ~~δ_γ and δ_β~~ $\delta_{\alpha,\beta}$ the L-curve analysis ~~(not shown)~~ suggests a value of 1×10^{-5} .[1] (see Table A1 for all parameters statistics and units). We used those values to conduct the rest of the experiments presented in Sect. 4.2 and 4.3.

4.2 Model output computed with different ice velocity observations

We use the regularisation parameters found in the previous section and run all stages of the error model framework (all methods

415 in Sect. 2) twice using different satellite velocity products for each run; MEaSUREs and ITS_LIVE. We compare the observed ice velocity from both products in Fig. 4b and find significant differences (≥ 100 m.yr^{-1}), especially at the ice margins. The assimilated states of `FEniCS_ice` reproduce these differences as shown in Fig. 4a, ~~were~~ where we show modelled velocity differences between the two runs. Consequently~~;~~, the inverted parameters from both runs are also different (see figures 4c and d)~~and thus are the projections of VAF (see Fig 5a)~~. Differences in the output from both inversions are particularly large at the

420 ice margins, and in the case of the ice stiffness parameter β, the largest differences are found at the Crosson and Dotson ice shelves (see Fig 4d). The projections of VAF differ as well: after 40 years, the difference in VAF is approximately 3.9×10^{11} m^3, or 390 km^3 – nearly 15% of overall VAF loss (Fig 5a).

Fig. 5b shows estimated posterior uncertainty of VAF loss after 40 years from our Hessian-based framework in our L-curve informed workflows. The uncertainty estimates are on the order of ~~10^{10}~~ 10 km^3 ~~, or~~ (10 km^3 ~~. Meanwhile, the difference in~~

425 ~~VAF loss between the MEaSUREs- and ITS_LIVE-based trajectories is ~O(10^{11}~~ for the ITS_LIVE-constrained inversion and 16 km^3 ~~), i.e.~~ for MEaSUREs) – an order of magnitude ~~larger~~ smaller than the observed difference. These results are seemingly at odds. In other words, our error propagation framework suggests a forecast uncertainty that is one order of magnitude smaller

---

[1]~~For each regularisation parameter units please refer to Table 1.~~

than the variability in VAF found when using two different satellite velocity products. This leads to a contradiction, as it suggests that our observed variability in VAF loss is extremely unlikely (see Discussion). This contradiction could arise from one or more of the following: i) the regularisation suggested by the L-curve analysis is too strong, i.e. the prior is overly informative; ii) the observational error covariance matrix used for ITS_LIVE does not capture the true variability of the velocity field shown in Fig. 4b (see Sect. 3.1.3); and/or iii) there are too many data points informing our cost function. We address point (iii) in the next section whereas points (i) and (ii) are addressed in sections 5.2 and 5.1.

4.3 Effect of observational data subsampling

In this section we develop a metric to evaluate the quality of the model's inversions if we decrease the number of observations. In other words, we study the effect that different data densities have on the cost function performance. Similar techniques have been used to gather information on cross-correlations between two different sets of observations (e.g. covariance between winds speed observations in Desroziers et al., 2005), here we apply a similar diagnostic test to study the covariance between adjacent velocity observations from the same product. Additionally, Koziol et al. (2021) show that data density affects the posterior covariance of the control parameters for an idealised experiment, thus we use the results of this metric to test in Sect. 5.2, how data density affects the posterior uncertainty of VAF. The test has the following steps:

1. Produce several training data-sets of observed ice velocity by retaining different percentages of data points from a given set of observations (e.g. ITS_LIVE).

2. Use those training sets to compute inversions of α and β using our L-curve informed prior configuration.

3. Use the α and β results from step 2 to evaluate ~~the cost function~~ J^c_{mis} (Eq. 8) on velocity points that were not used to compute the inversions (i.e. using a validation data-set; observations from a different product such as MEaSUREs).

Based on both products metadata (see Sect. 3.1), we consider MEaSUREs and ITS_LIVE velocities to be two independent realisations of the state of the ice sheet at a given time and location. Hence, we use ITS_LIVE velocities for training and MEaSUREs for validation.

To construct the training sets, we divide the domain into cells of different sizes (different grid-spacing) and systematically drop observations by iterating over the x and y directions of the ITS_LIVE grid. We select sub-samples of the data by retaining corner and center observations from each cell per iteration – i.e. upper and middle cell-points. An example of a training data set is shown in Fig. 6a, where we retain only 1.6% of the velocity observations. To construct the validation set, we downscale MEaSUREs to the ITS_LIVE resolution and drop problematic data points (see figures 6d to f), i.e. locations where MEaSUREs and ITS_LIVE present velocity differences higher than 50 $\mathrm{m.yr}^{-1}$ (as shown in Fig. 4b).

Results of the test (see Fig. 7) demonstrate that our framework provides robust inversions for the drag and stiffness parameters α and β. Additionally, these results reveal that the value of ~~the cost function J^c~~ J^c_{mis} does not change significantly even when we retain only 1.6% of the data, which suggests that a large amount of data points may be redundant when inverting for the control parameters, with potential implications for the error propagation of VAF (see the results from Sect. 5.2). However,

observations are needed to inform the model in critical areas of the domain (i.e. at the grounding line or calving fronts) and if we drop observations in a random manner the performance of ~~the cost function~~ J^c_{mis} may decrease.

5   Results

Results from Sect. 4.2 show that if we use the prior strength suggested by the L-curves and the original ITS_LIVE velocity and standard deviation (STD), our framework underestimates the posterior uncertainty of VAF loss after 40 years by one order of magnitude (see Fig. 5b). More precisely, we calculate a posterior uncertainty which suggests that the difference in projected VAF (estimated by using two velocity products that nominally observe the same physical properties to calibrate our model), is extremely unlikely (see Fig. 5a). To achieve a posterior uncertainty that reflects the same order of magnitude – i.e. $\sim \mathrm{O}(10^{11}$ $\mathrm{m}^3)$, we carry out two more experiments aiming to understand how the uncertainty in forecast of VAF is affected by the use of different strengths of prior (Sect. 5.1) and different versions of the ITS_LIVE data (Sect. 5.2). Additionally, we study the impact of using different sliding laws on the posterior uncertainty of VAF (Sect. 5.3).

5.1   Impact of using different prior strengths on the posterior uncertainty of VAF

We keep the same velocity input for all model configurations trialed in this section (i.e. retaining only 1.6% of the ITS_LIVE data and adjusting the observations STD, as explained in Sect. 5.2), but vary the strengths of the prior. We experiment with the variance $\sigma^2_{c(0)}$ and auto-covariance length scale $l_{c(0)}$ of each control parameter instead of using priors suggested by L-curve analysis, as these definitions (see Sect. 2.6) have a more physical meaning. We calculate prior strengths using (18) and by making an informed guess on $\sigma^2_{c(0)}$ and $l_{c(0)}$ based on existing prior knowledge and physical concepts that define each control parameter.

From the literature (Pattyn, 2010; Khazendar et al., 2011; Still et al., 2022) we know that the spatial pattern of the ice stiffness is not uncorrelated. The advection of colder tributary glacier ice onto the ice shelf is well represented by the vast expanses of stiffer ice originating at the grounding line and extending downstream for tens of kilometers (see panels c and d of Fig. 3), whereas observed deformation patterns at the shear margins (Khazendar et al., 2011; Still et al., 2022) suggest the presence of weaker deformable ice, where the prominent formation of crevasses occur. Thus we keep an auto-covariance length scale $l_{\beta(0)}$ for β equal to 1 km in all prior configurations, as crevasses can be present within that length scale. In future studies this length scale could be set by conducting a detailed spatial statistical analysis of crevasse maps derived from remote sensing, which is beyond the scope of this study. For $\sigma^2_{\beta(0)}$ we trial variance values computed from the STD of the ice stiffness (B) initial guess (see details in Sect. 3).

For the sliding parameter we define auto-covariance length scales $l_{\alpha(0)}$, that are slightly larger than those assumed for β (i.e. 2 to 3 km); as observations from airborne radar over the ice sheet (De Rydt et al., 2013) verify that for fast-flowing ice streams, the surface topography carries important information about the bed with wavelengths between 1 and 20 times the mean ice thickness (≥ 1 km) thus controlling basal sliding at similar scales. Additionally, model experiments described in Gudmundsson (2008) show that the SSA overestimates the effects of bed slipperiness perturbations on the surface profile for

wavelengths less than about 5 to 10 times the mean ice thicknesses, the exact number depending on values of surface slope and slip ratio. Variance values are less intuitive, thus we trial $\sigma^2_{\alpha(0)}$ values over several orders of magnitude in order to vary the prior strength imposed on α.

495   The resultant prior strengths are shown in Table 1 and the estimated posterior uncertainty of VAF for each prior configuration is shown in Fig.8a (solid lines), both are ordered from weak to strong. ~~We find~~ For the weakest prior we find an uncertainty of approximately 160 km^3 – an order or magnitude larger than the estimates found previously. This suggests that a strong prior on the sliding parameter (such as the one suggested by the L-curve analysis), suppresses the error propagation from the satellite data onto projections of VAF. Most of our prior experiments focus on α however, we also trial our error propagation

500 framework changing the variance of β (see Table 1). Compared to prior experiments run on α, changing β priors show little influence in the posterior uncertainty of VAF. We also test weaker priors for both parameters, but these experiments lead to non-stable solutions of the time-dependent SSA as the parameters present undesirable and nonphysical features (not shown).

     Our low-rank approximation to the Hessian (16) makes the assumption that if additional eigenvectors were retained (i.e. if r were larger) the estimated posterior uncertainty $\sigma(Q_T)$ would not change considerably. To test this we examine the marginal

505 change in $\sigma(Q_T)$ for each r (see Fig. 8c), which exhibits an approximately exponential decay with r. To estimate the effect of the low-rank approximation we assume that the decay rate holds up to $r = N$, where N is the full problem size, and that all neglected terms in (16) make a negative contribution to $\sigma(Q_T)$ – i.e. we estimate the "worst case" where every extra eigenvector/eigenvalue calculated decreases the uncertainty (See Appendix B for details of this estimate). The resulting estimated STD's of VAF for an infinite number of eigenvalues σ^{est}_{full} are shown in the captions of Fig. 8c, and indicate that for all prior

510 strengths, even in the worst case, posterior uncertainties decrease by a small proportion and more importantly are not as small as those values seen in our L-curve investigations. We perform this same check in all remaining experiments and observe similar results (see Fig 8d and Fig. 10f).

     Finally, the retained eigenvectors from the Hessian can be interpreted as those modes in the parameter space that change the approximated posterior uncertainty relative to the prior uncertainty. We show in Fig. 9 the leading eigenvectors (see panels

515 a to f) for the prior configuration highlighted in Table 1. The dominant eigenvectors of the ice stiffness parameter field β (panels b, d, and f), suggest that we gained more information regarding this parameter at the grounding line. The eigenvector corresponding to the smallest eigenvalue (Fig. 9h), is increasingly more oscillatory (and thus informs at a smaller length scales in the parameter space) and is increasingly relatively less informed by the velocity observations; such patterns are also present in similar studies (e.g. Isaac et al., 2015).

520 **5.2   Impact of using different ice velocity observations on the posterior uncertainty of VAF**

Contrary to the previous section, here we keep the same prior strength for all the experiments but modify the velocity input. We modify the original ITS_LIVE data by decreasing the amount of observations and by adjusting the STD of each velocity

component to match the following condition:

$$vx_{std} \rightarrow \max(vx_{std}, \quad \text{abs}(vx_I - vx_M)), \tag{21}$$

$$vy_{std} \rightarrow \max(vy_{std}, \quad \text{abs}(vy_I - vy_M)). \tag{22}$$

where vx_{std} and vy_{std} are standard deviations of velocity components, and I and M subscripts refer to ITS _LIVE and MEaSUREs, respectively. In other words, where the original uncertainty of the data is small, we replace those coordinates STD with the absolute difference between MEaSUREs and ITS _LIVE velocities at that same location. Figures 6b and c show this error adjustment for each velocity component. We generate three versions of the ITS _LIVE data by; i) retaining all data points but adjusting the STD, ii) retaining only 1.6% of the data (inline with the results from the observational data subsampling test) and adjusting the STD and iii) retaining only 1.6% of the data but keeping the original STD.

We run our error propagation framework using these three data sets (and the weak prior configuration highlighted in Table 1) and compute VAF posterior uncertainties as shown in Fig. 8b (see solid lines); uncertainties in this figure are plotted from high to low (from dark to light blue) and represent a 95% confidence interval. The effect of retaining only 1.6% of the observational data leads to a slight decrease in posterior VAF uncertainty – which is counterintuitive as one would expect fewer observations to give a larger posterior calibration uncertainty. However, the VAF uncertainty (15) is derived from both calibration uncertainty and VAF sensitivity. The latter differs between the experiments, as can be seen from the projection of prior uncertainty on to VAF sensitivity (blue dashed lines). The adjustment of observational STD increases VAF uncertainty at approximately 5 to 15 years, but has less impact after. Importantly however, the overall impact on the posterior uncertainty of VAF loss after 40 years is small relative to the effect of changing the prior strength (see solid line differences between panels a and b of Fig. 8).

5.3   Impact of using different sliding laws on the posterior uncertainty of VAF

In previous configurations we use the Weertman-Budd sliding law, but due to the sensitivity of the time-dependent ice sheet model to the choice of sliding law (Brondex et al., 2019; Kazmierczak et al., 2022; Barnes and Gudmundsson, 2022) we trial our error propagation framework using the Cornford law and compare qualitative differences from both laws in Fig 10. We use the same velocity constraints (i.e. same as in Sect. 5.2) and consider only a single prior distribution~~(the~~. Ideally, we would need to investigate a range of priors for the Cornford law (as we did for Weertman-Budd in Sect. 5.1); but this is beyond the scope of our study. The prior distribution used is similar to the highlighted parameters in Table 1~~) but modifying $\sigma^2_{\alpha(0)}$ as basal stress does not scale with effective stress in the interior~~, but with a modified $\sigma^2_{\alpha(0)}$ – since in the interior the basal stress is independent of the effective stress (Cornford et al., 2020), and thus we expect variations of α to have a different scale (see last configuration in Table 1).

Using the inverse of the low rank update approximation for the cost function Hessian (Γ_{post}) we can estimate the posterior standard deviation (STD) of α and β. We divide the mesh into 'patches' of approximately 1 km in diameter, and for each patch we compute the mean of each control parameter. We treat this mean as a new quantity of interest (QoI) and compute its STD via the same framework as projections of VAF loss (see Sect. 2.5). Essentially, we visualise in panels a to c of Fig. 10 the posterior of a 'local average' of α and β.

For both sliding laws the sliding parameter α is more uncertain close to the grounding line and at the Bear Peninsula (see Fig. 10 panels a and b) where uncertainties from the ITS_LIVE product are higher (see Fig 2 panels b and c). The large uncertainty just at the grounding line in the Cornford results (relative to that of Weertman-Budd) is due to the insensitivity of basal stress to α when the ice is near ~~floatation~~ flotation (see sensitivity analysis below). For the ice stiffness parameter β the most uncertain areas of our domain are the grounding lines of the PSK glaciers and the Crosson ice shelf (see panel c of Fig. 10) – these are the areas where the speed from the two satellite products show significant differences (see Fig. 4b).

For both sliding laws, VAF uncertainty reaches a similar order of magnitude $\sim O(10^{11})$ m^3 at year 40 (Fig 10e). However, a quantitative comparison is somewhat misleading, as the impact of prior strength is is not investigated for the Cornford law. There are qualitative differences however: the posterior uncertainty of VAF for each sliding law saturates at a different rate, with the posterior uncertainty of the Cornford law configuration growing at a faster rate after year 10. We compare sensitivity maps of the model's VAF estimates to the basal friction coefficient α^2 at year 10 and 40, normalised to year 40 sensitivities for the respective experiment (see Fig. 11). VAF sensitivities at year 10 computed using the Weertman-Budd law have a higher sensitivity to the basal friction coefficient relative to those computed using the Cornford law – particularly at the grounding line of Kohler Glacier (see panels a and c of Fig. 11). Additionally, Fig. 11 shows that at year 40 both sliding laws have similar sensitivities. In this section we only show sensitivity maps for α^2, sensitivities to the ice stiffness are shown in Fig. A3 of Appendix A.

6 Discussion

The efficiency of our error propagation framework allows us to explore how different prior strengths, velocity inputs and sliding laws affect the uncertainty of VAF projections. We find that by choosing different satellite ice velocity products (that nominally observed the same physical properties to calibrate FEniCS_ice) our model leads to different estimates of VAF loss after 40 years (see Sect. 4.2). This effect may be less important for ice streams strongly coupled to ocean forcing (Lilien et al., 2019; Goldberg and Holland, 2022), but could be more influential for unstable margins (Joughin et al., 2014). ~~We use such difference to quantify (in a robust way) the variance that projections of VAF are expected to have after 40 years and identify prior strengths that can reproduce this expected variability (see Sect. 5.1). We demonstrate that if we use prior strengths suggested by the~~ The differences in VAF trajectories (shown in Fig. 5a) computed using the different velocity products, allow us to additionally identify issues in our initial prior probability densities computed using the L-curve criteria. The observed differences are extremely unlikely to be seen under the posterior densities informed by L-curve analysis, ~~as is typically done in ice-sheet calibration studies, our uncertainty quantification is not able to reproduce this level of variability~~ whereas they are far more likely under the posterior informed by physically-motivated priors. The regularisation suggested by the L-curves is thus too strong and thus suppresses the error propagation from the satellite data into the QoI, resulting in VAF projections with quantified uncertainties that are ~~smaller than those suggested by our 2-member "sample" of observed velocity fields. Additionally, our analysis~~ too low.

Our analysis also suggests that the error ~~given by the velocity data~~ provided with the velocity products cannot fully explain the variability in ice velocities observed in Fig. 4b and that large amounts of data points may be redundant, with implications for the error propagation of the QoI.

Our ice sheet flow model described in Sect. 2 can be thought of as a (nonlinear) *mapping* from a set of input fields, which might be unobservable or poorly known (α and β fields), to a set of output fields, which might correspond to observable quantities (e.g. satellite surface velocity observations). In `FEniCS_ice`, the parameter-to-observable map $\breve{f}$ is a composition of two functions: the solution of the SSA equations (see Sect. 3 of Koziol et al., 2021, for details) and the misfit term J^c_{mis} (8).

Our error propagation framework considers the ice sheet inverse problem as a linearised inverse problem; by linearisation we mean that $\breve{f}$ is linearised about the MAP point. Thus the framework relies on a number of key assumptions related to this and other issues:

1. (i) The observational errors and prior distributions are defined as Gaussian distributions, (ii) the parameter-to-observable map $\breve{f}$ is linear (or close to linear), and (iii) the Quantity of interest (i.e. VAF) at a given time depends linearly (or nearly linearly) on the control parameters – in other words, the parameter-to-QoI map is close to linear.

2. The difference between velocities predicted by the model and the observations is due only to measurement errors (we assume zero model error see Sect. 2.4 – or more precisely consider conditional posterior information given the model).

3. The observational error covariance matrix is diagonal, i.e. errors in observations do not correlate spatially.

4. The posterior covariance of the control parameters Γ_{post} is fully sampled with the number of eigenvectors and eigenvalues that we retain from the Hessian.

Note (1).(i,ii) above implies a Gaussian posterior. We already test (4) in Sect. 5.1 and address (1)-(3) in the following subsections.

6.1 Linear dependence of parameter-to-observable and Quantity of Interest maps with respect to the control parameters.

`FEniCS_ice` computes a second-order approximation to the posterior covariance of the control parameters Γ_{post} (via the eigendecomposition of the cost function Hessian evaluated at the MAP point, see 2.5) and propagates forward the associated calibration uncertainty in time-dependent estimates of VAF loss (our QoI). The posterior PDF of $\bar{c}$ is not guaranteed to be Gaussian due to the nonlinearity of the Stokes equations that describe $\breve{f}$. Furthermore, the propagation step (Eq. 15) is based on a linear transformation of a Gaussian random variable, and assumes that $Q_T(\bar{c})$, the parameter-to-QoI map, is well-described by linear sensitivities.

Petra et al. (2014) test the Gaussianity of the parameter-to-observable map by sampling from the posterior PDF of the hidden field $\bar{c}$ via ~~two~~ different Markov chain Monte Carlo (MCMC) sampling methods (Tierney, 1994); ~~the Newton MCMC method and~~ including a new stochastic Newton method with a MAP-based Hessian. They solve a two-dimensional flow-line ice sheet inverse problem with a moderate number of parameters (~ 100)~~and conclude that the most non-Gaussian behavior for $\breve{f}$ occurs~~

620 ~~in the directions of the domain with the largest variance in the observations, thus $\breve{f}$ deviates from a linear approximation in those directions (Petra et al., 2014; Isaac et al., 2015). Nevertheless, in the directions where the variance is large.~~ It is suggested that for control parameter directions which are more strongly informed by observations, and in the weak observational noise limit, the posterior ~~covariance of the control parameters will be significantly influenced by the prior – which is also defined as Gaussian thus it is expected that the Gaussianity assumption will hold and~~ may be closer to Gaussian, due to the weaker

625 influence of the non-linearity of $\breve{f}$~~will be weakly linear in such directions~~. It is further suggested that for directions which are weakly informed by observations, and hence for which the Gaussian prior dominates, the posterior may again be closer to Gaussian. Therefore, a Hessian-based approximation (such as the one describe in Sect. 2.5) to the posterior covariance of the parameters may be ~~appropriate~~ useful despite the nonlinearity of $\breve{f}$~~(Petra et al., 2014)~~.

Koziol et al. (2021) test the linearity of the `FEniCS_ice` parameter-to-QoI map for an idealised ice sheet flow problem

630 (Pattyn et al., 2008) through a simple Monte Carlo sampling of the posterior PDF of $\bar{c}$. The study finds strong agreement with the linearly propagated posterior covariance when there is a moderately strong prior, but slightly poorer agreement with a weak prior.

Unfortunately, due to the size of our parameter space, testing the Gaussianity of the posterior PDF of $\bar{c}$ is beyond the scope of our study. Similarly, sampling the posterior PDF to validate the propagation of calibration uncertainty to the QoI as in Koziol

635 et al. (2021) would be intractable for our more realistic setting. Instead, we develop a simple test to check the linearity of the parameter-to-QoI map and how this linearity is affected when we impose different strengths of prior. We test the linearity of the parameter-to-QoI mapping by using data from Sect. 4.2 to compute the following dot products:

$$\frac{\partial Q_T^I}{\partial \alpha_I} \cdot (\alpha_I - \alpha_M), \tag{23}$$

$$\frac{\partial Q_T^I}{\partial \beta_I} \cdot (\beta_I - \beta_M). \tag{24}$$

640 Here I and M indicate model output computed by either using ITS_LIVE or MEaSUREs velocities. We visualise the linearity of the VAF operator by plotting each dot product together with the absolute difference between VAF trajectories computed using ITS_LIVE and MEaSUREs. Additionally, we repeat this test for a stronger prior configuration by imposing a strong regularisation on β (stronger than the one suggested by the L-curve analysis, $\gamma_\beta = 100.0$).

Results from both tests are shown in Fig 12 and verify that VAF estimates over time are highly dependent on α and that the

645 linearity of the parameter-to-QoI map depends on the strength of the regularisation (as in Fig. 12b and in Koziol et al., 2021). The main objective of this study is to propagate calibration uncertainty into projections of VAF loss. We find that in order to do so, we must impose a weaker prior on the control parameters than widely-used methods (i.e. L-curve analysis) would suggest. But as shown above, in doing so we might need to compromise on the linearity of the parameter-to-QoI map. Moreover, as shown in Koziol et al. (2021), a weaker prior means a weaker spectral decay of the prior-preconditioned Hessian spectrum,

650 requiring to retain more of its eigenvectors (see also Sect. 5.1).

In other words, to avoid the prior probability from overwhelming the likelihood in our Bayesian inversion, we are required to examine a regime where we compromise the linearity of the time-dependent model in certain areas of the domain. Still, we expect that the framework can provide an "order-of-magnitude" estimate of the contribution of calibration uncertainty

to QoI uncertainty. Although not previously applied to a problem as large as the present study, Stochastic Newton MCMC (Martin et al., 2012; Petra et al., 2014), which does not rely on a Gaussian assumption, may provide a more robust estimate in such regimes. Importantly, to be tractable this method requires a reasonable estimate of the posterior density (the "proposal density") – and such an estimate can be provided using the low-rank Hessian approximation generated within our framework. Thus Stochastic Newton MCMC may be a viable approach for non-Gaussian uncertainty quantification in future studies.

6.2 Qualitative inspection of the model's structural and forcing uncertainty

We only quantify calibration (parametric) uncertainty in projections of marine ice sheet loss. We do not quantify structural or model uncertainty, i.e. errors that arise from the discretisation of the inverse problem (Barnes et al., 2021) or from the formulation of the model and its ability to represent the physics of the system (Hill et al., 2021). In Sect. 5.3 we trial our error propagation framework with different sliding laws and examine the implications for projections of VAF loss (see Fig. 10); though quantifying the likelihood of various sliding-law formulations is beyond the scope of our study.

In this section we look at uncertainty due to the use of different physics and discretisation to solve the ice-sheet momentum balance. We do this by using a second ice sheet model: the STREAMICE module of MITgcm (Goldberg and Heimbach, 2013), which solves a depth-integrated balance that accounts for vertical shearing (absent from the shallow ice-stream approximation; Goldberg, 2011). STREAMICE solves the momentum balance on a regular rectangular grid, a distinct discretisation from `FEniCS_ice`. With a uniform 500 m grid, we simulate with STREAMICE an instantaneous velocity field (without time evolution), using the inverted parameter fields of `FEniCS_ice` (interpolated to the STREAMICE grid) and the same geometry and boundary conditions described in Sect. 3. The particular fields of α and β are from our L-curve analysis (Sect. 4.1). We compare both models' surface velocities and find differences on the order of 100-200 myr^{-1}, particularly in the fastest-flowing ice areas and on the ice shelves (see Fig. A4). `FEniCS_ice` and STREAMICE have different approximations to Stokes flow, employ different treatments of the grounding line in their equations, and have very different resolution, which may lead to this disagreement. Barnes et al. (2021) find similar results, when two other adaptive mesh, finite-element SSA models are compared to STREAMICE through the same diagnostic experiment. The authors show that these diagnostic calculations are not indicative of the performance of the models in time-dependent simulations (see Fig. 6 of Barnes et al., 2021, where all models reach similar projections of VAF).

We emphasise this comparison cannot quantify structural uncertainty, but can inform us (qualitatively) of the effects of implementing different discretisations and grounding line formulations in the model numerics.

6.3 Relevance of calibration uncertainty versus structural and forcing uncertainty

As previously mentioned, structural uncertainty is neglected in our study and we use a very simple ocean forcing parameterisation, for which uncertainties are not considered. We make clear that our aim is to quantify calibration uncertainty alone; however, it is only worth doing so if the contribution of calibration uncertainty to forecast uncertainty is non-negligible, and/or the framework represents nontrivial steps toward incorporating these other sources of uncertainty. Regarding the former, the existing literature provides some clues as to whether calibration uncertainty is important. Goldberg and Holland (2022) carry

out coupled ice sheet-ocean modelling experiments for the PSK glaciers region, and show that the type of calibration of ice model parameters (i.e. whether fit to observed thinning is accounted for) strongly determines ice loss over 20-30 years; beyond this point, ice loss depends on far-field ocean conditions. For other catchments, this "crossover time" could be shorter, or longer – meaning that uncertainty in calibration could inform projection uncertainty on the multidecadal scale before it is overtaken by climate uncertainty. The short-term persistence of calibration errors is echoed in other types of cryospheric modelling: Aschwanden and Brinkerhoff (2022) showed that the introduction of satellite-based information strongly reduced uncertainty in short-term projections of Greenland ice loss, but that this relative information gain was greatly reduced by 2100, particularly under strong climate forcing scenarios. Still, calibration uncertainties should not be dismissed even if they are overwhelmed by climate forcing on long time scales: there are strong reasons why short-term (multidecadal) projections of ice loss are key for planning and mitigation (Bassis, 2022).

Moreover, our framework of estimating calibration uncertainty can easily be expanded to account for forcing uncertainty. Provided that forcing uncertainty is independent of parameter uncertainty, the contribution of forcing to projection uncertainty is additive, and can be found using an expression similar to Eq. 15. Importantly, such a calculation is independent from the estimation of posterior parameter uncertainty through eigendecomposition of the Hessian – which is by far the most costly component. This is not true of model uncertainty: our likelihood PDF $p(\breve{u}|\bar{c})$, which gives the probability of observable velocity conditioned on parameters and the model, and hence neglects model uncertainty. A potential way to incorporate model uncertainty – once it is quantified – is to adjust the observational error covariance used in the likelihood. A similar approach has been used in the Bayesian Error Approximation method of Babaniyi et al. (2021).

6.4 Accuracy of observational error model

We draw the tentative conclusion that, for our study area, the choice of prior distribution informed by L-curve analysis is overly informative and underestimates calibration uncertainty. This is based on the fact that, with such a prior, the posterior VAF is an order of magnitude smaller than the variability in VAF between two widely used velocity products as constraining data. Essentially, the two products are treated as a 2-member sample from a distribution describing the true surface velocities. While this is a very small sample size, our assessment makes the assumption that (i) the posterior VAF distribution is Gaussian (which is explored above) and (ii) the two members are likely sampling outcomes under our observational error model – and therefore that the variation of $\sim O(10^{11} \text{ m}^3)$ is not a statistical outlier – thus our Hessian-based assessment must be too small.

A further assumption in our assessment is that our observational error model is accurate. As described in Sect. 3.1.3, we use reported errors and standard deviations ~~as~~ to define diagonal terms in $\mathbf{\Gamma_{obs}}$ and assume zero spatial error covariance. Error magnitudes may be underestimated – although we somewhat account for this by adjusting observational errors based on differences between the products (Sect. 5.2). Additionally, not accounting for spatial error correlation could underestimate calibration uncertainty, as shown in the idealised experiments of Koziol et al. (2021). It is possible that improved assessments of spatial observational error covariance may be needed to accurately quantify calibration uncertainty when calibrating ice-sheet models with satellite-based data. Such approaches have been used in weather data assimilation ~~Tabeart et al. (2020)~~ (Tabeart et al., 2020).

7 Conclusions

This study set out to apply `FEniCS_ice` error propagation framework to a realistic setting in West Antarctica (which includes three ice streams: Pope, Smith and Kohler Glaciers) and infer from satellite velocity observations two important unknown parameters in ice dynamics and its uncertainties; the basal sliding friction coefficient (α) and the rheological parameter for describing ice stiffness (β). As in many other ice-sheet modelling studies we use a control method to calibrate grid-scale flow parameters. However, our framework augments the control method with a Hessian-based Bayesian inference approach, which characterises the posterior covariance of the inverted parameters. We project calibration uncertainty forward in time and onto projections of volume above flotation (VAF).

We find that by choosing different satellite ice velocity products (that nominally observed the same physical properties to calibrate `FEniCS_ice`) our model leads to different estimates of VAF after 40 years or to different projections of sea-level rise contribution. We use this difference in model output as an order-of-magnitude estimate of the variance that projections of VAF should have after 40 years and identify prior strengths that can reproduce that variability. We demonstrate that if we use prior strengths suggested by L-curves, as is typically done in ice-sheet calibration studies, our uncertainty quantification is not able to reproduce that same variability. The regularisation suggested by the L-curves is too strong (i.e. the information contained in the prior distribution computed via L-curve analysis dominates over the information contained in ice velocity observations) and thus suppresses the error propagation from the satellite data into VAF projections with quantified uncertainties that are smaller than those suggested by our 2-member "sample" of observed velocity fields. Therefore, we recommend to use the variance and length scale arising from a physical interpretation of the prior to define regularisation parameters; as these definitions (see Sect. 2.6) will inform the ice sheet model with a more realistic spatial variability regarding the basal sliding and ice stiffness parameters. Additionally, our experiments suggest that large amounts of data points may be redundant, with implications for the error propagation of VAF.

We inspect qualitatively one aspect of structural uncertainty by trialing our error propagation framework with two different sliding laws (Weertman-Budd and Cornford laws). The posterior uncertainty of VAF evolves differently for the two parameterisations, with the Weertman-Budd uncertainty saturating relatively quickly while that of Cornford steadily increases. This may be due to differing patterns of sensitivity of VAF to the sliding parameters, particularly near the grounding line.

Finally, our framework alone does not fully quantify sea-level rise forecast uncertainty, but represents an important step. Further improvements to our method could be to (i) quantify calibration uncertainty through Stochastic Newton Markov chain Monte Carlo (MCMC) using our Hessian eigendecomposition as the proposal density, (ii) take into account model error in the likelihood probability density function and (iii) take into account forcing uncertainty in the error propagation framework.

Code availability. The version of `tlm_adjoint` used in this manuscript is available in a permanent DOI repository (https://zenodo.org/badge/latestdoi/452296839). The `FEniCS_ice` software together with the application of the code to a real domain are coded in the Python language and licensed under the GPL-3.0 license. The latest version of the `FEniCS_ice` code is available on Github (https://github.com/EdiGlacUQ/fenics_ice), the documentation website of the model is under construction but a user guide is provided (https:

//github.com/EdiGlacUQ/fenics_ice/tree/main/user_guide). The code used to generate all figures and analyses of this study is available in a permanent DOI repository (https://doi.org/10.5281/zenodo.7615259) and on Github (https://github.com/bearecinos/smith_glacier) as well as the FEniCS_ice version used for this study (https://zenodo.org/badge/latestdoi/101511241). We have constructed a documentation website for the application of the model to the PSK Glaciers domain (https://github.com/bearecinos/smith_glacier/wiki) where we explain in detail the installation of the code, the preparation of input data and how to run and visualise the experiments presented in this study.

Data availability. The output data of the model is available in the following permanent DOI repository https://doi.org/10.5281/zenodo.7612243. Information about how to read and plot the data can be found in the Smith repository wiki, see: https://github.com/bearecinos/smith_glacier/wiki.

Appendix A: Extra figures and tables

A1 ITS_LIVE 6 month offset speed change

To study the 6 month offset between ITS_LIVE and MEaSUREs velocities – i.e. from July 2014 to December 2014, we subtract from the 2018 ITS_LIVE velocity mosaic, data acquired in 2014 and divide this by 8 in order get the monthly changes (see Fig. A1a). The effect of the 6 month offset between both products is negligible compared to the difference observed in Fig. 4b and to the speed ratio shown in Fig. A1b. However, there are significant differences (over a small area) at the calving front of the Crosson Ice shelf (see Fig. A1a).

A2 δ_α *L*-curve

In addition to the experiments shown in Sect. 4.1, we constructed a *L*-curve **only for** δ_α and use the same result for δ_β. Results are shown in Fig. A2 and suggest a value for δ_α of 1×10^{-4} compared to values in Table 1, any $\delta_\alpha \geq 1 \times 10^{-7}$ will result in a stronger prior thus we choose $\delta_\alpha = 1 \times 10^{-5}$ for the error propagation experiments shown in Fig. 3 e and f.

A3 Sensitivity of VAF to the ice stiffness (B)

We compare sensitivity maps of the model's VAF estimates to the ice stiffness B (or β^2) at year 10 and 40 (see Fig. A3). VAF projections are more sensitive to the ice stiffness at the grounding line of the PSK Glaciers and at the Crosson Ice shelf. In future studies, these sensitivity maps and the ice stiffness spatial distribution could be correlated to detailed spatial maps of crevasses in the area.

Appendix B: Convergence of the estimated posterior uncertainty $\sigma(Q_T)$ with the number of eigenvectors

For each successive eigenvalue-eigenvector pair (λ_r, C_r) we construct the low-rank approximation to the posterior covariance using Eq. 16, and find the associated approximation to $\sigma(Q_T)$ by projecting the estimated covariance on to the QoI (Eq. 15).

We refer to this iterate here as σ_r – the posterior QoI uncertainty using the leading r eigenvectors – and to the difference $\sigma_r - \sigma_{r-1}$ as $\Delta \sigma_r$.

We observe that, for sufficiently large r, the absolute change with r can be represented reasonably well by an exponential decay i.e.

$$|\Delta \sigma_r| = d_0 b^r \tag{B1}$$

for some b smaller than 1 (Fig. 8(c,d)). Assuming this to hold as r becomes large, we can estimate a lower bound for $\sigma(Q_T) = \sigma_N$ (where N is the parameter dimension) with a geometric sum. Specifically, we find the d_0 and b that best fit $|\Delta \sigma_r|$ for $r_{th} \leq r \leq r_M$, where r_{th} is inferred from the decay of $|\Delta \sigma_r|$ and r_N is the number of eigenpairs retained (in our case 10^4). The relationship given by Eq. B1 then implies for $M > N$

$$\sigma_M = \sigma_N + \sum_{r=N+1}^{M-1} \Delta \sigma_r$$
$$\leq \sigma_N - d_0 b^N \left(\frac{1 - b^{M-N}}{1 - b} \right)$$
$$< \sigma_N - \frac{d_0 b^N}{1 - b}. \tag{B2}$$

In Sect. 5 we use this result (with $r_{th} = 3000$) to estimate a lower bound for the posterior uncertainty of Q_T without low-rank approximation. The calculation is done at the final time i.e. for $T = 40$ a only. We emphasise that this calculation is purely heuristic, and we are unaware of a theoretical lower bound for σ_M. Due to the tendency of the shallow-shelf approximation to filter high spatial frequencies in basal parameters (Gudmundsson, 2008), it is unlikely that $|\Delta \sigma_r|$ will decay more slowly than predicted by Eq. B1, and it may even decay more quickly. However due to the large size of the parameter space (10^5) it is not tractable to find the full spectrum, so the estimate is not testable for this problem.

Finally, other studies use eigenvalue magnitude as a criterion for truncating the spectrum (e.g., Isaac et al., 2015). More specifically, eigenvectors are retained up to an index r such that $\frac{\lambda_r}{\lambda_r + 1} \ll 1$. We note this constraint alone does not ensure that the contribution to QoI uncertainty arising from the truncated part of the spectrum is negligible, even if the marginal contribution associated with each individual eigenpair is small.

Author contributions. BR, DG and JRM jointly conceived the study and contributed equally to the writing and development of the experiments presented in this manuscript. DG and JRM developed the mathematical framework implemented in `FEniCS_ice`. BR develop the Smith Glacier repository and its documentation, which consist of several experiments with `FEniCS_ice` over Smith, Pope, and Kohler Glaciers. JT provided significant contributions to the `FEniCS_ice` software and to the pre-processing tasks module of the Smith Glacier repository. BR and DG performed the data analysis presented in this study with input from the other authors.

Competing interests. The authors declare they have no competing interests.

Acknowledgements. ~~BR is~~ All authors are supported by NERC standard grant NE/T001607/1 (QUoRUM). The computations were realised on the computing facilities of the University of Edinburgh, Scotland, UK. We acknowledge Conrad Koziol contributions as developer of the `FEniCS_ice` software. We calculate that the carbon footprint of the computations is approximate 100.13 kg CO2e (equivalent to 2 flights from London to Paris); this calculation its done using Lannelongue et al. (2021) and based on the amount of times that we ran a complete model workflow on our specific hardware. However, this does not account for permanent storage used.

810

References

Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N.: The FEniCS Project Version 1.5, Archive of Numerical Software, 3, 9–23, 2015.

Altena, B., Kääb, A., and Wouters, B.: Correlation dispersion as a measure to better estimate uncertainty in remotely sensed glacier displacements, The Cryosphere, 16, 2285–2300, https://doi.org/10.5194/tc-16-2285-2022, 2022.

Arthern, R. J.: Exploring the use of transformation group priors and the method of maximum relative entropy for Bayesian glaciological inversions, Journal of Glaciology, 61, 947–962, https://doi.org/10.3189/2015JoG15J050, 2015.

Arthern, R. J., Winebrenner, D. P., and Vaughan, D. G.: Antarctic snow accumulation mapped using polarization of 4.3-cm wavelength microwave emission, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/https://doi.org/10.1029/2004JD005667, 2006.

Asay-Davis, X. S., Cornford, S. L., Durand, G., Galton-Fenzi, B. K., Gladstone, R. M., Gudmundsson, G. H., Hattermann, T., Holland, D. M., Holland, D., Holland, P. R., Martin, D. F., Mathiot, P., Pattyn, F., and Seroussi, H.: Experimental design for three interrelated marine ice sheet and ocean model intercomparison projects: MISMIP v. 3 (MISMIP+), ISOMIP v. 2 (ISOMIP+) and MISOMIP v. 1 (MISOMIP1), Geoscientific Model Development, 9, 2471–2497, https://doi.org/10.5194/gmd-9-2471-2016, 2016.

Aschwanden, A. and Brinkerhoff, D.: Calibrated Mass Loss Predictions for the Greenland Ice Sheet, Geophysical Research Letters, 49, e2022GL099 058, https://doi.org/10.1029/2022GL099058, 2022.

Aschwanden, A., Bartholomaus, T. C., Brinkerhoff, D. J., and Truffer, M.: Brief communication: A roadmap towards credible projections of ice sheet contribution to sea level, The Cryosphere, 15, 5705–5715, https://doi.org/10.5194/tc-15-5705-2021, 2021.

Babaniyi, O., Nicholson, R., Villa, U., and Petra, N.: Inferring the basal sliding coefficient field for the Stokes ice sheet model under rheological uncertainty, The Cryosphere, 15, 1731–1750, https://doi.org/10.5194/tc-15-1731-2021, 2021.

Barnes, J. M. and Gudmundsson, G. H.: The predictive power of ice sheet models and the regional sensitivity of ice loss to basal sliding parameterisations: a case study of Pine Island and Thwaites glaciers, West Antarctica, The Cryosphere, 16, 4291–4304, https://doi.org/10.5194/tc-16-4291-2022, 2022.

Barnes, J. M., Dias dos Santos, T., Goldberg, D., Gudmundsson, G. H., Morlighem, M., and De Rydt, J.: The transferability of adjoint inversion products between different ice flow models, The Cryosphere, 15, 1975–2000, https://doi.org/10.5194/tc-15-1975-2021, 2021.

Bassis, J.: Quit Worrying About Uncertainty in Sea Level Projections, Eos, Transactions American Geophysical Union, 102, https://doi.org/https://doi.org/10.1029/2021EO210632, 2022.

Brinkerhoff, D., Aschwanden, A., and Fahnestock, M.: Constraining subglacial processes from surface velocity observations using surrogate-based Bayesian inference, Journal of Glaciology, 67, 385–403, https://doi.org/10.1017/jog.2020.112, 2021.

Brinkerhoff, D. J.: Variational inference at glacier scale, Journal of Computational Physics, 459, 111 095, https://doi.org/10.1016/j.jcp.2022.111095, 2022.

Brondex, J., Gillet-Chaulet, F., and Gagliardini, O.: Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law, The Cryosphere, 13, 177–195, https://doi.org/10.5194/tc-13-177-2019, 2019.

Budd, W. F. and Jenssen, D.: Numerical Modelling of the Large-Scale Basal Water Flux under the West Antarctic Ice Sheet, in: Dynamics of the West Antarctic Ice Sheet, edited by Van der Veen, C. J. and Oerlemans, J., pp. 293–320, Springer Netherlands, Dordrecht, 1987.

Budd, W. F., Keage, P. L., and Blundy, N. A.: Empirical Studies of Ice Sliding, Journal of Glaciology, 23, 157–170, https://doi.org/10.3189/S0022143000029804, 1979.

850 Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G.: A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion, arXiv e-prints, arXiv:1308.1313, 2013.

Cornford, S. L., Martin, D. F., Payne, A. J., Ng, E. G., Le Brocq, A. M., Gladstone, R. M., Edwards, T. L., Shannon, S. R., Agosta, C., van den Broeke, M. R., Hellmer, H. H., Krinner, G., Ligtenberg, S. R. M., Timmermann, R., and Vaughan, D. G.: Century-scale simulations of the response of the West Antarctic Ice Sheet to a warming climate, The Cryosphere, 9, 1579–1600, https://doi.org/10.5194/tc-9-1579-2015,

855 2015.

Cornford, S. L., Seroussi, H., Asay-Davis, X. S., Gudmundsson, G. H., Arthern, R., Borstad, C., Christmann, J., Dias dos Santos, T., Feldmann, J., Goldberg, D., Hoffman, M. J., Humbert, A., Kleiner, T., Leguy, G., Lipscomb, W. H., Merino, N., Durand, G., Morlighem, M., Pollard, D., Rückamp, M., Williams, C. R., and Yu, H.: Results of the third Marine Ice Sheet Model Intercomparison Project (MISMIP+), The Cryosphere, 14, 2283–2301, https://doi.org/10.5194/tc-14-2283-2020, 2020.

860 Cuffey, K. and Paterson, W.: The Physics of Glaciers, 4th Edition, Academic Press, 2010.

De Rydt, J., Gudmundsson, G. H., Corr, H. F. J., and Christoffersen, P.: Surface undulations of Antarctic ice streams tightly controlled by bedrock topography, The Cryosphere, 7, 407–417, https://doi.org/10.5194/tc-7-407-2013, 2013.

DeConto, R. M. and Pollard, D.: Contribution of Antarctica to past and future sea-level rise, Nature, 531, 591–597, https://doi.org/10.1038/nature17145, 2016.

865 Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/https://doi.org/10.1256/qj.05.108, 2005.

Dobrzynski, C.: MMG3D: User Guide, Technical Report RT-0422, INRIA, https://hal.inria.fr/hal-00681813, 2012.

Dutrieux, P., De Rydt, J., Jenkins, A., Holland, P., Ha, H., Lee, S., Steig, E., Ding, Q., Abrahamsen, E., and Schröder, M.: Strong Sensitivity of Pine Island Ice-Shelf Melting to Climatic Variability, Science, 343, 174–178, https://doi.org/10.1126/science.1244341, 2014.

870 Favier, L., Durand, G., Cornford, S. L., Gudmundsson, G. H., Gagliardini, O., Gillet-Chaulet, F., Zwinger, T., Payne, A., and Brocq, A. M. L.: Retreat of Pine Island Glacier controlled by marine ice-sheet instability, Nature Climate Change, 4, 117–121, https://doi.org/10.1038/nclimate2094, 2014.

Fürst, J. J., Durand, G., Gillet-Chaulet, F., Merino, N., Tavard, L., Mouginot, J., Gourmelen, N., and Gagliardini, O.: Assimilation of Antarctic velocity observations provides evidence for uncharted pinning points, The Cryosphere, 9, 1427–1443, https://doi.org/10.5194/tc-9-1427-

875 2015, 2015.

Gardner, A. S., Moholdt, G., Scambos, T., Fahnstock, M., Ligtenberg, S., van den Broeke, M., and Nilsson, J.: Increased West Antarctic and unchanged East Antarctic ice discharge over the last 7 years, The Cryosphere, 12, 521–547, https://doi.org/10.5194/tc-12-521-2018, 2018.

Gardner, A. S., Fahnestock, M. A., and Scambos, T. A.: ITS_LIVE Regional Glacier and Ice Sheet Surface Velocities. Data archived at National Snow and Ice Data Center, https://doi.org/10.5067/6II6VW8LLWJ7, accessed: June 14, 2023., 2019.

880 Geuzaine, C. and Remacle, J.-F.: Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities, International Journal for Numerical Methods in Engineering, 79, 1309–1331, https://doi.org/https://doi.org/10.1002/nme.2579, 2009.

Gillet-Chaulet, F., Gagliardini, O., Seddik, H., Nodet, M., Durand, G., Ritz, C., Zwinger, T., Greve, R., and Vaughan, D. G.: Greenland ice sheet contribution to sea-level rise from a new-generation ice-sheet model, The Cryosphere, 6, 1561–1576, https://doi.org/10.5194/tc-6-1561-2012, 2012.

885 Glen, J. W. and Perutz, M. F.: The creep of polycrystalline ice, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 228, 519–538, https://doi.org/10.1098/rspa.1955.0066, 1955.

Goldberg, D. N.: A variationally derived, depth-integrated approximation to a higher-order glaciological flow model, Journal of Glaciology, 57, 157–170, https://doi.org/10.3189/002214311795306763, 2011.

Goldberg, D. N. and Heimbach, P.: Parameter and state estimation with a time-dependent adjoint marine ice sheet model, The Cryosphere, 7, 1659–1678, https://doi.org/10.5194/tc-7-1659-2013, 2013.

Goldberg, D. N. and Holland, P. R.: The Relative Impacts of Initialization and Climate Forcing in Coupled Ice Sheet-Ocean Modeling: Application to Pope, Smith, and Kohler Glaciers, Journal of Geophysical Research: Earth Surface, 127, e2021JF006 570, https://doi.org/https://doi.org/10.1029/2021JF006570, e2021JF006570 2021JF006570, 2022.

Gudmundsson, G. H.: Analytical solutions for the surface response to small amplitude perturbations in boundary data in the shallow-ice-stream approximation, The Cryosphere, 2, 77–93, https://doi.org/10.5194/tc-2-77-2008, 2008.

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

Hansen, P. C.: Analysis of Discrete Ill-Posed Problems by Means of the L-Curve, SIAM Review, 34, 561–580, https://doi.org/10.1137/1034115, 1992.

Hansen, P. C.: The L-Curve and Its Use in the Numerical Treatment of Inverse Problems, vol. 4, pp. 119–142, WIT Press, 2001.

Hernandez, V., Roman, J. E., and Vidal, V.: SLEPc: A Scalable and Flexible Toolkit for the Solution of Eigenvalue Problems, ACM Trans. Math. Softw., 31, 351–362, https://doi.org/10.1145/1089014.1089019, 2005.

Hernandez, V., Roman, J. E., Vidal, V., and A., T.: Krylov-Schur Methods in SLEPc, Tech. rep., Universidad Politecnica de Valencia, 2007.

Hill, E. A., Rosier, S. H. R., Gudmundsson, G. H., and Collins, M.: Quantifying the potential future contribution to global mean sea level from the Filchner–Ronne basin, Antarctica, The Cryosphere, 15, 4675–4702, https://doi.org/10.5194/tc-15-4675-2021, 2021.

Hock, R., Maussion, F., Marzeion, B., and Nowicki, S.: What is the global glacier ice volume outside the ice sheets?, Journal of Glaciology, p. 1–7, https://doi.org/10.1017/jog.2023.1, 2023.

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, 2015.

Jacobs, S. S., Jenkins, A., Giulivi, C. F., and Dutrieux, P.: Stronger ocean circulation and increased melting under Pine Island Glacier ice shelf, Nature Geoscience, 4, 519–523, https://doi.org/10.1038/ngeo1188, 2011.

Jay-Allemand, M., Gillet-Chaulet, F., Gagliardini, O., and Nodet, M.: Investigating changes in basal conditions of Variegated Glacier prior to and during its 1982–1983 surge, The Cryosphere, 5, 659–672, https://doi.org/10.5194/tc-5-659-2011, 2011.

Jenkins, A.: A Simple Model of the Ice Shelf-Ocean Boundary Layer and Current, Journal of Physical Oceanography, 46, 1785–1803, https://doi.org/10.1175/JPO-D-15-0194.1, 2016.

Jenkins, A., Shoosmith, D., Dutrieux, P., Jacobs, S., Kim, T. W., Le, S. H., Ha, H. K., and Stammerjohn, S.: West Antarctic Ice Sheet retreat in the Amundsen Sea driven by decadal oceanic variability, Nat. Geoscience, 11, 733–738, https://doi.org/https://doi.org/10.1038/s41561-018-0207-4DO, 2018.

Joughin, I., Smith, B., and Holland, D. M.: Sensitivity of 21st Century Sea Level to Ocean-Induced Thinning of Pine Island Glacier, Antarctica, Geophys. Res. Lett., 37, L20 502, https://doi.org/10.1029/2010GL044819, 2010.

Joughin, I., Smith, B. E., and Medley, B.: Marine Ice Sheet Collapse Potentially Under Way for the Thwaites Glacier Basin, West Antarctica, Science, 344, 735–738, https://doi.org/10.1126/science.1249055, 2014.

Kalmikov, A. G. and Heimbach, P.: A Hessian-Based Method for Uncertainty Quantification in Global Ocean State Estimation, SIAM Journal on Scientific Computing, 36, S267–S295, https://doi.org/10.1137/130925311, 2014.

Kazmierczak, E., Sun, S., Coulon, V., and Pattyn, F.: Subglacial hydrology modulates basal sliding response of the Antarctic ice sheet to climate forcing, The Cryosphere, 16, 4537–4552, https://doi.org/10.5194/tc-16-4537-2022, 2022.

Khazendar, A., Rignot, E., and Larour, E.: Acceleration and spatial rheology of Larsen C Ice Shelf, Antarctic Peninsula, Geophysical Research Letters, 38, https://doi.org/https://doi.org/10.1029/2011GL046775, 2011.

Koziol, C. P., Todd, J. A., Goldberg, D. N., and Maddison, J. R.: fenics_ice 1.0: a framework for quantifying initialization uncertainty for time-dependent ice sheet models, Geoscientific Model Development, 14, 5843–5861, https://doi.org/10.5194/gmd-14-5843-2021, 2021.

Lannelongue, L., Grealey, J., and Inouye, M.: Green Algorithms: Quantifying the Carbon Footprint of Computation, Advanced Science, 8, 2100 707, https://doi.org/https://doi.org/10.1002/advs.202100707, 2021.

Levermann, A., Winkelmann, R., Albrecht, T., Goelzer, H., Golledge, N. R., Greve, R., Huybrechts, P., Jordan, J., Leguy, G., Martin, D., Morlighem, M., Pattyn, F., Pollard, D., Quiquet, A., Rodehacke, C., Seroussi, H., Sutter, J., Zhang, T., Van Breedam, J., Calov, R., DeConto, R., Dumas, C., Garbe, J., Gudmundsson, G. H., Hoffman, M. J., Humbert, A., Kleiner, T., Lipscomb, W. H., Meinshausen, M., Ng, E., Nowicki, S. M. J., Perego, M., Price, S. F., Saito, F., Schlegel, N.-J., Sun, S., and van de Wal, R. S. W.: Projecting Antarctica's contribution to future sea level rise from basal ice shelf melt using linear response functions of 16 ice sheet models (LARMIP-2), Earth System Dynamics, 11, 35–76, https://doi.org/10.5194/esd-11-35-2020, 2020.

Lilien, D. A., Joughin, I., Smith, B., and Gourmelen, N.: Melt at grounding line controls observed and future retreat of Smith, Pope, and Kohler glaciers, The Cryosphere, 13, 2817–2834, https://doi.org/https://doi.org/10.5194/tc-13-2817-2019, 2019.

Lindgren, F., Rue, H., and Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73, 423–498, https://doi.org/https://doi.org/10.1111/j.1467-9868.2011.00777.x, 2011.

MacAyeal, D. R.: Large-scale ice flow over a viscous basal sediment: Theory and application to Ice Stream B, Antarctica, Journal of Geophysical Research, 94, 4071–4087, 1989.

Macayeal, D. R.: The Basal Stress Distribution of Ice Stream E, Antarctica, Inferred by Control Methods, JOURNAL OF GEOPHYSICAL RESEARCH, 97, 595–603, 1992.

MacAyeal, D. R., Bindschadler, R. A., and Scambos, T. A.: Basal friction of ice stream E, West Antarctica, Journal of Glaciology, 41, 247–262, https://doi.org/10.3189/S0022143000016154, 1995.

Maddison, J. R., Goldberg, D. N., and Goddard, B. D.: Automated Calculation of Higher Order Partial Differential Equation Constrained Derivative Information, SIAM Journal on Scientific Computing, 41, C417–C445, https://doi.org/10.1137/18M1209465, 2019.

Martin, J., Wilcox, L. C., Burstedde, C., and Ghattas, O.: A Stochastic Newton MCMC Method for Large-Scale Statistical Inverse Problems with Application to Seismic Inversion, SIAM Journal on Scientific Computing, 34, A1460–A1487, https://doi.org/10.1137/110845598, 2012.

Morales, J. L. and Nocedal, J.: Remark on "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound Constrained Optimization", ACM Trans. Math. Softw., 38, https://doi.org/10.1145/2049662.2049669, 2011.

Morlighem, M., Rignot, E., Seroussi, G., Larour, E., Ben Dhia, H., and Aubry, D.: Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica, Geophys. Res. Lett., 37, L14 502, https://doi.org/10.1029/2010GL043853, 2010.

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., Goel, V., Greenbaum, J. S., Gudmundsson, H., Guo, J., Helm, V., Hofstede, C., Howat, I., Humbert, A., Jokat, W., Karlsson, N. B., Lee, W. S., Matsuoka, K., Millan, R., Mouginot, J., Paden, J., Pattyn, F., Roberts, J., Rosier, S., Ruppel, A., Seroussi, H., Smith, E. C., Steinhage, D., Sun, B., Broeke, M. R. v. d., Ommen, T. D. v., Wessem, M. v., and Young, D. A.: Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet, Nature Geoscience, 13, 132–137, https://doi.org/10.1038/s41561-019-0510-8, 2020.

Mouginot, J., Rignot, E., Scheuchl, B., and Millan, R.: Comprehensive Annual Ice Sheet Velocity Mapping Using Landsat-8, Sentinel-1, and RADARSAT-2 Data, Remote Sensing, 9, https://doi.org/10.3390/rs9040364, 2017.

Pattyn, F.: Antarctic subglacial conditions inferred from a hybrid ice sheet/ice stream model, Earth and Planetary Science Letters, 295, 451–461, https://doi.org/https://doi.org/10.1016/j.epsl.2010.04.025, 2010.

Pattyn, F., Perichon, L., Aschwanden, A., Breuer, B., de Smedt, B., Gagliardini, O., Gudmundsson, G. H., Hindmarsh, R. C. A., Hubbard, A., Johnson, J. V., Kleiner, T., Konovalov, Y., Martin, C., Payne, A. J., Pollard, D., Price, S., Rückamp, M., Saito, F., Souček, O., Sugiyama, S., and Zwinger, T.: Benchmark experiments for higher-order and full-Stokes ice sheet models (ISMIP HOM), The Cryosphere, Volume 2, Issue 2, 2008, pp.95-108, 2, 95–108, 2008.

Petra, N., Martin, J., Stadler, G., and Ghattas, O.: A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems, Part II: Stochastic Newton MCMC with Application to Ice Sheet Flow Inverse Problems, SIAM Journal on Scientific Computing, 36, A1525–A1555, https://doi.org/10.1137/130934805, 2014.

Rignot, E., Mouginot, J., and Scheuchl, B.: Ice Flow of the Antarctic Ice Sheet, Science, 333, 1427–1430, https://doi.org/10.1126/science.1208336, 2011.

Rignot, E., Mouginot, J., and Scheuchl, B.: MEaSUREs InSAR-Based Ice Velocity of the Amundsen Sea Embayment, Antarctica, Version 1, https://doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0545.001, 2014.

Rignot, E., Mouginot, J., and Scheuchl, B.: MEaSUREs InSAR-Based Antarctica Ice Velocity Map, Version 2, https://doi.org/10.5067/D7GK8F5J8M8R, 2017.

Robel, A. A., Seroussi, H., and Roe, G. H.: Marine ice sheet instability amplifies and skews uncertainty in projections of future sea-level rise, Proceedings of the National Academy of Sciences, 116, 14 887–14 892, https://doi.org/10.1073/pnas.1904822116, 2019.

Scheuchl, B., Mouginot, J., Rignot, E., Morlighem, M., and Khazendar, A.: Grounding line retreat of Pope, Smith, and Kohler Glaciers, West Antarctica, measured with Sentinel-1a radar interferometry data, Geophysical Research Letters, 43, 8572–8579, https://doi.org/https://doi.org/10.1002/2016GL069287, 2016.

Schoof, C.: A variational approach to ice stream flow, Journal of Fluid Mechanics, 556, 227–251, https://doi.org/10.1017/S0022112006009591, 2006.

Seddik, H., Greve, R., Zwinger, T., and Sugiyama, S.: Regional modeling of the Shirase drainage basin, East Antarctica: full Stokes vs. shallow ice dynamics, The Cryosphere, 11, 2213–2229, https://doi.org/10.5194/tc-11-2213-2017, 2017.

Sergienko, O. V., MacAyeal, D. R., and Thom, J. E.: Reconstruction of snow/firn thermal diffusivities from observed temperature variation: application to iceberg C16, Ross Sea, Antarctica, 2004–07, Annals of Glaciology, 49, 91–95, https://doi.org/10.3189/172756408787814906, 2008.

Seroussi, H., Nakayama, Y., Larour, E., Menemenlis, D., Morlighem, M., Rignot, E., and Khazendar, A.: Continued retreat of Thwaites Glacier, West Antarctica, controlled by bed topography and ocean circulation, Geophysical Research Letters, pp. 6191—-6199, https://doi.org/10.1002/2017GL072910, 2017.

Shapero, D. R., Badgeley, J. A., Hoffman, A. O., and Joughin, I. R.: icepack: a new glacier flow modeling package in Python, version 1.0, Geoscientific Model Development, 14, 4593–4616, https://doi.org/10.5194/gmd-14-4593-2021, 2021.

1000 Still, H., Hulbe, C., Forbes, M., Prior, D. J., Bowman, M. H., Boucinhas, B., Craw, L., Kim, D., Lutz, F., Mulvaney, R., and Thomas, R. E.: Tidal Modulation of a Lateral Shear Margin: Priestley Glacier, Antarctica, Frontiers in Earth Science, 10, https://doi.org/10.3389/feart.2022.828313, 2022.

Stuart, A. M.: Inverse problems: A Bayesian perspective, Acta Numerica, 19, 451–559, https://doi.org/10.1017/S0962492910000061, 2010.

Tabeart, J. M., Dance, S. L., Lawless, A. S., Migliorini, S., Nichols, N. K., Smith, F., and Waller, J. A.: The impact of using reconditioned
1005 correlated observation-error covariance matrices in the Met Office 1D-Var system, Quarterly Journal of the Royal Meteorological Society, 146, 1372–1390, https://doi.org/https://doi.org/10.1002/qj.3741, 2020.

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

Thacker, W. C.: The role of the Hessian matrix in fitting models to measurements, J. Geophys. Res., 94, 6177–6196,
1010 https://doi.org/10.1029/JC094iC05p06177, 1989.

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

Tsai, C.-Y., Forest, C. E., and Pollard, D.: Assessing the contribution of internal climate variability to anthropogenic changes in ice sheet volume, Geophysical Research Letters, 44, 6261–6268, https://doi.org/https://doi.org/10.1002/2017GL073443, 2017.

1015 Waddington, E. D., Neumann, T. A., Koutnik, M. R., Marshall, H.-P., and Morse, D. L.: Inference of accumulation-rate patterns from deep layers in glaciers and ice sheets, Journal of Glaciology, 53, 694–712, https://doi.org/10.3189/002214307784409351, 2007.

Weertman, J.: On the Sliding of Glaciers, Journal of Glaciology, 3, 33–38, https://doi.org/10.3189/S0022143000024709, 1957.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization, ACM Trans. Math. Softw., 23, https://doi.org/10.1145/279232.279236, 1997.

Table 1. Prior strength configurations used in Sect. 5.1 and 5.3, based on the pointwise standard deviation $\sigma_{c(0)}$ and auto-covariance length scale $l_{c(0)}$ of each control parameter, ordered from weak to strong (except for the Cornford sliding law experiment).

Prior configurations	$\sigma_{\alpha(0)}$	$l_{\alpha(0)}$	γ_α	δ_α	$\sigma_{\beta(0)}$	$l_{\beta(0)}$	γ_β	δ_β
weak	1000	3000	0.85	9.40×10^{-8}	30	1000	9.4	9.40×10^{-6}
	510	2000	1.1	2.80×10^{-7}	30	1000	9.4	9.40×10^{-6}
$\downarrow$	**500**	**3000**	**1.7**	**1.90×10^{-7}**	**30**	**1000**	**9.4**	**9.40×10^{-6}**
	500	3000	1.7	1.90×10^{-7}	60	1000	4.7	4.70×10^{-6}
strong	150	3000	5.6	6.20×10^{-7}	30	1000	9.4	9.40×10^{-6}
Cornford sliding law	3000	3000	0.28	3.1×10^{-8}	30	1000	9.4	9.40×10^{-6}

The configuration in bold is also used in the experiments of Sect. 5.2 and 5.3. The units of $\sigma_{\alpha(0)}$ are $\mathrm{m}^{-1/6}$ $\mathrm{yr}^{1/6}$ $\mathrm{Pa}^{1/2}$ and $\sigma_{\beta(0)}$ are $\mathrm{Pa}^{1/2}$ $\mathrm{yr}^{1/6}$. The unit of the auto-covariance length scale $l_{c(0)}$ is reported in m. Following (18), the units of γ_α are $\mathrm{m}^{7/6}$ $\mathrm{yr}^{-1/6}$ $\mathrm{Pa}^{-1/2}$, γ_β are m $\mathrm{Pa}^{-1/2}$ $\mathrm{yr}^{-1/4}$, δ_α are $\mathrm{m}^{-5/6}$ $\mathrm{yr}^{-1/6}$ $\mathrm{Pa}^{-1/2}$ and δ_α are m^{-1} $\mathrm{Pa}^{-1/2}$ $\mathrm{yr}^{-1/4}$.
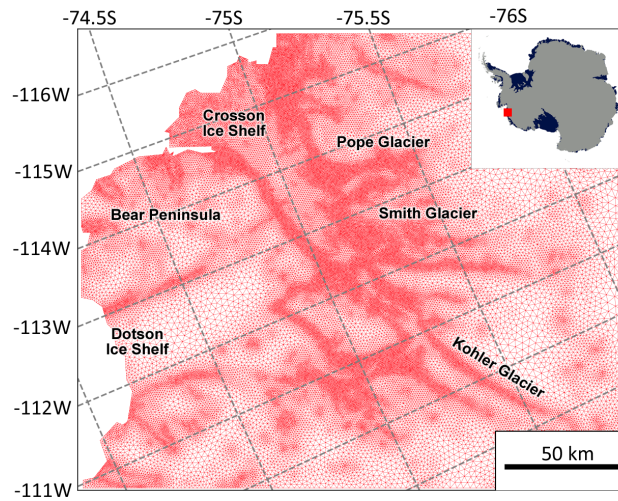
Figure 1. Variable resolution mesh of the ice streams region. The resolution depends on observed strain rates derived by using satellite velocity data (MEaSUREs v1.0 1996 - 2012, Rignot et al., 2014) and BedMachine Antarctica v2.0 (Morlighem et al., 2020). The boundaries to the East and South are entirely ice-ice boundaries, whereas the North and West features calving fronts where ice meets the ocean.
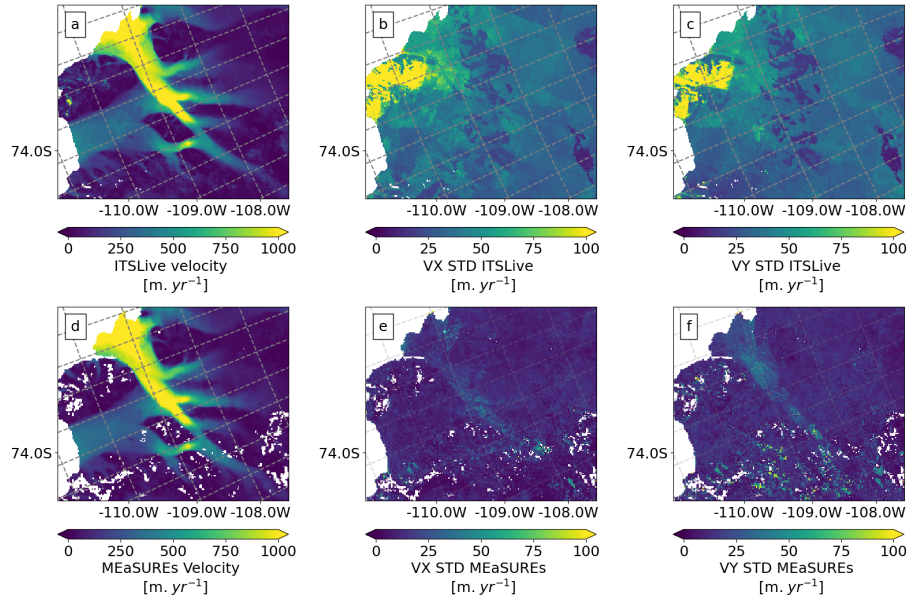
Figure 2. Observational input data. Satellite surface velocity observations (vector magnitude) and standard deviation (STD) of the velocity components (vx and vy) from ITS_LIVE (**a, b, and c.** Gardner et al., 2019, 2018) acquired from January to December 2014, and from MEaSUREs v2.0 (**d, e, and f.** Rignot et al., 2017) acquired from July 2013 to July 2014. For details on observational error model (STD estimates) see Sect. 3.1.3.
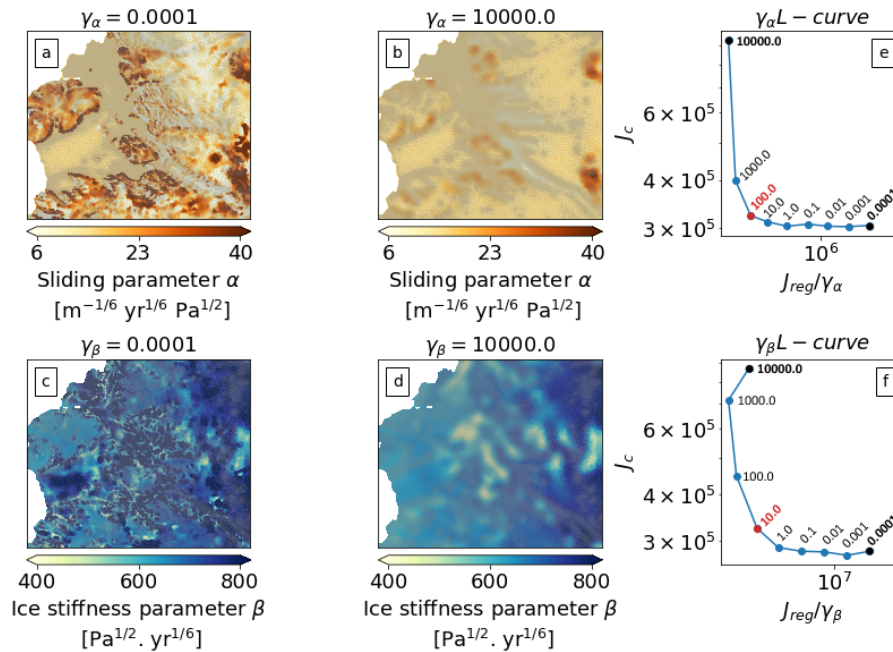
Figure 3. *L*-curve analysis output. **a and b:** Sliding parameter (α) computed using extreme γ_α values (bold values in panel e). **c and d:** Ice stiffness parameter (β) computed using extreme γ_β values (bold values in panel g). **e and g:** *L*-curve analysis for γ_α and γ_β, the optimal values suggested by the *L*-curves are highlighted in red.
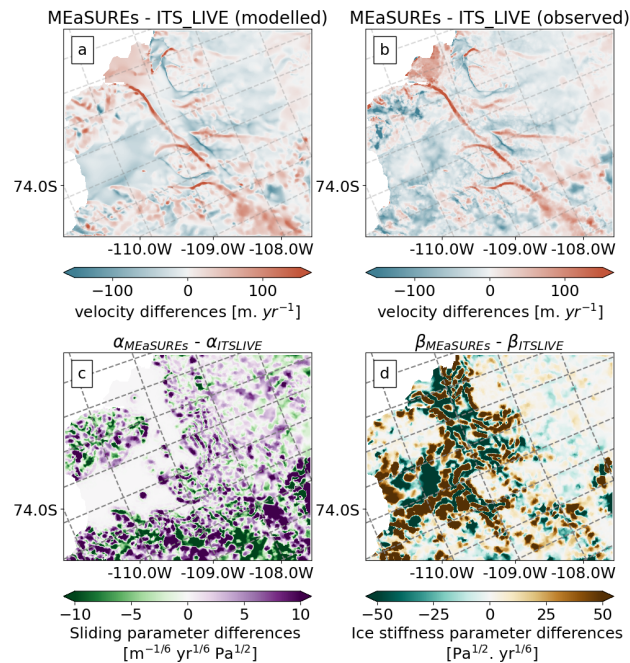
Figure 4. Inversion output differences after using two different satellite velocity products (MEaSUREs and ITSlive) to calibrate the ice dynamic parameters. **a:** Modelled velocity differences **b:** Observed velocity differences. **c:** Sliding parameter (α) differences. **d:** Ice stiffness parameter differences (β).
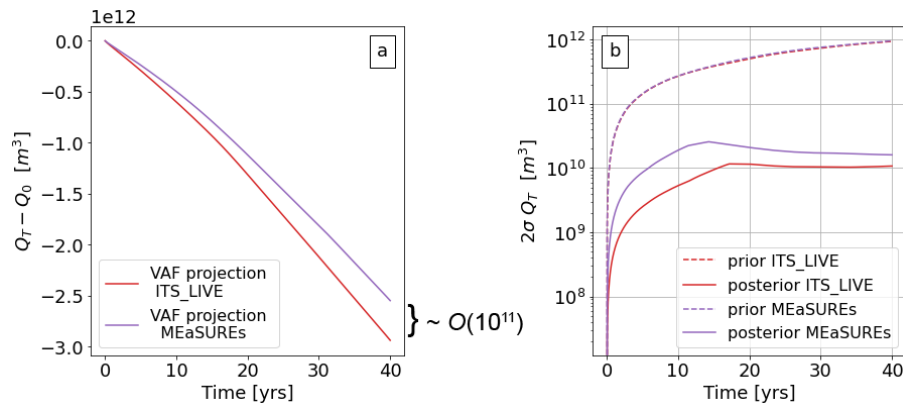
Figure 5. a Trajectories of change in VAF ~~trajectories~~ (i.e. $Q_T - Q_0$) using different velocity products and the regularisation terms suggested by the L-curves ($\gamma_\alpha = 100$, $\gamma_\beta = 10$). **b**: Hessian-based prior (dash lines) and posterior (solid line) uncertainties of VAF over time ($2\sigma Q_T$, represents 95% confidence interval).
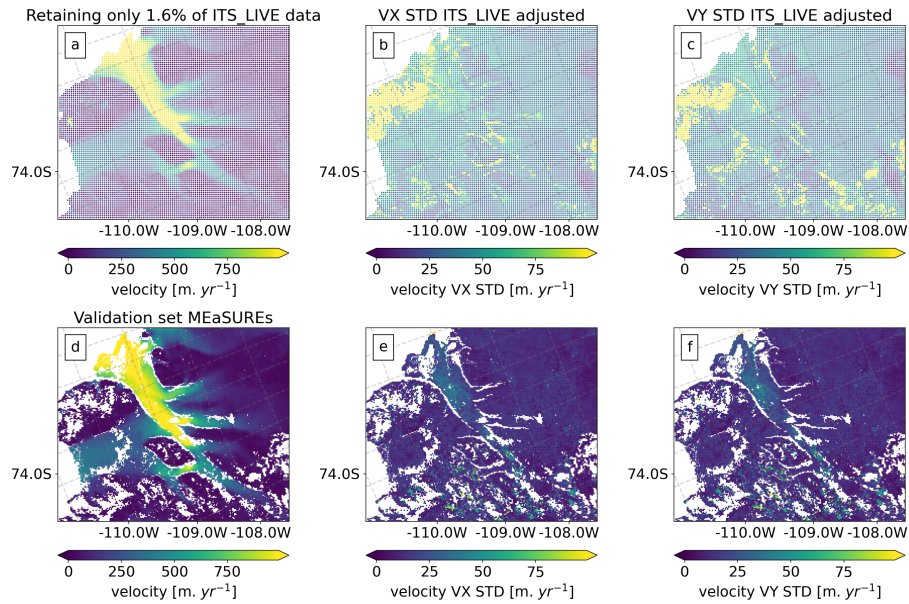
Figure 6. Overview of the input data used for the observational data subsampling test and for the experiments presented in Sect. 5. **a:** example of a training data set from ITS_LIVE where only 1.6% of the data points are retained, **b and c:** ITS_LIVE uncertainty in the x and y direction with the same data density as in (a) and with the STD of each component adjusted (see Sect. 5.2 for details). **d to f:** MEaSUREs validation data set used for the test.
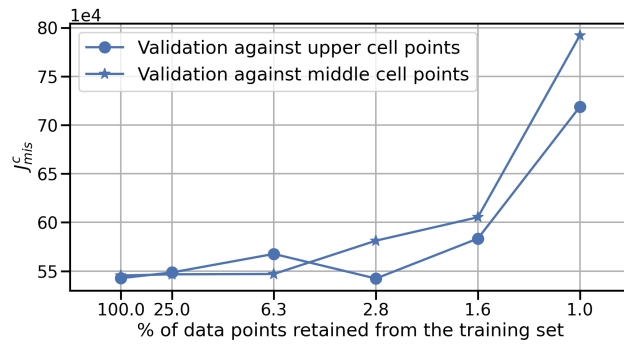
Figure 7. Observational data subsampling results. $\mathcal{J}^c$ $\mathcal{J}^c_{mis}$ performance if retaining a different number of observations for each training set.
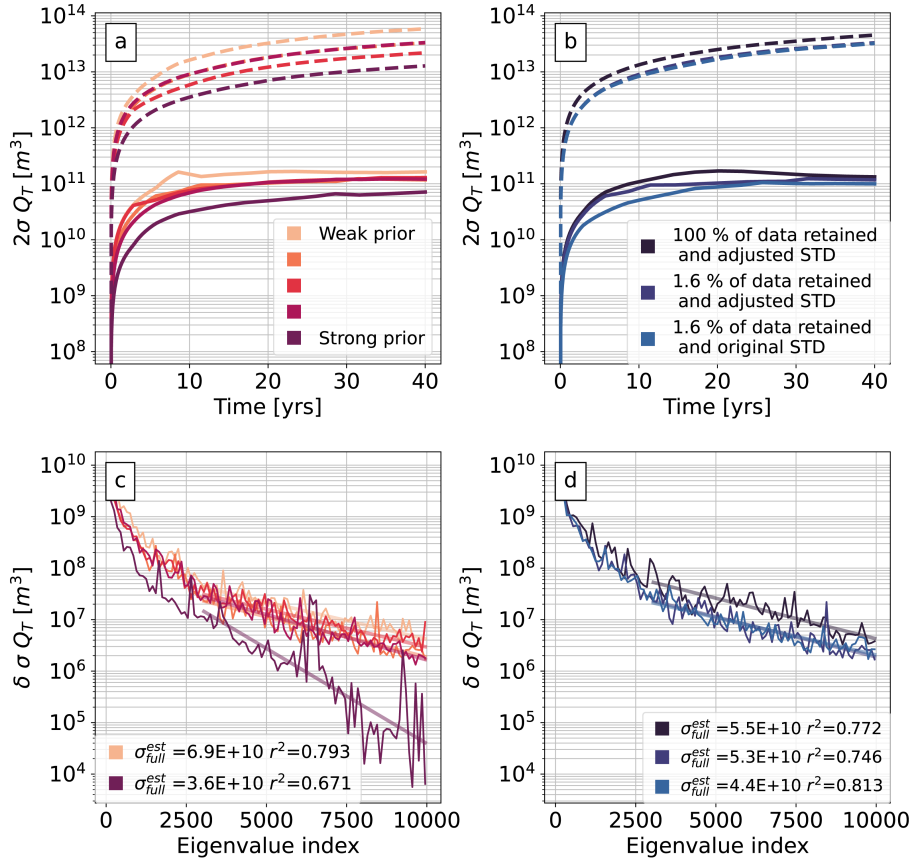
Figure 8. Hessian-based prior (dash lines) and posterior (solid line) uncertainties of VAF over time ($2\sigma Q_T$, represents 95% confidence interval) computed using; **a:** different strengths of prior and **b:** different versions of the ITS_LIVE data (i.e. different data density and STD, see details in Sect.5.2). **c and d**: Rate of change for the posterior uncertainty of VAF ($\delta\sigma Q_T$) against the number of eigenvalues calculated; statistics are shown in the lower corners, i.e. σ_{full}^{est} the estimated STD of VAF for an infinite number of eigenvalues and the decreasing trend coefficient of determination (r^2).
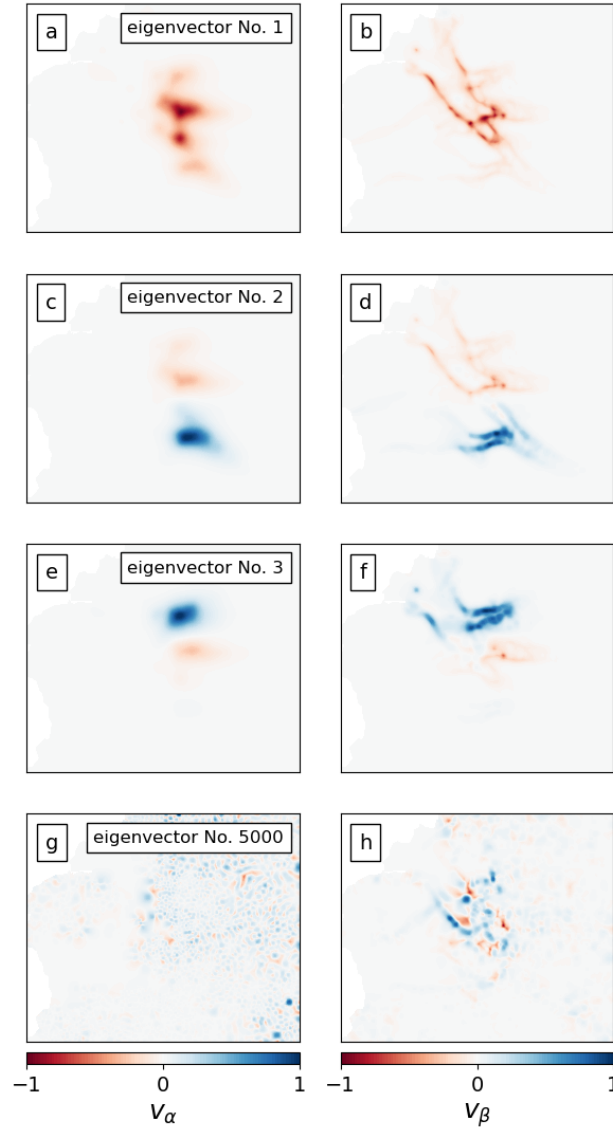
Figure 9. Eigenvectors (v) of the Hessian. Each eigenvector has an α component (right column) and a β component (second column). Each component is scaled to have a maximum magnitude of one~~and the scaling factor is shown in each vector legend~~. Ordered from large to small (from top to bottom), these v correspond to the 1st, 2nd, 3rd, and 5000th eigenvalues.
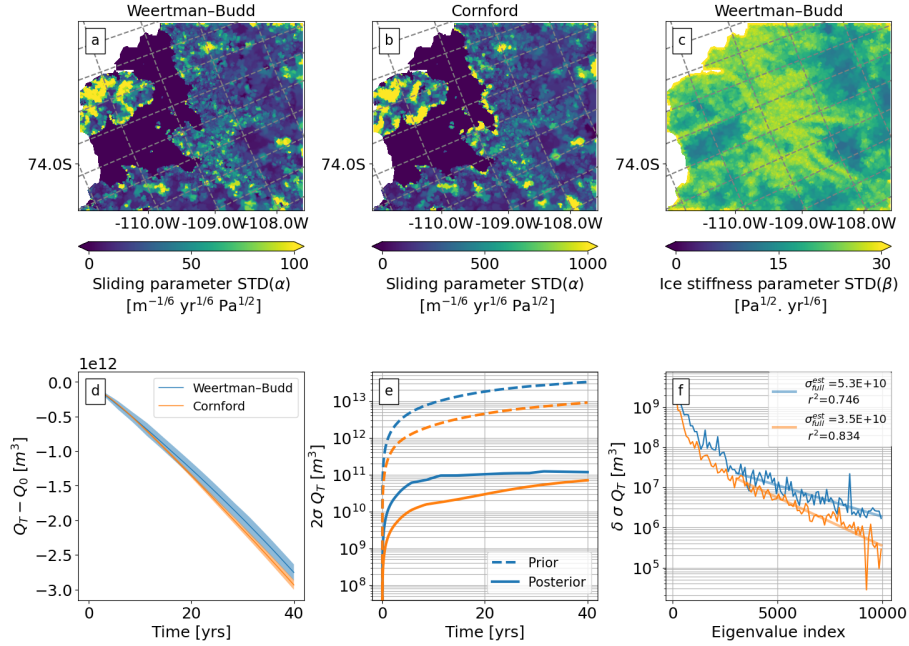
Figure 10. Model output when using two different sliding laws. Point-wise STD of the sliding parameter α for the **(a)** Weertman-Budd and **(b)** Cornford law. **c:** Point-wise STD of the ice stiffness parameter β (independent of the sliding law). **d:** ~~Trajectories of change in~~ VAF ~~trajectories~~ $(Q_T - Q_0)$ using the different sliding laws and the highlighted weak prior from Table 1. **e:** Hessian-based prior (dash lines) and posterior (solid lines) uncertainties of VAF over time ($2\sigma Q_T$, represents 95% confidence interval). **f:** Rate of change for the posterior uncertainty of VAF ($\delta\sigma Q_T$) against the number of eigenvalues calculated; statistics are shown in the lower corners, i.e. σ_{full}^{est} the estimated STD of VAF for an infinite number of eigenvalues and the decreasing trend coefficient of determination (r^2).
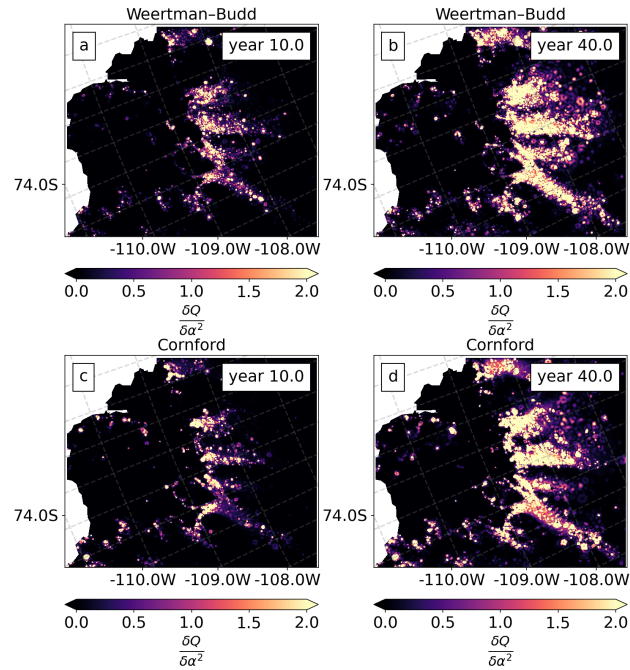
Figure 11. Sensitivity maps of the model's change in volume above flotation (Q_T) to the basal friction coefficient α^2; for year 10 and year 40, when using two different sliding laws. Units of α^2 are $\mathrm{m}^{-1/3}\ \mathrm{yr}^{1/3}$ Pa. **a and b:** Weertman-Budd and **c and d:** Cornford law. These visualise the node-wise sensitivity given the choice of mesh and finite element discretisation.
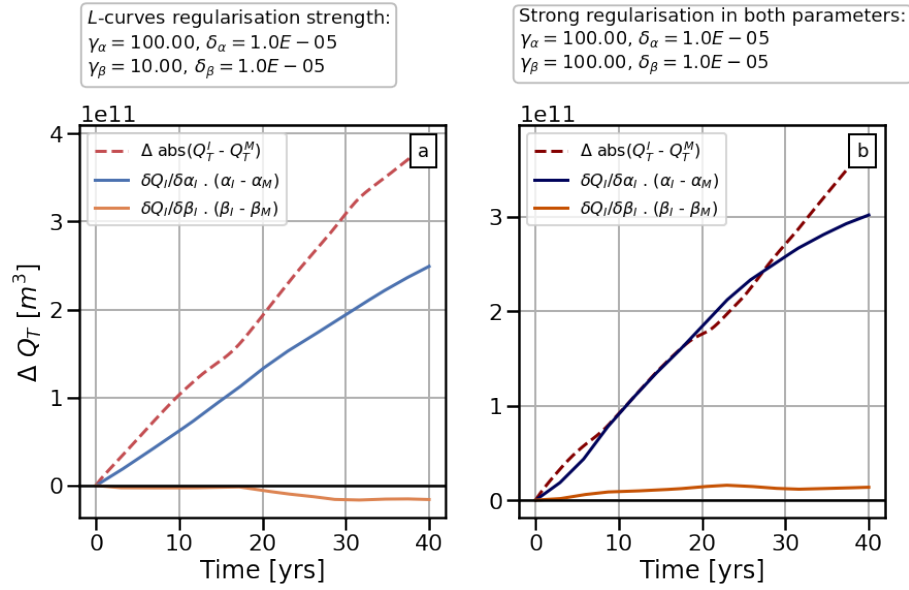
Figure 12. Linearity test results. Absolute difference of ~~VAF~~ trajectories of VAF (Q_T) estimated using different satellite velocity products (red dotted lines) and dot products results (solid lines) from (23) and (24) plotted in blue and yellow respectively. For this figure we use model output from Sect. 4.2. **a:** Linearity test using the regularisation strength suggested by the L-curves and **b:** Linearity test using a stronger regularisation on β.

Table A1. *L*-curves configurations used in Sect. 4.1. Reference values for β parameters are obtained from variance values computed from the STD of the ice stiffness (B) initial guess (see details in Sect. 3). Reference values for α parameters (first guess) are chosen heuristically and looking to maintain the linearity of the VAF operator and stable solutions of the time-dependent SSA.

Prior configurations	$\sigma_{\alpha(0)}$	$l_{\alpha(0)}$	γ_α	δ_α	$\sigma_{\beta(0)}$	$l_{\beta(0)}$	γ_β	δ_β
γ_α *L*-curves	-	-	-	1.9×10^{-7}	30	1000	9.4	9.4×10^{-6}
γ_β *L*-curves	65	23000	100	1.9×10^{-7}	-	-	-	9.4×10^{-6}
δ_α *L*-curves	-	-	100	-	30	1000	9.4	9.4×10^{-6}
ITS_LIVE and MEaSUREs *L*-curve workflow	8.9	3200	100	1×10^{-5}	28	1000	10	1×10^{-5}

The "-" indicates that those parameters were vary by several orders of magnitude while constructing the *L*-curves. The units of $\sigma_{\alpha(0)}$ are $\mathrm{m}^{-1/6} \, \mathrm{yr}^{1/6} \, \mathrm{Pa}^{1/2}$ and $\sigma_{\beta(0)}$ are $\mathrm{Pa}^{1/2} \, \mathrm{yr}^{1/6}$. The unit of the auto-covariance length scale $l_{c(0)}$ is reported in m. Following (18), the units of γ_α are $\mathrm{m}^{7/6} \, \mathrm{yr}^{-1/6} \, \mathrm{Pa}^{-1/2}$, γ_β are m $\mathrm{Pa}^{-1/2} \, \mathrm{yr}^{-1/4}$, δ_α are $\mathrm{m}^{-5/6} \, \mathrm{yr}^{-1/6} \, \mathrm{Pa}^{-1/2}$ and δ_α are $\mathrm{m}^{-1} \, \mathrm{Pa}^{-1/2} \, \mathrm{yr}^{-1/4}$.
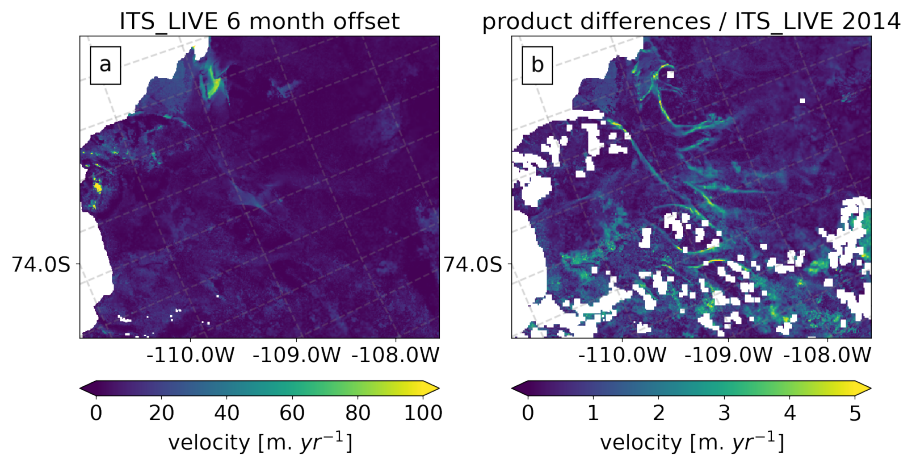
Figure A1. Speed comparisons. **a:** 6 month offset speed change from ITS_LIVE 2014 (Jul-Dec) and **b:** Speed ratio of the difference between the two products (ITS_LIVE 2014 - MEaSUREs acquired from July 2013 to July 2014) and ITS_LIVE 2014. Empty pixels in panel b are due to gaps in the MEaSUREs dataset, which increase when the data is interpolated to the ITS_LIVE grid.
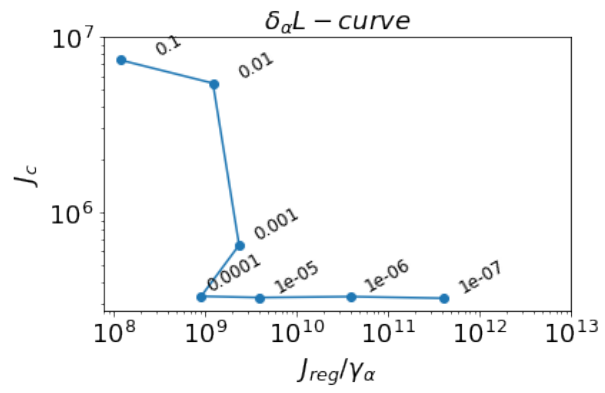
Figure A2. Sensitivity maps of the model's volume above flotation (VAF) to the ice stiffness (B or β^2)L-curve for δ_α.Units of B are $^{1/3}$
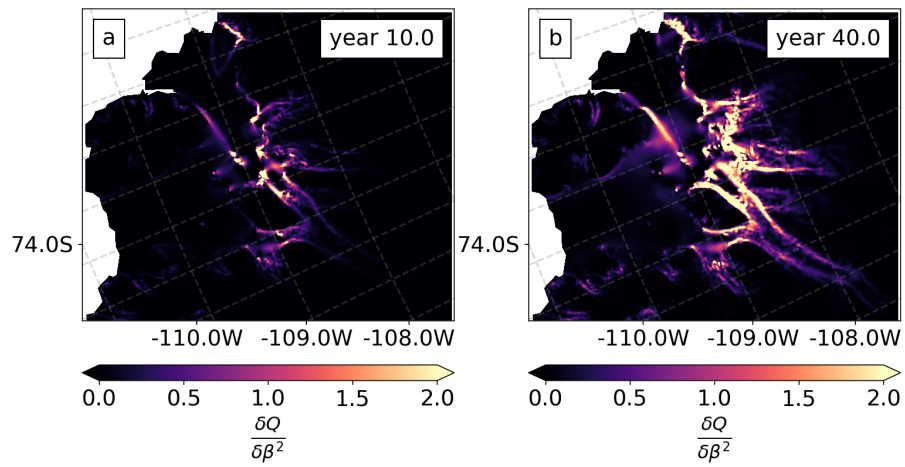
Figure A3. Sensitivity maps of the model's change in volume above flotation (Q_T) to the ice stiffness (B or β^2). Units of B are $\mathrm{Pa\ yr}^{1/3}$
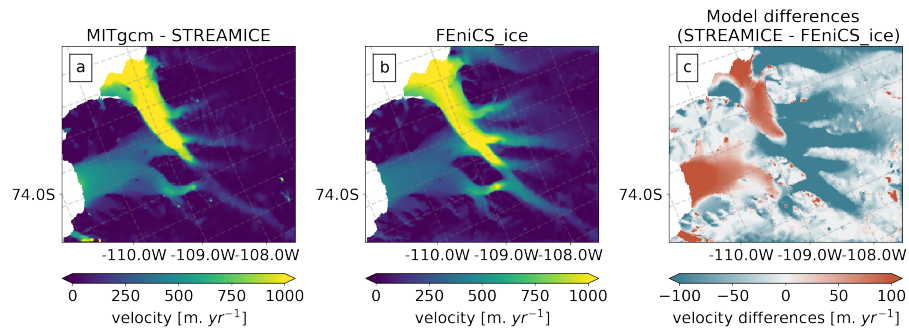
Figure A4. Models surface velocity comparison. Surface velocities are calculated by using `FEniCS_ice` inversions of α and β calibrated with ITS_LIVE and the highlighted weak prior configuration from Table 1. **a:** STREAMICE. **b:** `FEniCS_ice`. **c:** Difference between both models.