The authors would like to express their thanks to both referees for providing such thoughtful comments on our manuscript -- and also to the Associate Editor for overseeing our manuscript.

We apologise for the length of time required for our response and resubmission. The delay was primarily as a result of our implementation of the Gauss-Newton Hessian approximation as asked after by Rev. 2, which was nontrivial but we feel makes a contribution to the manuscript and to the code.

Line numbers referenced below refer to the revised version (not the difference report or the original version)

Daniel N Goldberg on behalf of authors

Comments from Reviewer 1

The uncertainty in model data is propagated to a Quantity of Interest (QoI) in ice sheet simulations using a Bayesian approach. This kind of analysis has been published previously but the authors extend it here to time dependent QoIs. They show how to reduce the uncertainty in the prior by utilizing Bayes' formula for the posterior. The method is tested with the SSA ice sheet equations solving an ISMIP-HOM problem. Parameters are varied in numerical computations. Examples of possible future work are given in the final section. The code for the experiments is available for free.

We thank the reviewer for this synopsis. We also point out that (to our knowledge) the application of Algorithmic DIfferentiation to this problem (UQ of ice-sheet initialisation) is new.

The paper is suitable for GMD and is improved if the comments below are addressed somehow in a revised version.

General comments

Sect. 2.4: Γ_{post} depends on how r is chosen. If there is a gap in the eigenvalue distribution then r can be chosen such that the gap is between λ_r and λ_{r+1}. But in general there is no gap. How should r be chosen in a general case? This question is related to the choice of γ, see Fig 4a. If we believe in prior data then γ should be large but if we don't then give data lower weight (or should γ be viewed only as a regularization parameter?).

We agree the “choosing” or r was a point not addressed. In this work the problems are small enough for us to recover the entire spectrum but in general this needs to be decided. Isaac et al (2015) suggest a strategy of ensuring lambda_r is O(1). However a more pragmatic strategy we are considering is to choose r so that eq (22) (the

variance of the QoI) has negligible change as r increases. We mention both strategies now.

Choosing an appropriate value for the prior parameter γ is in general very important beyond the scope of this idealised study. We explain (line 347) that, in line with other non-probabilistic inverse modelling studies, we carry out an L-curve analysis, but only as a guideline for choosing values of \gamma to examine. In general, it does have meaning, and we feel the background sections of the manuscript express this sufficiently.

Sect. 4, (33): Why is this particular QoI chosen? A bit longer motivation would be welcome.

We now give more motivation for this choice -- and also "call forward" to a passage that is (now) in the discussion (line 562) about the choice of QoI and its sensitivity to small scale variability.

Sect. 4: How is the prior c_0 chosen? Maybe this is mentioned somewhere but it could be repeated here. A discussion of how to select c_0 and its impact on the posterior result would be interesting.

We now say:

"In our experiments C_0, the prior value of C, is uniformly zero -- indicating we assume no preconceieved notion of its mean value, only its spatial variability (implied by γ)."

It is beyond the scope of this idealised experiment, but in general C_0 should depend on independent assessments e.g. the bed elevation R may be sought as an unknown, but independent assessments exist. However, we do not mention this as we do not introduce additional unknown fields in our simple idealised framework.

Sect. 5.3: One would expect that the linear approximation of QoI should work for sufficiently small perturbations. Maybe the perturbations are too large when the regularization is small (γ_1) and for smaller perturbations the approximation will work.

Yes, we agree -- and actually we feel this is echoed in our passage: "It is likely that the small-scale noise inherent in the low-regularisation samples (\textit{cf} Fig. 8) impacts the Quantity of Interest strongly enough that the linear approximation in Eq. 22 breaks down "

Sect. 6: The issues above with choice of r, γ, c_0 and QoI could also be discussed in the last section.

Thank you for the suggestion. We have decided to address these issues in the relevant sections of the methodology (see passages and line number references above)

Specific comments

line 93: Euclidean inner product of a with $\Gamma^{-1}_{obs}b$? Next line $\|a\|_{\Gamma^{-1}_{obs}}$?

Thank you, fixed!

105: Is there a weight missing in front of the prior term? Maybe γ?

This is as we intended it. Parameters such as γ are meant to be included in \Gamma_{prior} ie. see eq 11.

110: Tell that you maximize over $\bar{c}$.

Done (though actually we minimize the negative log posterior)

145: Define δ. $\matcal{L}$ is defined for a function in (9). In (10) $\matcal{L}$ is applied to a vector c, at least c is a vector after the last equality in (10). Should J_{reg}^c depend on a weight too?

Thank you for pointing out this is undefined, we now make clear that γ and δ are scalars (and thus $\delta(\cdot)$ implies multiplication). You also raise an important point, γ and δ in general can vary spatially, though in the present they are treated as constant. We now write:

"where γ and δ are positive scalars which are in general spatially varying, though in the present study we consider only constants."

These "weights" however are absorbed into the definition of $\mathbf{L}$.

149: Should the definition of L have only one $\mathcal{L}$ in the integral? What are the bars over ϕ?

You are correct, apology for the oversight which is fixed. As described in 2.1, the overbar is our convention for a coefficient vector of a finite element function.

171: Mention that the eigenvalues are ordered such that $\lambda_i\ge\lambda_{i+1}$.

Done.

179: leading eigenmodes -> leading eigenmodes with large eigenvalues.

With our implementation of the previous suggestion (that the eigenvalue ordering is in descending order), this is now implied, so we do not make this change.

242: Specify parameters B and n

We are not sure if you meant to define these or give the values we used, but we now do both (lines 255-257)

251, 253: Specify δ which is different from δ in (9). Tell what o.w is.

Apologies, \delta is a symbol commonly used for this quantity (now defined just after the equation), we have changed to \eta which is not used elsewhere in the paper, and defined it (line 265).

o.w. = otherwise. We have simply written "otherwise"

321, 322: Is weight γ missing here? Is c a function in the integral after the first equality and a vector in the second term after the second equality?

γ is "absorbed" into the definition of the prior covariance (eq 11).

As for the formatting of c this was an oversight -- thank you. In the first equality it is a finite element function and we use the capital C as previously in the subsection. In the second it is a coefficient vector and we use \overline{c} (cf line 338).

515: The uncertainty in the QoI is not only lower than the uncertainty in the parameters due to the filtering but also due to the choice of QoI. With a different QoI it may be larger even with filtering.

This is a good point -- and we feel it is now addressed by the paragraph which *now* follows this one (though, as your correctly point out below, we had forgotten to move to the discussion).

Technical corrections

line 138: J_c -> J^c

fixed.

176: H -> $\bar{H}_{mis}$?

We believe this is correct as written. We have added additional equalities to make this clearer.

223: (Section ?)

fixed.

302, 304: C^2 or C here?

C^2 is what was intended. Eqn (30) is consistent with eq (24). Eq 31 is modified to make clear we are defining C.

385: say something about ..... intervals?

Apologies, we overlooked this leading up to submission -- it is now removed.

413: maybe move ....?

We overlooked this but now it is moved and addresses nicely a comment you make above.

665, 677: missing journals

Comments from Reviewer 2

This paper presents a mathematical method and a software tool for quantifying uncertainty in glacier flow model projections. The methods consists of (1) using automatic differentiation to compute the action of the full Hessian of the negative log posterior, (2) computing the smallest few eigenvalue / eigenvector pairs of the Hessian in order to find out which directions in parameter space are most important to sample, and (3) computing the uncertainty in a given output quantity of interest by linearizing the model physics around the most probable state. The authors then test this method on a synthetic test problem, the ISMIP-HOM test case. The authors also address two key shortcomings of how data assimilation in glaciology is usually practiced: accounting for discrete (as opposed to dense) spatial observations and correlation between measurements. Overall, the method is described fairly well albeit with a few points that could use clarifying. I have a few concerns about how generalizable the method is, but in any case the paper and the software are a valuable contribution that I recommend for publication with minor revisions.

Many thanks for your supportive comments and concise summary.

Regarding generalisability, we cannot address this point within the context of our manuscript, though since the time of its writing we have been applying the methodology to a more realistic domain and setting. Still, we are glad to hear that even with the idealised experiments here you feel it is a valuable contribution.

The authors make excellent use of low-rank approximations to the Hessian. This trick has appeared in the literature before (although it's not as widely used as it should be). How does their approach compare to that of, say, Petra et al. 2014 or the hIPPYlib code?

In truth the approach is very similar to that of Petra 2014 (which we had neglected to cite, thank you for raising this), which actually follows from their "Part 1" paper, Bui-Thanh et al 2013. The two approaches are very similar and differ subtly, but not profoundly, and essentially do the same thing. We prefer the formulation of Isaac et al 2015, which we follow (but also make explicit some ideas not in their manuscript, such as eq 17). We have already stated that our analysis follows Isaac 2015, but we now state:

"The following low-rank approximation follows from \cite{Isaac2015} and similar approaches are used in \cite{BuiThanh2013} and \cite{petra2014}"

I have two real concerns with the approach, although these don't change my overall opinion of this very good paper. First, the method relies on the assumption of quasi-linearity in several places. While the authors check that this assumption wasn't violated for their particular test case, it's difficult to assess whether this would generalize to other problems. By contrast, the stochastic Newton approach in Petra et al. 2014 uses an assumption of quasi-linearity only locally, to bootstrap a more sophisticated Monte Carlo sampling algorithm.

We agree that this approach is limited by the assumption of Gaussianity (ie. quasi-linearity) and we do caveat with the fact this assumption is made, in several places. It is worth noting that Petra 2014's Stochastic Newton MCMC method actually does rely on a low-rank Hessian approximation, which is a central part of the methods discussed in our manuscript as well. While we did not mention Petra 2014 originally, there is a mention of Martin et al 2012, who applied Stochastic Newton MCMC to seismic inversion -- though this was a bit buried in line 422 (of the original manuscript). We now cite Petra 2014 here as well.

The argument we make is that if one wanted to carry out MCMC on a state space this large, they would likely need to use a method such as that of Martin et al 2012 (or Petra et al 2014), which involves (as you mention) a local Hessian based approximation to the posterior pdf -- but even so, these methods are beyond the scope of the current study. To make this more clear we now add text to the discussion making this clear, and note that Stochastic Newton MCMC using our framework could be a future possibility.

Second, the authors make a big deal about using the full Hessian instead of the Gauss-Newton approximation while at the same time relying on quasi-linearity, and yet the Gauss-Newton approximation is works best when the dynamics are almost linear. At the end of the paper, the authors state that they did not use a time-dependent control method

because it's computationally expensive. Would a time-dependent method have been feasible if the authors had instead used the Gauss-Newton approximation? Establishing whether the full Hessian is really necessary is a very important point. Other researchers might want to emulate the techniques described in this paper and yet they might be using other modeling frameworks for which the Gauss-Newton approximation is feasible to implement while the full Hessian is not.

Thank you for this. Based on your feedback a decision was made to implement the Gauss-Newton Hessian within fenics_ice. We now have an additional section before the Discussion explaining its construction and providing results of a simple test, as well as a short paragraph in the discussion regarding the implications for future work -- but we stress that the results do not necessarily extend to more realistic settings.

General comments

45, "gradient-based optimisation": The point of this sentence is to state that computing the MAP estimate doesn't give any information about parametric uncertainty. The fact that you used a gradient-based algorithm concerns more the "how" than the "what"; it isn't really important and you could just cut this phrase entirely. You could have used a derivative-free optimization algorithm -- it's a dreadfully awful idea, but you could do it!

We agree that the MAP point does not tell us anything about uncertainty, we would not completely agree that this is the point we are making -- a point which we feel is addressed in the following paragraph regarding the seemingly contradictory aims of large-scale and UQ. Therefore we feel the efficiency of control methods in finding the MAP point relative to MC methods is in fact relevant. We remove the phrase, now stating

“Although control methods might efficiently provide estimates of parameter fields, they do not provide parametric uncertainty.”

57-60: How much better is using the full Hessian than using the Gauss-Newton approximation? This isn't immediately apparent from the text or from the sources you cite here.

We do not know of another study which addresses this question in the context of UQ. As mentioned, though, we now have a section devoted to this. We of course cannot address this question generally but we can and do for our experiments.

97: By taking the parameter-to-observation map f to be a map from R^n to R^m, you're assuming a "discretize, then optimize" mindset. It might clarify the points you make about mesh dependence later on if you instead define it as a map from some function space Q to R^m -- the "optimize, then discretize" mindset. See Gunzburger 2002.

We appreciate what you are saying here, but at the same time feel that our approach is very different from the "discretize, then optimize" approach taken by, e.g. the line-by-line differentiation of ocean and climate models by source-to-source transformation (e.g. Kalmikov and Heimbach 2014) -- we mention this because this is the type of approach which "discretize, then optimize" typically refers to. Since tlm_adjoint differentiates at the level of finite element variational forms, at least in simple cases discretization and algorithmic differentiation commute, in which case the approaches are equivalent.

Regardless, though, we believe our approach in this paper -- to consider functions (and their higher-order derivatives) in terms of coefficient vectors -- greatly simplifies things relative to trying to express relationships as differential forms, which we feel would add complexity to the text and make it more difficult for interested audiences to follow. We are very careful to use distinct notation for finite element coefficient vectors to distinguish from the actual functions, and feel that we have struck a balance between discrete and continuous.

The parameter-to-observation map that you've written down encapsulates both the physics and the measurement process. This is just a suggestion, but it might help the exposition to instead define f as the composition of two maps. First, there's a function g that takes the parameters to the observable fields, like the ice velocity. This function g is basically just "solve the shallow shelf equations". Next, there's a function h that takes the observable fields to the actual observations. When doing the "wrong thing" that you point out later, h is the identity map and the norm is an L^2 norm. When doing the correct that that you have actually implemented, h evaluates the observable fields at a bunch of discrete points and packs these observations into a vector, and the model-data misfit is a discrete sum of squared errors.

Thank you for noting this, as this is a subtle point in the manuscript not made explicit. Introducing another symbol altogether would add complexity to a number of expressions which we feel outweighs the benefits of making this idea clear symbolically. We add a paragraph now in section 4 where Numerical Experiments are described:

“Our parameter-to-observable map $\vobs{f}$ is really a composition of two functions: the first finds the solution to the momentum balance (Eq. \ref{eq:weak_form}) as a finite-element function, and the second interpolates the function to discrete locations. If the misfit cost were to be expressed as the weighted L_2 norm of the model-data misfit as in Eq. \ref{eq:control_mis}, then the interpolation function is replaced by the identity.”

135-139: Using a control method does not necessarily imply that you're writing the model-data misfit as a squared L^2 norm, it's just a sinful thing that many glaciologists (including me) have done because it's easier.

We agree with this, and we did not mean to imply that all control methods do this -- rather those which are specifically referenced in the introduction do. To avoid ambiguity we now write

“By contrast with Bayesian methods, the control methods generally used in glaciological data assimilation” .. and cite a subset of those cited in the introduction (to avoid length/repetition)

It's also worth noting that when you say "mesh dependence", certain readers are immediately going to think of something other than what you describe here. In the PDE-constrained optimization literature, mesh dependence refers to what happens when you use a bad optimization algorithm based on using the vector of coefficients of the derivative obtained from the adjoint method as a descent direction. This can give really obvious mesh imprinting artifacts in the results, especially with higher-order finite element bases. (The minimal right thing to do is to multiply by the inverse of the mass matrix. You mention using the BFGS method later, and taking the H_0 matrix to be the inverse mass matrix works there as well.) This problem is more a question of how you solve a particular optimization problem. The mesh dependence that you're talking about is much more serious -- by neglecting the discreteness of the observations, going to a finer mesh implicitly assumes that you magically have more measurements than you did before. At a higher level, what you've done is tackle the fact that everyone has been solving the wrong problem, irrespective of how they were solving it. Since some readers will immediately associate with the first case, it might be good to either (1) clarify the distinction with a reference to, say, Schwedes et al. 2017, or, (2) if you don't feel like talking about that, use a different phrase besides "mesh dependence".

In fact, by mesh dependence we actually intend the meaning to be with respect to the computational mesh, i.e. how things change (or do not change) when the mesh itself is resolved and finite-element degrees of freedom increase -- which we believe is distinct from the two types of mesh dependence you mention. In the experiment you refer to the observational spacing and location remains the same (we believe this is clear, since Figs 4c and 4d relate to the refinement of the observation array.) We would like to retain “mesh (in)dependence” as it was the same wording used in Isaac et al 2015 (their figure 5), and we now cite this paper where it is mentioned.

140-146: Including the delta term is essentially adding the prior information that you think the mean of the parameters is zero. I don't think this is a good prior in all cases. Are there other ways to get a prior with bounded covariance that don't make this assumption? For example, you could use the Moreau-Yosida regularization of bounds constraints, which

instead assumes that the parameters don't wander too far outside a preset interval but which provides no constraints within that interval.

We agree it is not the only way to impose this prior mean, and agree it is perhaps not the best approach in all cases, but as mentioned in our response to referee 1 we feel that a prior mean of zero is appropriate. The elliptic operator we use for a prior does have strong advantages though such as the ease of inverting which needs to be done many times. We point out this is not the only way to impose this and reference an optimisation study which uses the regularisation method you mention.

260-264: It's easy to get the impression from this paragraph that you're doing time-dependent inversions, which is only dispelled in the conclusion on line 552. You should probably state this earlier in the text.

We now add " In this study, we do not consider initialisations based on time-varying data (i.e. the misfit cost function J_{mis}^c does not depend on time-varying fields), so the continuity function is only involved with finding a Quantity of Interest and propagation of initialisation uncertainty."

270-273: How do you know how many Picard iterations is adequate? Why not use another globalization strategy, like damping / line search or trust regions?

We do not know (or specify) a priori a set number of picard iterations, rather we specify a relative tolerance, which is chosen empirically. We now state this as well as the value used for our study.

274: Why did you use BFGS when you can calculate a Hessian-vector product? Why not Newton-Krylov?

Full Hessian-vector multiplication costs the tangent-linear + second order adjoint, as well as the usual forward + first order adjoint, so as a baseline would be more expensive, and we are not sure that a Newton-Krylov approach would on the whole offer an advantage. While we have not carried out a detailed study of Newton-Krylov methods for optimisation of the cost function, results with a Newton-CG approach were not very promising. Given this, we feel the passage is best kept as-is.

341: The L-curve is fine, but you might want to mention the discrepancy principle or other more statistically-motivated approaches. See Habermann et al. 2013.

We now note that there are alternative approaches to determining the optimal level of regularisation and cite a few references which implement these, including Habermann 2013.

484-490: This is really great, I haven't seen anyone address this issue before.

Thank you, to our knowledge it had not been examined in the context of this problem.

537-538: I don't think the technical hurdles are minor at all because of exactly what you say in the next sentence.

I would remove this statement from the text.

We meant to say that it is relatively trivial to add another "control" field into the fenics_ice framework and implement Hessian-vector products with respect to this joint space -- not to overcome the issue of equifinality and/or quantify overfitting (in fact this is very difficult). To avoid confusion we modify to

"The version of \texttt{fenics_ice} presented in this study is not capable of joint inversions or of Hessian-vector products with multiple parameter fields, however the technical hurdles to implementation are minor."

550-551: It might be worth citing some of the work that Karen Willcox and her group have done on multi-fidelity modeling and UQ.

What we were suggesting was much less sophisticated than a multifidelity approach but we do now add a reference.

Technical corrections

Several of the authors' "note to self" comments remain in the manuscript.

Thank you, these were overlooked and now actioned.

484: "isotroptically" -> "isotropically"

Thank you, fixed.

References

Bui-Thanh, T., Ghattas, O., Martin, J., & Stadler, G. (2013). A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, *35*(6), A2494-A2523.

Isaac, T., Petra, N., Stadler, G., & Ghattas, O. (2015). Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *Journal of Computational Physics*, *296*, 348-368.

Kalmikov, A. G., & Heimbach, P. (2014). A Hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM Journal on Scientific Computing*, *36*(5), S267-S295.

Martin, J., Wilcox, L. C., Burstedde, C., & Ghattas, O. (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, *34*(3), A1460-A1487.

Petra, N., Martin, J., Stadler, G., & Ghattas, O. (2014). 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*(4), A1525-A1555.

fenics_ice 1.0: A framework for quantifying initialisation uncertainty for time-dependent ice-sheet models

Conrad Koziol[1], Joe Todd[1], Daniel N Goldberg[1], and James R Maddison[2,3]

[1]School of GeoSciences, Univ. of Edinburgh, City of Edinburgh, United Kingdom
[2]School of Mathematics, Univ. of Edinburgh, City of Edinburgh, United Kingdom
[3]Maxwell Institute for Mathematical Sciences, Univ. of Edinburgh, City of Edinburgh, United Kingdom

Correspondence: D N Goldberg (dan.goldberg@ed.ac.uk)

Abstract.

Mass loss due to dynamic changes in ice sheets is a significant contributor to sea level rise, and this contribution is expected to increase in the future. Numerical codes simulating the evolution of ice sheets can potentially quantify this future contribution. However, the uncertainty inherent in these models propagates into projections of sea level rise, and hence is crucial to understand. Key variables of ice sheet models, such as basal drag or ice stiffness, are typically initialized using inversion methodologies to ensure that models match present observations. Such inversions often involve tens or hundreds of thousands of parameters, with unknown uncertainties and dependencies. The computationally intensive nature of inversions along with their high number of parameters mean traditional methods such as Monte Carlo are expensive for uncertainty quantification. Here we develop a framework to estimate the posterior uncertainty of inversions, and project them onto sea level change projections over the decadal timescale. The framework treats parametric uncertainty as multivariate Gaussian, and exploits the equivalence between the Hessian of the model and the inverse covariance of the parameter set. The former is computed efficiently via algorithmic differentiation, and the posterior covariance is propagated in time using a time-dependent model adjoint to produce projection error bars. This work represents an important step in quantifying the internal uncertainty of projections of ice-sheet models.

1 Introduction

The dynamics of ice sheets are strongly controlled by a number of physical properties which are difficult (or intractable) to observe directly, such as basal traction and ice stiffness (Arthern et al., 2015). This poses challenges in terms of future ice-sheet projections, as the behaviour of ice sheets often depends strongly on these (spatially varying) properties. There are two principal approaches that have been taken by ice-sheet modellers to approach these challenges: control methods and sampling-based uncertainty quantification. Below, we discuss these approaches in the context of ice-sheet modelling.

Control methods (MacAyeal, 1992), sometimes referred to simply as "inverse methods" in a glacial flow-modelling context, consist of the minimisation of a "cost" function involving some global measure of model-data misfit, as well as regularisation cost terms which penalise nonphysical behaviour (e.g. high variability at small scales or strong deviation from prior knowledge). A strong benefit of control methods is their ability to estimate hidden properties at the grid scale through large-scale

optimisation techniques. Such methods have been used extensively to calibrate ice-sheet models to observations (e.g., Rommelaere, 1997; Vieli and Payne, 2003; Larour et al., 2005; Sergienko et al., 2008; Morlighem et al., 2010; Joughin et al., 2010; Fürst et al., 2015; Cornford et al., 2015).

Uncertainty quantification (UQ) in projections of ice-sheet behaviour is a crucial challenge in ice-sheet modelling. Studies of fast-flowing Antarctic glaciers have shown that uncertainties in the parameters controlling ice flow can lead to large variability in modelled behaviour (Nias et al., 2016). Thus it is of great importance to quantify how this parametric uncertainty translates into uncertainty in projections. In some cases, this uncertainty may be exogenous to the dynamics of the ice sheet model: for instance, uncertainty in ocean-driven ice shelf melt, while a likely important contributor to ice-sheet projection uncertainty (Robel et al., 2019), arises from incomplete knowledge of the ocean system rather than the dynamics of the ice model itself. This is in contrast to parameters that must be constrained via calibration; their uncertainties derive from observational uncertainty, uncertainty in model physics, and *a priori* knowledge.

The uncertainty associated with ice-sheet model calibration can be quantified through Bayesian inference, in which prior knowledge is "updated" with observational evidence. Such methods have been applied to continental-scale ice-sheet models and models of coupled ice-ocean interactions (Gladstone et al., 2012; Ritz et al., 2015; Deconto and Pollard, 2016). In these Bayesian studies, the dimension of the parameter space is small (i.e. less than $\sim$20). Though the methods of these studies differ, they share the common feature of generation of a large ensemble (thousands of runs) through sampling of a parameter space. Bayesian methods are then applied in conjunction with observational data to find likelihood information for the parameters, and associated probability distributions of ice-sheet behaviour.

Applying such ensemble-based Bayesian methods to glacial flow models and parameter sets of dimension $\sim \mathcal{O}(10^4 - 10^6)$ (a dimension size typical of control methods) is prohibited by computational expense. Although control methods ~~efficiently might~~ might efficiently provide estimates of parameter fields, they do not provide parametric uncertainty. While it can be shown that such methods provide the *most likely* parameter field (often referred to as the Maximum A Posteriori, or *MAP*, estimate) (Raymond and Gudmundsson, 2009; Isaac et al., 2015), the *covariance* of the joint probability distribution – necessary for assessing uncertainty in calibrated model behaviour at the MAP point – cannot be inferred.

Thus, there is at present a disconnect between the dual aims of (i) modelling ice sheets as realistically as possible, i.e. through the implementation of higher-order stresses and without making limiting assumptions regarding "hidden" properties of the ice sheet, and (ii) uncertainty quantification (UQ) of models by approximate inference by reducing the dimensionality of the set of parameters.

By augmenting control methods using a Hessian-based Bayesian approach, it is possible to quantify parametric uncertainty without sacrificing parameter dimension or model fidelity. Just as control methods can be interpreted as returning the *mode* of a joint posterior probability distribution, it can be shown that, under certain assumptions, the *covariance* of the distribution can be characterised by the inverse of the *Hessian* (the matrix of second derivatives) of the cost function with respect to the parameters (Thacker, 1989; Kalmikov and Heimbach, 2014; Isaac et al., 2015). For a nonlinear model, calculating the Hessian involves model second derivatives with respect to parameters, which can be challenging for complex models; in many cases, second-derivative information is ignored and the Hessian is approximated using first-derivative information only (Kaminski

et al., 2015); such an approximation is referred to as the Gauss-Newton Hessian (Chen, 2011). Some studies retain second-derivative information, however, using variational methods (Isaac et al., 2015) or Algorithmic Differentiation (AD) software (Kalmikov and Heimbach, 2014).

Once determined, the Hessian-based parameter covariance can then be used to quantify the variance of a scalar Quantity of Interest (QoI) of the calibrated model (e.g., ice-sheet sea level contribution over a specified period). One approach to this is projecting the parameter covariance on to a linearised model prediction (e.g., Kalmikov and Heimbach, 2014). Isaac et al. (2015) employ this methodology in a finite-element ice flow model, but since their model is time-independent, uncertainty estimates cannot be projected forward in time.

In this study we introduce a framework for time-dependent ice-sheet uncertainty quantification, and apply it to an idealised ice-sheet flow problem (Pattyn et al., 2008). Beginning with a cost-function optimisation for sliding parameters given noisy ice-sheet velocity data, we then generate a low-rank approximation to the posterior covariance of the sliding parameters through the use of the cost-function Hessian. In our work, the Hessian is calculated through AD, using the "complete" Hessian rather than the Gauss-Newton approximation. We then project the covariance on a linearisation of the time-dependent ice-sheet model (again using AD to generate the linearisation) to estimate the growth of QoI uncertainty over time. We also apply a method of sampling the posterior distribution, and use this to validate our calculation of time-dependent QoI uncertainty for an idealised problem.

2 Methodology

2.1 Symbolic convention

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 $\overline{c}$; general vectors and vector-valued functions as $\breve{d} \in \mathbb{R}^q$; and matrices as $\mathbf{E}$.

2.2 Mathematical Framework

An ice-sheet flow model can be thought of as a (nonlinear) mapping from a set of input fields, which might be unobservable or poorly known (such as bed friction) to a set of output fields, which might correspond to observable quantities (such as surface velocity). Here, our focus is on the probability distribution function (PDF) of a "hidden" field C *conditioned* on an observational field U, i.e. $p(C|U)$; and our aim is to determine properties of this conditional distribution through Bayes' theorem:

$$p(C|U) = \frac{p(U|C)p(C)}{p(U)}. \tag{1}$$

$p(U)$, the unconditional distribution of observations, is effectively a normalisation constant which we do not consider further.

As described in Section 3, our ice-sheet flow model is a finite-element model, meaning C can be described by a vector of finite dimension. We furthermore consider discrete observations, meaning U can be described by a finite-dimensional vector

as well (in general with different dimension from C). We assume that observational errors follow a Gaussian distribution. Referring to the vector of observations as $\breve{u}_{obs} \in \mathbb{R}^m$, this is expressed as

$$-\log\left(p(\breve{u}_{obs})\right) = \frac{1}{2}\langle \breve{u}_{obs} - \breve{u}_{true}, \breve{u}_{obs} - \breve{u}_{true}\rangle_{\mathbf{\Gamma}_{obs}^{-1}} \equiv \frac{1}{2}\|\breve{u}_{obs} - \breve{u}_{true}\|^2_{\mathbf{\Gamma}_{obs}^{-1}}. \tag{2}$$

Here, $\langle \breve{a}, \breve{b}\rangle_{\mathbf{\Gamma}_{obs}^{-1}}$ is the Euclidean inner product of $\breve{a}$ with ~~$\mathbf{\Gamma}_{obs}\breve{b}$~~ $\mathbf{\Gamma}_{obs}^{-1}\breve{b}$, where $\mathbf{\Gamma}_{obs} \in \mathrm{Sym}^+(m)$ (the set of real symmetric positive definite $m \times m$ matrices) is the observational *covariance matrix*, and ~~$\|\breve{a}\|_{\mathbf{\Gamma}_{obs}}$ is the associated norm~~ $\|\breve{a}\|_{\mathbf{\Gamma}_{obs}^{-1}}$ is the norm associated with its inverse. If the parameter field is represented by the vector $\bar{c} \in \mathbb{R}^n$, then the conditional PDF $p(U|C)$ satisfies

$$-\log\left(p(\breve{u}_{obs}|\bar{c})\right) = \frac{1}{2}\|\breve{u}_{obs} - \breve{f}(\bar{c})\|^2_{\mathbf{\Gamma}_{obs}^{-1}}, \tag{3}$$

where $\breve{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a function from the space of parameter fields to the space of observations, i.e. our ice-sheet flow model. Note that the above construction equates $\breve{f}(\bar{c})$ with the "truth", i.e. it assumes zero model error. In general model error is extremely difficult to constrain, and doing so is beyond the scope of our study; however, in Section 7 we discuss potential strategies to incorporate model error into our framework.

The distribution $p(C)$ in Eq. 1 is the *prior* PDF of $\bar{c}$, which expresses knowledge of C prior to consideration of ice-sheet observations and physics – for instance, the autocorrelation scale of basal friction, which may be inferred from proxies such as the presence of basal water inferred from ice-penetrating radar. If the prior PDF is Gaussian, then the distribution of $\bar{c}$ conditioned on $\breve{u}_{obs}$ satisfies

$$-\log\left(p(\bar{c}|\breve{u}_{obs})\right) = \frac{1}{2}\|\breve{u}_{obs} - \breve{f}(\bar{c})\|^2_{\mathbf{\Gamma}_{obs}^{-1}} + \frac{1}{2}\|\bar{c} - \bar{c}_0\|^2_{\mathbf{\Gamma}_{prior}^{-1}}, \tag{4}$$

where $\bar{c}_0$ is the prior mean and $\mathbf{\Gamma}_{prior} \in \mathrm{Sym}^+(n)$ is the prior covariance. This conditional distribution is referred to as the *posterior* distribution, or p_{post}. If $\breve{f}$ is linear, p_{post} is Gaussian, with mean μ and covariance $\mathbf{\Gamma}$ given by

$$\begin{aligned}
\mu_{post,lin} =& \mathbf{\Gamma}_{post,lin}\left(\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right)^T \mathbf{\Gamma}_{obs}^{-1}(\breve{u}_{obs} - \breve{f}_0) + \mathbf{\Gamma}_{post,lin}^{-1}\bar{c}_0\right), \\
\mathbf{\Gamma}_{post,lin} =& \left(\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right)^T \mathbf{\Gamma}_{obs}^{-1}\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right) + \mathbf{\Gamma}_{prior}^{-1}\right)^{-1}.
\end{aligned} \tag{5}$$

(The above can be derived by ~~maximizing~~ minimizing Eq. 4 with respect to $\bar{c}$ with $\breve{f} = \breve{f}_0 + (\partial \breve{f}/\partial \bar{c}_0)(\bar{c} - \bar{c}_0)$.)

Models of ice-sheet dynamics are in general nonlinear, however, and Eq . 5 does not strictly apply. Instead we use a quadratic approximation to the negative log posterior (Bui-Thanh et al., 2013; Isaac et al., 2015; Kalmikov and Heimbach, 2014). Such an approximation considers a second-order Taylor expansion of $-\log(p_{post})$ about the mode of the posterior, or equivalently about the Maximum a Posteriori (MAP) estimate $\bar{c}_{MAP}$. This leads to a Gaussian distribution with mean $\bar{c}_{MAP}$ and covariance

$$\mathbf{\Gamma}_{post} = \left(\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right)^T \mathbf{\Gamma}_{obs}^{-1}\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right) + \mathbf{\Gamma}_{prior}^{-1} + \left(\frac{\partial^2 \breve{f}}{\partial \bar{c}^2}\right)\mathbf{\Gamma}_{obs}^{-1}(\breve{u}_{obs} - \breve{f}(\bar{c}))\right)^{-1}. \tag{6}$$

Eq. 6 differs from the covariance given by Eq. 5 in that derivatives of $\breve{f}$ depend on $\overline{c}_{MAP}$, and in the final term involving second derivatives of $\breve{f}$. Essentially, p_{post} is approximated by the Gaussian distribution with the local covariance at $\overline{c}_{MAP}$. While this is insufficient to calculate global properties of p_{post} such as skew, it gives insight into the directions in parameter space which are most (and least) constrained – information which can be propagated to model projections.

2.3 Relation to control methods

By contrast with Bayesian methods, the control methods ~~discussed in Section 1~~ generally used in glaciological data assimilation (Morlighem et al., 2010; Joughin et al., 2010; Cornford et al., 2015) find the parameter set which gives the best fit to observations. This is done by minimizing 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 the square-integral of the misfit between the surface velocity of the ice model and remotely-sensed observations, normalised by the observational error. These terms are discretised to implement the control method. If the ice-sheet model is solved via a finite element scheme, then the misfit cost can be written

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

Here $\breve{u}$ and $\breve{u}_{obs}$ are nodal values of the finite-element representations of modelled and observed velocities; $\mathbf{D}_\sigma$ is a diagonal matrix containing standard errors of the $\breve{u}_{obs}$ measurements; and $\mathbf{M}$ is the *mass matrix* corresponding to the finite element basis ϕ_i: $M_{ij} = \int_\Omega \phi_i \phi_j dA$, where Ω is the computational domain. J_{reg}, the regularisation cost, is imposed to prevent instabilities, and is generally 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). In other words, regularisation imposes smoothness on the control parameter field, which otherwise may exhibit variability at scales not strongly determined by the observations. Such a term can generally be written as a positive definite quadratic form of $\overline{c}$.

J^c is thus a functional with a form similar to Eq. 4, i.e. the negative log posterior. In this sense, solving the control problem is equivalent to finding $\overline{c}_{MAP}$. However, there are important differences between J^c_{mis} and the first term of Eq. 4. The former is an L_2 inner product (which, with standard continuous finite elements, introduces mesh dependent factors in the covariance) while the latter is an inner product involving values at a fixed set of observation points (which does not). Identifying ~~J_c~~ J^c as a negative log posterior therefore implies observational errors that are changed by factors related to grid cell areas.

Our framework effectively uses a control method – but one which allows calculation of the posterior covariance after the MAP point is found. As such we use a fixed set of points, as described above, in our misfit cost term. Thus, the Hessian of the cost function of our control method is equal to the inverse of the posterior covariance given by Eq. 6. However, our form of J^c_{reg} does not involve the square integral of the gradient of $\overline{c}$, as Bui-Thanh et al. (2013) note this can lead to unbounded prior covariances as the numerical grid is refined. These authors recommend a discretization of a differential operator of the form

$$\mathcal{L}(\cdot) \equiv \gamma\nabla^2(\cdot) - \delta(\cdot) \tag{9}$$

where ~~the~~ γ and δ are positive scalars which are in general spatially varying, though in the present study we consider only constants. The second term on the right hand side ensures the operator is invertible; though there are other ways of doing this (e.g., Keuthen and Ulbrich, 2015), it is a computationally simple approach. Isaac et al. (2015) use the same definition for their prior, which we adopt in our study as well. Hence, our regularisation term is

$$J^c_{reg} = \int_\Omega \frac{1}{2}(\mathcal{L}(c))^2 dA = \frac{1}{2}\|c\|^2_{\mathbf{L}\mathbf{M}^{-1}\mathbf{L}} \tag{10}$$

where $\mathbf{L}$ is the operator on the finite element space such that ~~$\overline{\phi}_i^T \mathbf{L}\overline{\phi}_j = \int_\Omega \mathcal{L}(\phi_i)\mathcal{L}(\phi_j)dA$~~ $\overline{\phi}_i^T \mathbf{L}\overline{\phi}_j = \int_\Omega \phi_i \mathcal{L}(\phi_j)dA$ for all ϕ_i, ϕ_j. Thus, in the Bayesian interpretation of the control method optimisation, the prior covariance is given by

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

2.4 Low rank approximation

In the previous section we establish that the posterior covariance is equivalent to the inverse of the Hessian of the (suitably defined) cost function. With a large parameter space, though, calculating the complete Hessian (and its inverse) can become computationally intractable. Still, in many cases, the constraints on parameter space provided by observations can be described by a subspace of lower dimension. In the present study, our idealised examples are small enough that the full Hessian can be calculated; but to provide scalable code we seek an approximation to the posterior covariance that exploits this low-rank structure.

The following low-rank approximation follows from Isaac et al. (2015) and similar approaches are used in Bui-Thanh et al. (2013) and Petra et al. (2014). We define the term

$$\left(\frac{\partial \breve{f}}{\partial \overline{c}}\right)^T \mathbf{\Gamma}^{-1}_{obs}\left(\frac{\partial \breve{f}}{\partial \overline{c}}\right) + \left(\frac{\partial^2 \breve{f}}{\partial \overline{c}^2}\right)\mathbf{\Gamma}^{-1}_{obs}(\breve{u}_{obs} - \breve{f}(\overline{c}))$$

from Eq. 6 as $\mathbf{H_{mis}}$, the Hessian of the misfit component of the negative log posterior (or, equivalently, of the misfit cost term). Eq. 6 can be written

$$\mathbf{\Gamma}_{post} = \left(\mathbf{H}_{mis} + \mathbf{\Gamma}^{-1}_{prior}\right)^{-1}. \tag{12}$$

This can be rearranged:

$$\mathbf{\Gamma}_{post} = \left(\mathbf{\Gamma}_{prior}\mathbf{H}_{mis} + \mathbf{I}\right)^{-1}\mathbf{\Gamma}_{prior}. \tag{13}$$

The term $\tilde{\mathbf{H}}_{mis} \equiv \mathbf{\Gamma}_{prior}\mathbf{H}_{mis}$ is referred to as the "prior-preconditioned Hessian", and it has the eigendecomposition

$$\tilde{\mathbf{H}}_{mis} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1} \tag{14}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues and $\mathbf{C}$ contains the corresponding eigenvectors. $\tilde{\mathbf{H}}_{mis}$ is not in general symmetric positive semidefinite (even though $\mathbf{H}_{mis}$ and $\mathbf{\Gamma}_{prior}$ both are), but Eq. 14 can be written as

$$\mathbf{\Gamma}_{prior}^{\frac{1}{2}}\mathbf{H}_{mis}\mathbf{\Gamma}_{prior}^{\frac{1}{2}} = \mathbf{\Gamma}_{prior}^{-\frac{1}{2}}\mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}\mathbf{\Gamma}_{prior}^{\frac{1}{2}} \tag{15}$$

i.e. an eigendecomposition of the symmetric matrix $\mathbf{\Gamma}_{prior}^{\frac{1}{2}}\mathbf{H}_{mis}\mathbf{\Gamma}_{prior}^{\frac{1}{2}}$. Thus the eigenvalues in $\mathbf{\Lambda}$ are real-valued, and the eigenvectors $\mathbf{C}$ can be chosen to be $\mathbf{\Gamma}_{prior}^{-1}$-orthogonal, i.e. such that

$$\mathbf{C}^T\mathbf{\Gamma}_{prior}^{-1}\mathbf{C} = \mathbf{I}. \tag{16}$$

While $\mathbf{H}_{mis}$ could be eigendecomposed directly, decomposing $\tilde{\mathbf{H}}_{mis}$ better informs uncertainty quantification~~: for an~~. We assume an ordering of the eigenvalues λ_i such that $\lambda_{i+1} \leq \lambda_i$. For an eigenvector $\bar{c}_k$ with eigenvalue λ_k, the negative log posterior probability density evaluated at $\bar{c} = \bar{c}_k + \bar{c}_{MAP}$ is

$$\begin{aligned}
\langle \bar{c}_k, \mathbf{H}\bar{c}_k\rangle &= \langle \bar{c}_k, (\mathbf{H}_{mis} + \mathbf{\Gamma}_{\mathbf{prior}}^{-\mathbf{1}})\bar{c}_k\rangle \\
&= \langle \bar{c}_k, (\mathbf{\Gamma}_{prior}^{-1}\tilde{\mathbf{H}}_{mis} + \mathbf{\Gamma}_{\mathbf{prior}}^{-\mathbf{1}})\bar{c}_k\rangle \\
&= \lambda_k\langle \bar{c}_k, \mathbf{\Gamma}_{prior}^{-1}\bar{c}_k\rangle + \langle \bar{c}_k, \mathbf{\Gamma}_{prior}^{-1}\bar{c}_k\rangle \\
&= (1+\lambda_k)\langle \bar{c}_k, \mathbf{\Gamma}_{\mathbf{prior}}^{-\mathbf{1}}\bar{c}_k\rangle \\
&= (1+\lambda_k).
\end{aligned} \tag{17}$$

In other words, the leading eigenmodes of $\tilde{\mathbf{H}}_{mis}$ correspond to those directions in which the posterior uncertainty is reduced by the most, relative to the prior uncertainty in those directions. Thus one can truncate the eigendecomposition, neglecting eigenmodes for which the data provides minimal information. The Sherman–Morrison–Woodbury matrix inversion lemma gives

$$\mathbf{\Gamma}_{post} = (\mathbf{I} - \mathbf{C}\mathbf{D}\mathbf{C}^{-1})\mathbf{\Gamma}_{prior} \tag{18}$$

where $\mathbf{D}$ is a diagonal matrix with entries $d_{kk} = \lambda_k/(1+\lambda_k)$, and with Eq. 16 this becomes

$$\mathbf{\Gamma}_{post} = \mathbf{\Gamma}_{prior} - \mathbf{C}\mathbf{D}\mathbf{C}^T. \tag{19}$$

This can then be approximated

$$\mathbf{\Gamma}_{post} \sim \mathbf{\Gamma}_{prior} - \mathbf{C}_r\mathbf{D}_r\mathbf{C}_r^T. \tag{20}$$

where $\mathbf{C}_r$ represents the first r columns of $\mathbf{C}$ and similarly for $\mathbf{D}_r$~~, with the rationale of neglecting posterior information in~~.

In this study, the ~~directions where it has minimal effect~~problems considered are sufficiently small that we calculate all eigenvalues, i.e. we do not carry out a low-rank approximation. In general, though, a strategy for deciding r is needed. Isaac et al. (2015) recommends choosing r such that $\lambda_r \ll 1$, which may in some cases result in a large value for r. A more pragmatic approach would be to choose r such that Eq. 22, the QoI variance (see Section 2.5) has negligible change when approximating with additional eigenvalue/eigenvector pairs.

2.5 Propagation of errors

Often of interest is how the observational data constrains outputs of a calibrated model, as opposed to how they constrain the calibrated parameters themselves. (A simple analogy is an extrapolation using a regression curve, which is generally of more interest than the regression parameters.) Such an output is termed a *Quantity of Interest* (QoI) Q, an example of which is the loss of ice volume above floatation (VAF), the volume of ice that can contribute to sea level, at a certain time horizon. Here we write $Q_T(\overline{c})$ to indicate the value of Q based on the output of the calibrated model at time horizon T.

The distribution of Q_T can be assessed by sampling from the posterior distribution of $\overline{c}$, although such sampling might be slow to converge. Alternatively an additional linear assumption can be made. Neglecting higher-order terms, Q_T can be expanded around $\overline{c}_{MAP}$:

$$Q_T = Q_T(\overline{c}_{MAP}) + \left(\frac{\partial Q_T}{\partial \overline{c}}\right)(\overline{c} - \overline{c}_{MAP}). \tag{21}$$

As this is an affine transformation of a Gaussian random variable, Q_T has a mean of $Q_T(\overline{c}_{MAP})$ and a variance of

$$\sigma^2(Q_T) = \left(\frac{\partial Q_T}{\partial \overline{c}}\right)^T \mathbf{\Gamma}_{post} \left(\frac{\partial Q_T}{\partial \overline{c}}\right) \tag{22}$$

If $\frac{\partial Q_T}{\partial \overline{c}}$ can be found at a number of times T along a model trajectory, then the growth of uncertainty along this trajectory arising from parametric uncertainty can be assessed.

Note the assumption of linearity in Eq. 21 is in general false due to the nonlinear momentum and mass balance equations that define a time-dependent ice-sheet model. For the idealised experiments conducted in this paper, we compare the above estimate for the variance with that derived from sampling the posterior.

3 Numerical approach

In this study we use a new numerical code, `fenics_ice`. `fenics_ice` is a Python code which implements the time-dependent Shallow Shelf Approximation (SSA; MacAyeal (1989)). The SSA is an approximation to the complete Stokes stress balance thought to govern ice flow. In the approximation the vertical stress balance is assumed to be hydrostatic, such that normal stress is in balance with the weight of the ice column. Additionally, flow is assumed to be depth-independent. These approximations reduce a three-dimensional saddle-point problem to a two-dimensional convex elliptic problem, enabling a more efficient solve. The nonlinear power-law rheology of the full Stokes problem is retained however. Despite these simplifications, the SSA describe flow of fast flowing ice streams and floating ice shelves well (Gagliardini et al., 2010; Cornford et al., 2020).

`fenics_ice` makes use of two sophisticated numerical libraries: `FEniCS` (Logg et al., 2012; Alnæs et al., 2015), an automated finite element method equation solver, and `tlm_adjoint` (Maddison et al., 2019), a library which implements automated differentiation of numerical partial differential equation solvers. `FEniCS` is a widely-used software library which abstracts the user away from low-level operations such as element-level operations. Rather, the weak form of the equation is written in Unified Form Language (UFL; Alnæs et al. (2014)), and `FEniCS` generates optimised low-level code which solves

the related finite-element problem with specified parameters (e.g. the order of the basis functions). `tlm_adjoint` is a library which implements high-level algorithmic differentiation of codes written with `FEniCS` or Firedrake.

`tlm_adjoint` is used for several of the operations detailed in Section 2. It facilitates the minimization of the model-data misfit cost J^c (Section 2.3) with respect to $\overline{x}$ (which is equivalent to finding the mode of the posterior density of $\overline{x}$). The higher-order derivative capabilities of `tlm_adjoint` furthermore enable efficient computation of the product of the Hessian of J^c with arbitrary vectors, enabling an iterative eigendecomposition of the prior-preconditioned Hessian as described in 2.4. Finally, `tlm_adjoint`'s time-dependent capabilities enable differentiation of the temporal trajectory of the Quantity of Interest Q_T, enabling projections of posterior uncertainty as described in 2.5. In our experiments in the present study, our cost function J^c is time-independent – but `tlm_adjoint` does allow for efficient calculation of Hessian-vector products for time-varying functionals (Maddison et al. (2019), their Section 4.2) – meaning time-varying data constraints can be considered with `fenics_ice`. Currently `fenics_ice` calls SLEPc for the solution of the generalised eigenvalue problem

$$\mathbf{H}_{mis}\mathbf{C} = \mathbf{\Gamma}_{prior}\mathbf{C}\mathbf{\Lambda} \tag{23}$$

ensuring real-valued eigenvalues – though in future versions of `fenics_ice` randomized algorithms of the type used by Villa et al. (2018) can be used without loss of generality.

`fenics_ice` solves the Shallow Shelf Approximation by implementing the corresponding variational principle (Schoof, 2006; Dukowicz et al., 2010; Shapero et al., 2021):

$$\begin{aligned}
&\int_\Omega 2H\nu\nabla\boldsymbol{\phi} : (\varepsilon_u + \mathrm{Tr}(\varepsilon_u)\mathbf{I})dA + \int_\Omega C^2\chi\boldsymbol{\phi}\cdot\boldsymbol{u}dA \\
&+ \int_\Omega \mathcal{W}\nabla R\cdot\boldsymbol{\phi} - \mathcal{F}\nabla\cdot\boldsymbol{\phi}dA \\
&+ \int_{\Gamma_c} (\boldsymbol{\phi}\cdot\boldsymbol{n})\cdot\left(\frac{1}{2}\rho g(H^2 - (\rho_w/\rho)|z_b^-|^2) - \mathcal{F}\right)dA = 0.
\end{aligned} \tag{24}$$

Here $\boldsymbol{\phi}$ is a vector-valued test function, and $\boldsymbol{u}$ is the depth-integrated horizontal velocity vector. ε_u is the horizontal strain-rate tensor $\frac{1}{2}(\nabla\boldsymbol{u} + \boldsymbol{u}^T)$, $\mathbf{I}$ is the 2×2 identity tensor, and ":" represents the Frobenius inner product. H is ice-sheet thickness (the elevation difference between the surface, z_s, and the base, z_b). ν is ice viscosity, which depends on the strain rate tensor:

$$\begin{aligned}
\nu &= \frac{1}{2}B\varepsilon_e^{\frac{1-n}{2n}}, \\
\varepsilon_e &= \varepsilon_{11}^2 + \varepsilon_{22}^2 + \varepsilon_{12}^2 + \varepsilon_{11}\varepsilon_{22}.
\end{aligned}$$

B is generally referred to as the "stiffness" of ice, and is thought to depend on ice temperature. In all experiments in this study, B is spatially constant and corresponds to a temperature of -12.5°C (Cuffey and Paterson, 2010). n is a measure of the degree of strain weakening of ice, with a widely accepted value of 3 (Glen, 1955). C is a real-valued, spatially varying sliding coefficient, and χ is a function that indicates where ice is grounded according to the hydrostatic condition:

$$H > (-\rho_w/\rho)R \equiv H_f \tag{25}$$

where ρ and ρ_w are ice and ocean densities, respectively, and R is bed elevation (note that $z_b = R$ when this condition is satisfied). In our code C can in general depend locally on velocity and thickness, though in this study we consider only a linear sliding law, i.e. one in which C varies only with location.

$\mathcal{F}$ is defined as

$$\mathcal{F} = \begin{cases} \frac{1}{2}\rho g H^2 & H > H_f, \\ \frac{1}{2}\rho g(\eta H^2 + (1-\eta)H_f^2 & \text{otherwise}, \end{cases} \tag{26}$$

where $\eta = (1 - \rho/\rho_w)$, and $\mathcal{W}$ as

$$\mathcal{W} = \begin{cases} \rho g H & H > H_f, \\ \rho g H_f & \text{otherwise}. \end{cases} \tag{27}$$

Γ_c is defined as the calving boundary, i.e. the boundary along which the ice sheet terminates in the ocean (or in a cliff on dry land), and $\boldsymbol{n}$ is the outward normal vector at this boundary. Finally, $|z_b^-|$ indicates the negative part of the ice base, i.e. it is zero when $z_b = R > 0$. The third integral of Eq. 24 is the weak form of the *driving stress* of the ice sheet, $\boldsymbol{\tau}_d = \rho g H \nabla z_s$. Although in our experiments in this study we consider only grounded ice, the full weak form is shown for completeness. The form of the driving stress term used here, $\nabla\mathcal{F} + \mathcal{W}\nabla R$, is not standard in glacial flow modelling, but it is equivalent to the more common form when thickness is represented by a continuous finite-element function.

In addition to the momentum balance, the continuity equation is solved:

$$H_t + \nabla \cdot (H\boldsymbol{u}) = b. \tag{28}$$

Here b represents localised changes in mass at the surface or the base of the ice sheet, i.e. accumulation due to snowfall or basal melting of the ice shelf by the ocean (though in the present study, surface b=0). The continuity equation is solved using a first-order upwind scheme which is implicit in H and explicit in $\boldsymbol{u}$. In this study, we do not consider initialisations based on time-varying data (i.e. the misfit cost function J^c_{mis} does not depend on time-varying fields), so the continuity function is only involved with finding a Quantity of Interest and propagation of initialisation uncertainty.

We discretize velocity ($\boldsymbol{u}$) using 1st-order continuous Lagrange elements on a triangular mesh. In the present study thickness (H) is discretized with 1st-order continuous Lagrange elements as well – although we point out that formulation (Eq. 26 and Eq. 27), together with an appropriate discretization for the continuity equation (28), will allow for discontinuous Galerkin elements (which have been found in more realistic experiments with `fenics_ice` to improve stability of time-dependent simulations). Eq. 24 is solved for $\boldsymbol{u}$ with a Newton iteration, with the Jacobian calculated at the level of the weak equation form using core `FEniCS` features. In the early iterations of the Newton solve, the dependence of ν on $\boldsymbol{u}$ is ignored in the Jacobian. This "linear" fixed-point iteration (often referred to in glacial modelling as Picard iteration, Hindmarsh and Payne (1996)) aids the Newton solver as it has a larger radius of convergence~~, once the Picard iteration is suitably converged~~. Once the nonlinear residual has decreased by a specified amount (a relative tolerance of 10^{-3} is used for this study), the full Newton iteration is applied.

To carry out an inversion, a cost function is minimized using the L-BFGS-B algorithm (Zhu et al., 1997; Morales and Nocedal, 2011) supplied with SciPy 1.5.2 (although note that no bounds on the controls are used). SLEPc (Hernandez et al., 2005) is used to implement the eigendecomposition described in Section 2.4, using a Krylov-Schur method. Rather than solve the eigenvalue problem (Eq. 14), we solve the Generalised Hermitian Eigenvalue Problem

$$\mathbf{H}_{mis}\mathbf{C} = \Gamma_{prior}^{-1}\mathbf{C}\mathbf{\Lambda}, \tag{29}$$

which guarantees real-valued eigenvalues. (Despite Λ being real-valued, the application of SLEPc to the non-hermitian eigenvalue problem Eq. 14 represents eigenvectors as imaginary, effectively doubling the memory requirements.)

4 Numerical Experiments

In this study, we aim to do the following:

1. Establish that control-method optimisations can be carried out with `fenics_ice`
2. Calculate eigendecompositions of the prior-preconditioned model-misfit Hessian as described in 2.4, examining the impacts of regularisation, resolution, and spatial density and autocorrelation of observations $\breve{y}_{obs}$ on the reduction of variance in the posterior.
3. Propagate the posterior uncertainty on to a Quantity of Interest Q_T as in 2.5
4. Establish, through simple Monte Carlo sampling, that the variance found through Eq. 22 is accurate.

Control method optimisations using ice-sheet models have been done extensively, with parameter sets of very high dimension (e.g., Cornford et al., 2015; Goldberg et al., 2015; Isaac et al., 2015), so our results regarding (1) above simply demonstrate the capabilities of `fenics_ice` but are not novel. Isaac et al. (2015) carries out eigendecompositions of the prior-preconditioned model-misfit Hessian and projects the associated uncertainty on to a Quantity of Interest – however, their QoI is time-independent. Importantly, Hessian-based Uncertainty Quantification has not been implemented for a model of ice dynamics using Algorithmic Differentiation before. Moreover, a time-dependent QoI has not been considered, nor has the impact of observational data density on the posterior uncertainty.

To investigate these and similar factors comprehensively, as well as validate the assumption of Gaussian statistics that leads to Eq 22, requires a model setup that is relatively inexpensive to run. We therefore choose one of the simplest frameworks possible for our numerical experiments, that of the Benchmark experiments for higher-order ice sheet models (ISMIP-HOM) intercomparison (Pattyn et al., 2008). We adopt the experiment ISMIP-C, a time-independent experiment in which an ice sheet slides across a doubly-periodic domain with constant thickness and a basal frictional factor that varies sinusoidally in both horizontal dimensions. The relation between velocity and basal shear stress is linear:

$$\boldsymbol{\tau}_b = -C^2(x,y)\boldsymbol{u} \tag{30}$$

where C is the factor from the second integral of Eq. 24, which has the form

$$C^2(x,y) = 1000 + 1000 \sin\frac{2\pi x}{L_x} \sin\frac{2\pi y}{L_y} \sqrt{1000 + 1000 \sin\left(\frac{2\pi x}{L_x}\right) \sin\left(\frac{2\pi y}{L_y}\right)} \tag{31}$$

with units of Pa (m/a)$^{-1}$, where L_x and L_y are experimental parameters. In this ISMIP-C specification, thickness is constant (H = 1000 m) and a shallow surface slope of 0.1° is imposed. In this ISMIP-HOM intercomparison, SSA models compared well with Stokes models for L_x, L_y over ∼40 km, so this is the value we use in our study. A regular triangular mesh is used to solve the model. Unless otherwise stated, cell diameter of the mesh is 1.33 km.

In our experiments, the momentum balance (Eq. 24) is solved on a highly refined grid and taken to be the "truth". To generate synthetic observations, values are interpolated to predefined locations. Observational error is then simulated by adding Gaussian random noise to these values. (These synthetic observations correspond to $\breve{u}_{obs}$ in Eq. 2.) In this study observational points occur at regular intervals, though our code allows for arbitrary distributions of observation points. Unless stated otherwise, in this study observational data points are spaced 2 km apart, with the velocity vector components coincident, and observational uncertainties are mutually independent with a standard deviation of 1 m/a. The regular spacing of observational points is not realistic and other studies use randomly scattered locations (e.g., Isaac et al., 2015); however, this choice is in line with the idealised nature of our study and furthermore allows comprehensive investigation of the effects of observational density (Section 5.4).

Our parameter-to-observable map $\breve{f}$ is really a composition of two functions: the first finds the solution to the momentum balance (Eq. 24) as a finite-element function, and the second interpolates the function to discrete locations. If the misfit cost were to be expressed as the weighted L_2 norm of the model-data misfit as in Eq. 8, then the interpolation function is replaced by the identity.

An inverse solution $\overline{c}$ is then found using a control method, where $\overline{c}$ is the vector of nodal coefficients of C. Below we refer to C as the *sliding parameter*. Note that minimisation with respect to C and not C^2. Our cost function J^c_{IS} is composed of a misfit term equal to the negative log density of observed velocities conditioned on $\overline{c}$ (*cf.* Eq. 2), and the regularisation operator is the discretised form of Eq. 9:

$$\begin{aligned} J^c_{IS} &= \frac{1}{2}\|\breve{\boldsymbol{u}}_{obs} - \breve{\boldsymbol{f}}(\overline{c})\|^2_{\boldsymbol{\Gamma}^{-1}_{u,obs}} + \frac{1}{2}\int_\Omega \|\mathcal{L}(c - cC - C_0)\|^2 dA \\ &= \frac{1}{2}\|\breve{\boldsymbol{u}}_{obs} - \breve{\boldsymbol{f}}(\overline{c})\|^2_{\boldsymbol{\Gamma}^{-1}_{u,obs}} + \frac{1}{2}\|c - c\overline{c} - \overline{c}_0\|^2_{\mathbf{L}\mathbf{M}^{-1}\mathbf{L}} \end{aligned} \tag{32}$$

where $\mathbf{L}$ is as described in Section 2.3. In many studies, the optimal value for γ, the regularisation parameter, is determined heuristically through an L-curve analysis (e.g., Gillet-Chaulet et al., 2012) – although there are alternative approaches (e.g., Waddington et al., 2007; Habermann et al., 2013). Here we examine, for different values of γ, the degree of uncertainty reduction associated with the cost-function optimisation. In other words, we seek the posterior density of $\overline{c}$, the coefficient vector of the finite-element function C. (We conduct an L-curve analysis, but only as a guideline for which values of γ to examine.) In our experiments C_0, the prior value of C, is uniformly zero – indicating we have no preconceieved notion of its mean value, only its spatial variability (implied by γ).

ISMIP-C does not prescribe a time-dependent component, but it is straightforward to evolve the thickness H (which is initially uniform) according to Eq. 28, where $m = 0$. We define a Quantity of Interest Q_T^{IS} as

$$Q_T^{IS} = \int_\Omega (H(T) - H_0)^4 dA. \tag{33}$$

Unlike Volume above Floatation, the example given in Section 2.5, Q_T^{IS} has no strong physical or societal significance. However, it is convenient to calculate and sufficiently nontrivial and nonlinear that the effects of uncertainty in C, as well as the strength of the prior covariance, can be seen. Moreover, Volume above Floatation is insensitive to small scale variability in thickness – but there may be scientific motivation to study Quantities of Interest which do take such variability into account (see Section 7). Thus we choose Eq. 33 as a QoI which is straightforward but also an indicator of thickness variability.

In our error propagation we evolve the ISMIP-C thickness for 30 years and use the time-dependent adjoint capabilities of `fenics_ice` to find $\partial Q_T^{IS}/\partial \overline{c}$ for discrete values of T over this period, and uncertainty at these times is found using Eq. 22; an uncertainty "trajectory" is then found for Q^{IS} via interpolation. Our results regarding the uncertainty of Q^{IS}, and the quadratic approximation inherent in Eq. 22, are then tested via sampling from the posterior as described in Section 5.3.1.

5 Results

5.1 Parameter uncertainties

5.1.1 Effect of regularisation

An L-curve for our inversion results (Fig. 1) shows the behaviour of regularisation cost and model-data misfit as γ is varied over 3 orders of magnitude. In all inversions, $\overline{c}$ is initialised assuming a point-wise balance between driving stress and basal drag arising from interpolated velocity observations, and J^c is lowered from the initial value by a factor of $\sim 10^3$ (meaning the probability density associated with C, proportional to e^{-J^c}, is increased by a factor of approximately 10^{400}).

While misfit does not vary greatly in a proportional sense, it suggests $\gamma = 10$ as a reasonable tradeoff between misfit and regularisation. Fig. 2 displays results of an inversion with a "strong" level of regularisation ($\gamma = 50$; referred to below as the γ_{50} experiment). The resulting C is relatively smooth (Fig. 2(a)), and the misfit is generally small though with some outliers (Fig. 2(d)). (Misfit is displayed as a histogram of errors – obtained by interpolating the finite-element solution to the sampled velocity locations – rather than as a spatially continuous function, to emphasize the discrete nature of the model-data misfit.) Fig. 3 gives equivalent results for a "weak" regularisation inversion ($\gamma = 1$; referred to below as the γ_1 experiment). The misfit distribution is similar but the inverted sliding parameter is significantly noisier, as a result of weaker constraints on these "noisy" modes by the prior.

The effect of regularisation on reduction of uncertainty can be seen from examining the eigenvalues defined by Eq. 14. More precisely, the ratio $1/(1 + \lambda_i)$, where λ_i is the ith leading eigenvalue, is examined. As shown in Section 2.4, this ratio gives the reduction in variance of the associated eigenvector in the posterior PDF relative to the prior distribution. In Fig. 4(a) this quantity is shown for the eigenvalue spectra corresponding to γ = 1, 10, and 50. For all inversions, uncertainty reduction is

several orders of magnitude for the leading eigenvalues, but the tails of the spectra are quite different. In the case of strong regularisation, there is little reduction in variance beyond $i \sim 100$, while in the weakly regularised case there is considerable reduction across the entire spectrum. This discrepancy can be interpreted as the prior providing so little information in the low-regularisation case that the information provided by the inversion reduces uncertainty across all modes. The comparison of eigenvalue spectra across experiments is only meaningful to the extent that the corresponding eigenvectors are equivalent. A comparison between the four leading eigenvectors in the high- and low-regularisation experiments (Fig. 5) shows they are not equivalent but have the same overall structure. (Differences arise due to $\bar{c}_{MAP}$ but also due to differences in $\mathbf{\Gamma}_{prior}$.)

Approximating the posterior covariance of $\bar{c}$, Γ_{post}, also allows an estimation of Σ_C, the *pointwise* variance of C. This is done via calculation of the square root of the diagonals entries of Γ_{post}, i.e. the standard deviation of the marginal distributions of the coefficients of $\bar{c}$. Σ_C is shown for the inversions discussed above in Figs. 2(b) and 3(b). Pointwise uncertainties in γ_1 are 5-10 times larger than in γ_{50}. For γ_{50} there is a clear pattern of higher uncertainty where the bed is weaker (i.e. C is smaller), though for γ_1 it is difficult to discern any pattern.

5.1.2 Effect of resolution

The impacts of grid resolution on eigenvalue spectra are investigated (Fig. 4(b)). In Isaac et al. (2015), it was shown that the leading eigenvalues were independent of the numerical mesh, implying that the leading eigenvectors – the patterns for which uncertainties are quantified – are not dependent on the dimension of the parameter space (which would be an undesirable property). Our spectra suggest that at 2 km resolution, there is mesh dependence; but the spectra for 1.33 and 1 km resolution are in close agreement, suggesting mesh independence (Isaac et al., 2015). Consistent values of γ and δ are used for these experiments, meaning the results of the L-curve in Fig. 1 are not dependent on model resolution.

5.2 Propagation of uncertainties

5.3 Linear propagation of uncertainties

The low-rank approximation of the posterior covariance of $\bar{c}$ found with Eq. 20 can be used to estimate the uncertainty of Q_T^{IS} using the Eq. 22. To do so, $\frac{\partial Q_T^{IS}}{\partial \bar{c}}$ must be found, which is done using Algorithmic Differentiation of the time-dependent model as described in Section 4. Figs. 2(e) and 3(e) show $\frac{\partial Q_{30}^{IS}}{\partial \bar{c}}$ arising from their respective inversions. There is small-scale noise in the low-regularisation experiment (γ_1), but the general pattern and magnitude between the two gradients is similar, with strengthening of weak-bedded areas and weakening of strong-bedded areas both leading to an increase in the fourth-order moment of thickness. The gradient of Q_T^{IS} with respect to $\bar{c}$ is found for intermediate values of T over the 30-year interval, with $\sigma(Q_T^{IS})$ calculated at these times – which can then be linearly interpolated to find a trajectory of uncertainty. In our experiments we find the gradient of Q_T^{IS} every 6 years~~, but . .. say something about how it is not time-limited to find gradients on shorter intervals~~. In Fig. 6(a) these trajectories are shown for the γ_{50} and γ_1 experiments, plotted as a 1-σ error interval around the calculated trajectory of Q_T^{IS}.

The trajectory of uncertainties for γ_{50} and γ_1 can also be seen in Fig 6(b), compared against the trajectories of

$$\left(\left(\frac{\partial Q_T^{IS}}{\partial \overline{c}}\right)^T \mathbf{\Gamma}_{prior}\left(\frac{\partial Q_T^{IS}}{\partial \overline{c}}\right)\right)^{1/2} \tag{34}$$

i.e. the prior uncertainty linearly projected along the trajectory of Q_T^{IS}. This uncertainty measure is not physically meaningful as it depends on the calculated $\partial Q_T^{IS}/\partial \overline{c}$, which in turn depends on the inversion for $\overline{c}$ and the related trajectory of Q_T^{IS} – and a random sample from the prior distribution of $\overline{c}$ is unlikely to yield such a trajectory. Still, it serves as a measure of decrease in uncertainty arising from the information encapsulated in the observations and model physics.

Q_T^{IS} is greater in magnitude in the γ = 1 experiment than in the γ = 50 experiment at all times – and it can be seen from the uncertainty of the γ_1 trajectory that this difference is statistically significant. The two experiments have differing (inverse) solutions, with the γ_1 inversion favoring a closer fit to noisy observations at the cost of small-scale variability in the inverse solution. Our quantity of interest (the fourth-order moment of thickness) is sensitive to this small-scale variability, so uniformity of trajectories of Q_T^{IS} would not be expected. At the same time, the level of QoI uncertainty in the γ_1 trajectory relative to that of the γ_{50} QoI uncertainty is much smaller than the relative magnitudes of the inversion uncertainties (*cf.* Figs 2(b), 3(b)) would suggest. This can be rationalised by considering Eq. 22: uncertainty in the QoI will depend on the extent to which uncertain parameter modes project on to the gradient of the QoI with respect to the parameters. While the γ_1 inversion results are overall more uncertain, the leading order modes are still constrained quite strongly. Thus, while Q_T^{IS} is to a degree sensitive to small-scale variability it may still filter the most uncertain modes of the γ_1 inversion, resulting in a smaller QoI uncertainty than expected. In fact, it can be seen from Fig. 6(b) that despite the large differences in prior distributions between γ_{50} and γ_1, the projections of the respective prior covariances along the trajectory of Q_T^{IS} are very similar, suggesting that the gradient of Q_T^{IS} does not project strongly on the modes which are poorly constrained in the γ_1 experiment.

~~The sensitivity of QoIs to small-scale variability is significant because not all glaciologically motivated QoIs are expected to have such sensitivities. For instance, the QoI considered by Isaac et al. (2015) was a contour integral of volume flux over the boundary of the domain, equivalent to a rate of change of ice volume – and such a quantity might be less sensitive to velocity gradients and small-scale thickness change in the domain interior. On the other hand, a forecast focused on the impact of evolving surface elevation on proliferation of surface lakes, or on surface fractures, might be very sensitive to such variability. Therefore, when considering parametric uncertainty, it should also be considered whether the nature of this uncertainty impacts the uncertainty of the intended Quantity of Interest. *maybe move to discussion*~~

5.3.1 Direct sampling of QoI uncertainties

Ideally, the assumptions implicit in the calculation of QoI uncertainties shown in Fig. 6(a) would be tested through unbiased sampling from the prior distributions of $\overline{c}$; followed by using the sampled parameters to initialise the time-dependent model and generating a sample of trajectories of Q_T^{IS}; and finally scaling the probability of each member of the ensemble based on the observational likelihood function $p(\breve{\boldsymbol{u}}_{obs}|\overline{c})$. However, given the dimension of the space containing $\overline{c}$ (equal to 900 in our idealised experiment; but on the order of 10^4-10^5 in more realistic experiments), the number of samples required to

ensure nonnegligible likelihoods would not be tractable without a sophisticated sampling strategy such as Markov Chain Monte Carlo (MCMC) methods (Tierney, 1994) (and even then may require approximations similar to those described above ~~(Martin et al., 2012)~~(Martin et al., 2012; Petra et al., 2014)). However, such approaches are beyond the scope of this study.

The assumptions in our propagation of observational and prior uncertainty to Quantity of Interest uncertainty are (i) that of Gaussianity of the distribution of $\overline{c}$ and (ii) that of linearity of the map from $\overline{c}$ to QoI. While (i) cannot be tested for the reasons stated above, (ii) can be tested by sampling from the calculated posterior distribution of $\overline{c}$, initialising the time-dependent model, and finding the ensemble variance and standard deviation of Q_T^{IS}. Our strategy for sampling from the posterior is described below, and is based on the derivation in Bui-Thanh et al. (2013).

A randomly sampled vector $\overline{x}$ will have covariance $\mathbf{\Gamma}_{post}$ and mean $\overline{c}_{MAP}$ if it is generated via

$$\overline{x} = \overline{c}_{MAP} + \mathbf{K}\overline{N} \tag{35}$$

where $\overline{N}$ is a sample from a multivariate normal distribution $\mathbb{N} \sim \mathcal{N}(0, \mathbf{I})$ of the same dimension as $\overline{c}$, and $\mathbf{K}$ is such that $\mathbf{K}\mathbf{K}^T = \mathbf{\Gamma}_{post}$. Hence it is required to find a suitable $\mathbf{K}$. We restate the generalised eigenvalue problem $\mathbf{H}_{mis}\mathbf{C} = \Gamma_{prior}^{-1}\mathbf{C}\mathbf{\Lambda}$. Since $\mathbf{C}$ is orthogonal in the inverse prior covariance (*cf.* Eq. 16), the identity matrix $\mathbf{I}$ can be spectrally decomposed in $\overline{c}_i$ (the columns of $\mathbf{C}$):

$$\left(\sum \overline{c}_i \overline{c}_i^T \mathbf{\Gamma}_{prior}^{-1}\right) = \mathbf{I}. \tag{36}$$

Rearranging gives $\sum \overline{c}_i \overline{c}_i^T = \mathbf{\Gamma}_{prior}$, and so (*cf.* Eq. 20)

$$\begin{aligned}\mathbf{\Gamma}_{post} &\sim \mathbf{\Gamma}_{prior} - \mathbf{C}_r \mathbf{D}_r \mathbf{C}_r^T \\ &= \sum_{i=1}^{n} \overline{c}_i \overline{c}_i^T - \sum_{i=1}^{r} \overline{c}_i \overline{c}_i^T \left(\frac{\lambda_i}{1+\lambda_i}\right) \\ &= \sum_{i=r+1}^{n} \overline{c}_i \overline{c}_i^T + \sum_{i=1}^{r} \overline{c}_i \overline{c}_i^T \left(\frac{1}{1+\lambda_i}\right).\end{aligned} \tag{37}$$

We define the matrix $\mathbf{B}$:

$$\begin{aligned}\mathbf{B} &= \mathbf{\Gamma}_{prior} + \sum_{i=1}^{r} \overline{c}_i \overline{c}_i^T \left(\frac{\lambda_i}{\sqrt{1+\lambda_i}} - 1\right) \\ &= \sum_{i=r+1}^{n} \overline{c}_i \overline{c}_i^T + \sum_{i=1}^{r} \overline{c}_i \overline{c}_i^T \left(\frac{1}{\sqrt{1+\lambda_i}}\right).\end{aligned} \tag{38}$$

And due to the $\mathbf{\Gamma}_{prior}^{-1}$-orthogonality of $\mathbf{C}$,

$$\begin{aligned}\mathbf{B}\mathbf{\Gamma}_{prior}^{-1}\mathbf{B}^T &= \sum_{i=r+1}^{n} \overline{c}_i \overline{c}_i^T + \sum_{i=1}^{r} \overline{c}_i \overline{c}_i^T \left(\frac{1}{1+\lambda_i}\right) \\ &= \mathbf{\Gamma}_{post}.\end{aligned} \tag{39}$$

Therefore a suitable $\mathbf{K}$ is given by (*cf.* Eq. 11)

$$\mathbf{B}\boldsymbol{\Gamma}_{prior}^{-1/2} = \mathbf{BLM}^{-1/2}. \tag{40}$$

The action of the square root of the mass matrix $\mathbf{M}$ is found by a Taylor series approach (Higham (2008), eq. 6.38). Fig. 7 shows a result of sampling from the posterior in the γ_{50} experiment. To the left (panel (a)), a realisation of the prior distribution, with mean zero and covariance $\boldsymbol{\Gamma}_{prior}$ is displayed. (This realisation is found similarly to that of the posterior, with the formula $\overline{N}\boldsymbol{\Gamma}_{prior}^{1/2}\cdot$.) To the right (panel (b)), a realisation of the posterior distribution is shown with the mean $\overline{c}_{MAP}$ removed. (Note that both samples are derived from the same realisation of $\mathbb{N}$.) From comparing the images it can be seen that variance is greatly reduced, particularly at medium-to-large scales. By contrast, when the posterior distribution of γ_1 is sampled, the result is very similar to the prior. Very little reduction of variance is visually apparent, especially at small scales.

Using this method of sampling the posterior, an ensemble of 1,000 30-year runs is carried out for both low and high regularisation experiments (γ_1 and γ_{50}, respectively), and standard deviations of Q_T^{IS} are calculated at discrete times. Values are plotted in Fig. 6(b). (For each such calculation, the variance quickly converged to the value shown, so it is unlikely that the Quantity of Interest is under-sampled.) In the γ_{50} experiment there is strong agreement between the sampled uncertainties and those found via projecting C uncertainty along the linearised QoI trajectory, suggesting the linear approximation inherent in Eq. 22 is appropriate. In contrast, there are large discrepancies in the γ_1 case. It is likely that the small-scale noise inherent in the low-regularisation samples (*cf* Fig. 8) impacts the Quantity of Interest strongly enough that the linear approximation in Eq. 22 breaks down – despite that this noise does not strongly affect the cost function J^c. As mentioned in Sec. 5.2, this may be due to the nature of the QoI. In the γ_{10} experiment (not shown), the disagreement in uncertainties is on the order of 30 – greater than for the γ_{50} experiment but far less than for γ_1.

5.4 Observational density and uncertainty

In all results presented to this point, the imposed locations observational data $\breve{u}_{obs}$, $\breve{v}_{obs}$ lie on a regular grid with a spacing of 2 km. Here we consider the effects of the observational spatial density on the reduction of uncertainty in $\overline{c}$.

5.4.1 Effect of observation spacing

Eigendecompositions of the prior-preconditioned misfit Hessian ($\hat{\mathbf{H}}_{mis}$) are carried out for observational spacings of 500 m, 1 km, 2 km, 4 km, and 8 km. (The 2 km case corresponds to the γ_{10} experiment in Fig. 4(a).) In all other respects the experiments are identical. Results are shown for comparison in Fig. 4(c). Increasing spatial density appears to reduce uncertainty: in the 500 m case, there is considerable uncertainty reduction even in cases where there is almost no reduction in coarser-observation cases. The result is intuitive: each increase in observational density quadruples the number of independent constraints, effectively adding more information (though a more sophisticated framework is required to quantify the information increase from a given observation, *e.g.*, Alexanderian et al. (2014)).

Comparison of eigenspectra relies on the corresponding eigenvectors being the same, or similar, between the experiments. As in the regularisation and resolution experiments, the eigenvectors depend on the exact form of $\hat{\mathbf{H}}_{mis}$ which in turn depends

on $\bar{c}_{MAP}$, which may differ between the experiments due to the differing number of points. However they are likely to be of similar patterns (on the basis of the results of Section 5.1.1).

5.4.2 Effect of observational covariance

The results described above imply that posterior uncertainty could be made arbitrarily small by increasing the spatial density of observations (although we do not examine observations more dense than 500 m). However, the decreasing uncertainty relies on the observations being statistically *independent*, which is unlikely to be the case as observations become more and more dense. We consider here the implications of a nonzero spatial covariance. Rather than imposing a realistic observational covariance matrix, we consider a simplified covariance structure in which correlations decay ~~isotroptically~~isotropically. That is, our observational covariance matrix $\mathbf{\Gamma}_{u,obs}$ is given by

$$\mathbf{\Gamma}_{u,obs}(i,j) = \begin{cases} \sigma^2_{u,obs} & i = j, \\ \sigma^2_{u,obs} e^{-\frac{|\boldsymbol{x}_i - \boldsymbol{x}_j|^2}{d^2_{auto}}} & o.w. \end{cases} \tag{41}$$

Here $\sigma_{u,obs}$ is the observational uncertainty and $\boldsymbol{x}_i$ is the position of observation i. (By contrast, $\mathbf{\Gamma}_{u,obs}$ in all experiments described above is a diagonal matrix with entries $\sigma^2_{u,obs}$.) A value of 1 m/a is used for $\sigma_{u,obs}$, as in all previous experiments; and d_{auto} is set to 750 m. We assert that the observations of orthogonal velocity components are independent, i.e. $\mathbf{\Gamma}_{u,obs}$ is block-diagonal with each block corresponding to a velocity component. While velocity component uncertainties are likely to correlate, introducing spatial correlation among the individual components already greatly changes the effect of observation spacing on uncertainty reduction, as seen in Fig. 4(d). When observational spacing is large compared to d_{auto}, an increase in density has a similar effect to that seen in the zero-spatial correlation case (Fig. 4(c)). But for observational spacing on the order of d_{auto}, additional observations have minimal effect.

6 Gauss-Newton approximation to the Hessian

Section 2.2 introduces the Hessian of the cost function, and gives an expression for the posterior covariance when the parameter-to-observable map $\breve{f}$ is linear (Eq. 5). The first term in brackets on the right hand side of Eq. 5, which we write here:

$$\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right)^T \mathbf{\Gamma}^{-1}_{obs} \left(\frac{\partial \breve{f}}{\partial \bar{c}}\right). \tag{42}$$

is the Hessian of the misfit cost function J^c_{inis} under the condition that $\breve{f}$ is linear. It is quite often used as an approximation to the Hessian for the purpose of covariance estimates even when $\breve{f}$ is nonlinear (Kaminski et al., 2015; Loose et al., 2020), and is referred to as the *Gauss-Newton* approximation to the Hessian (GNaH). The GNaH has the advantage of avoiding the complexity of finding second-order derivatives. It also has the property that the misfit Hessian (or rather, its approximation) is positive semidefinite – this is not necessarily true of the "full" Hessian even when the cost function J^c is minimised, meaning the eigendecomposition described in Section 2.4 can have negative eigenvalues.

In the context of our idealised experiments we calculate the GNaH (or rather, its action on a vector) to compare against the "full" Hessian. `tlm_adjoint` library's functionality is employed as follows. For a given finite-element coefficient vector $\bar{\xi}$, the GNaH action can be written

$$\left(\frac{\partial \breve{f}}{\partial \bar{c}}\right)^T \mathbf{\Gamma}_{obs}^{-1} \left(\frac{\partial \breve{f}}{\partial \bar{c}}\right) \bar{\xi} = \left(\frac{\partial \breve{f}}{\partial \bar{c}}\right)^T \breve{\eta} \tag{43}$$

where $\breve{\eta}$ is obtained through the action of the tangent linear model $\left(\partial \breve{f} / \partial \bar{c}\right)$ on $\bar{\xi}$ and the action of the inverse observational covariance on the result. The GNaH action is then obtained through the action of the adjoint of the Jacobian of the parameter-to-observable map, $\left(\partial \breve{f} / \partial \bar{c}\right)^T$.

In Fig. 9 we examine the effects of using the GNaH rather than the Hessian in our Hessian-based UQ framework. This is done for just a single experiment, with 1.33 km elements, 2 km observational spacing, and regularisation with $\gamma = 10$. Examining the uncertainty reduction (Fig. 9(a)), the first $\sim$80 eigenvalues are near-identical, but after this the uncertainty reduction approaches 1 much faster (or, equivalently, the eigenvalues decay much faster) with the GNaH. In terms posterior QoI uncertainty (Fig. 9(b)), $\sigma(Q_T^{IS})$ is slightly smaller with the GNaH, but the difference is very small and only visible at later times.

7 Discussion and Conclusions

The inversion of surface velocities for basal conditions is ubiquitous in ice-sheet modelling – but in most studies in which this is done, the uncertainty of the resulting parameter fields is not considered, and the implications of this parametric uncertainty on projection uncertainty is not quantified. We introduce `fenics_ice`, a numerical Python code which solves the Shallow Shelf Approximation (SSA) for ice-sheet dynamics. The code uses the `FEniCS` library to facilitate finite-element solution of partial differential equations. Algorithmic differentiation is implemented with the `tlm_adjoint` library, allowing for adjoint generation of the time-dependent and time-independent versions of the SSA. This feature is used to aid in inversions of surface velocity for parameter fields such as the basal sliding parameter. In addition, the `tlm_adjoint` library allows efficient second-order differentiation of the inversion cost function, allowing a low-rank approximation to the cost function Hessian. We utilise this ability to exploit the connection between the control-method inversions typically carried out with ice-sheet models, and a Bayesian characterisation of the uncertainty of the inverted parameter field. This interpretation allows us to form a local approximation to the posterior probability density at the maximum a posteriori (MAP) point. With a time-dependent Quantity of Interest (QoI) which depends on the outcome of the inversion, the adjoint features of `fenics_ice` allow linear propagation of parametric uncertainty to QoI uncertainty.

We apply our framework to a simple idealised test case, Experiment C of the ISMIP-HOM intercomparison protocol, involving an ice stream sliding across a doubly periodic domain with a varying basal friction parameter. An idealised time-varying QoI is defined, equivalent to the fourth moment of thickness in the domain, as thickness evolves due to mass continuity. The posterior probability density is examined, suggesting mesh independence (provided resolution is high enough). It is shown that the level of uncertainty reduction relative to the prior distribution depends on the amount of information in the prior (or,

equivalently, the degree of regularisation). Uncertainty of the QoI is found along its trajectory, and is found to increase with time and also found to be larger with less-constrained priors. However, the difference in the uncertainty of the QoI is far less than that of the parametric uncertainty, due to ~~filtering of~~ insensitivity of the QoI to high-frequency modes. ~~Testing the~~

Sampling from our posterior allows us to test the linearity of the parameter-to-QoI mapping, and this approximation is seen to be accurate with a moderately strong prior. However, even with the relatively modest problem sizes considered, testing the validity of our local Gaussian approximation of the posterior probability density ~~is~~ would require sophisticated sampling methods which are beyond the scope of our study. ~~However, sampling from our posterior allows us to test the linearity of the parameter-to-QoI mapping, and this approximation is seen to be accurate with a moderately strong prior.~~ It is worth noting though that one such method, Stochastic Newton MCMC Martin et al. (2012); Petra et al. (2014), relies on framework developed in this study (i.e. characterising the local behaviour of the posterior density through a Hessian-based approximation). Therefore it may be a viable approach for non-Gaussian uncertainty quantification in future iterations.

The sensitivity of QoIs to small-scale variability is significant because not all glaciologically motivated QoIs are expected to have such sensitivities. For instance, the QoI considered by Isaac et al. (2015) was a contour integral of volume flux over the boundary of the domain, equivalent to a rate of change of ice volume – and such a quantity might be less sensitive to velocity gradients and small-scale thickness change in the domain interior. On the other hand, a forecast focused on the impact of evolving surface elevation on proliferation of surface lakes, or on surface fractures, might be very sensitive to such variability. Therefore, when considering parametric uncertainty, it should also be considered whether the nature of this uncertainty impacts the uncertainty of the intended Quantity of Interest.

A key difference between our approach and the control-method inversions typically undertaken is the Euclidean inner product that appears in the misfit component of the cost function, as opposed to an area integral of velocity misfit. As discussed in Section 2.3, the latter formulation leads to difficulties with a Bayesian interpretation by conflating the observational error covariance with mesh-dependent factors. In our study observation locations are imposed on a regular grid. It is shown that, with statistically independent observations, posterior uncertainty is continually reduced as the observational grid becomes more dense. When a spatial correlation of observations is considered, however, there is little reduction of uncertainty when adding observations beyond a certain spatial density. This result is of significance to ice-sheet modelling: most ice-sheet model studies which calibrate parameters to velocity observations (including those mentioned in the Introduction) do not consider the spatial correlation of observations. As discussed in Section 2.3, these studies express the model-data misfit as an area integral – meaning that, effectively, observations in adjacent model grid cells are considered independent. If grid cells are sufficiently large, this is likely a suitable approximation – though with higher and higher resolutions being used in ice-sheet modelling studies (Cornford et al., 2013), it should be considered whether the spatial covariance of observations is such that it might affect results. Assessing such effects poses an additional challenge, however, as ice-sheet velocity products are not generally released with spatial error covariance information (Rignot et al., 2017).

Our study does not consider "joint" inversions, i.e. inversions with two or more parameter fields. With such inversions, complications can arise when both parameters affect the same observable, potentially leading to equifinality/ill-posedness. An example of such a pair is C, the sliding coefficient, and B, the ice stiffness in the nonlinear Glen's rheology (*cf.* Eq. 24), which

can both strongly affect ice speeds in a range of settings. The version of `fenics_ice` presented in this study is not capable of joint inversions or of Hessian-vector products with multiple parameter fields. ~~However,~~ , however the technical hurdles to implementation are minor. More importantly, though, Hessian-based Bayesian uncertainty quantification with multiple parameter fields has not, to our knowledge, been carried out in an ice-sheet modelling context, and may present difficulties due to a larger problem space or the equifinality issues mentioned above. (Instead of performing a joint inversion, Babaniyi et al. (2020) use a Bayesian Approximation Error framework, treating the stiffness parameter as a random variable.) Nonetheless, the investigation of joint inversions and uncertainty quantification is a future research aim for `fenics_ice`.

Model uncertainty is not accounted for in our characterisation of parametric uncertainty. In the expression for the posterior probability density (Eq. 4), the model-misfit term is expressed as the difference between observed and modelled velocity, and the uncertainty is assumed to arise from the observation platform. In fact, the discrepancy between modelled and observed velocity is the sum of observation error, ϵ_{obs} and model error, ϵ_{model}. This second error source can be considered a random variable, as it arises from incomplete knowledge about the ice-sheet basal environment and material properties of the ice, as well as the approximations inherent in the Shallow Shelf Approximation. Characterising this uncertainty is challenging as it requires both perfect knowledge of the basal sliding parameter and observations with negligible error, and is beyond the scope of our study. Future research, however, could involve using a model which implements the full Stokes solution (e.g., Gagliardini et al., 2010) to partially characterise this uncertainty, or could make use of a multi-fidelity approach (Khodabakhshi et al., 2021).

A number of Hessian-based uncertainty quantification studies use the Gauss-Newton approximation to the Hessian (*cf.* Section 6), avoiding computation of higher-order derivatives of the model, but few studies have investigated the impact of neglecting these higher-order terms. For our idealised experiment we have compared uncertainty reduction and QoI uncertainty with both Gauss-Newton and "full" Hessian computation. There is negligible difference in terms of QoI uncertainty, but it remains to be seen if this is the case for more realistic experiments.

Our study does not consider time-dependent inversions, i.e. control methods where the cost function is time-dependent. While the majority of cost-function inversions are time-independent, there are a growing number of studies carried out with time-dependent inversions (Larour et al., 2014; Goldberg et al., 2015) and it is possible that such methods may provide lower uncertainty in calibration of hidden parameters (simply by providing additional constraints) and hence in ice-sheet projections. `fenics_ice` (or rather `tlm_adjoint`) is capable of eigendecomposition of Hessian matrices of time-dependent cost functions (Maddison et al., 2019), but time-dependent Hessian-vector products are computationally expensive, requiring checkpointing and recomputation of both forward and reverse mode model information, and it is unlikely that full eigenvalue spectra can be found for even modestly sized problems. It is hopeful that for realistic problems of interest only a small fraction of eigenvalues will need to be found to accurately approximate the posterior covariance, ~~but~~ . Alternatively, the Gauss-Newton approximate Hessian might diminish some of the cost. Certainly more work is required in this area.

Code availability. The `fenics_ice` code can be obtained from https://doi.org/10.5281/zenodo.5153231 and is freely available under the LGPL-3.0 License. The branch containing the version of the code used for this manuscript is *GMD_branch*. Python scripts for running

all experiments and creating all figures in this manuscript can be found in the *example_cases* directory, and installation instructions for `fenics_ice` and dependencies can be found in the *user_guide* folder. The commit tag of `tlm_adjoint` used for the experiments in this manuscript is 79c54c00a3b4b69e19db633896f2b873dd82de4b.

Author contributions. CK and JT developed the `fenics_ice` software with considerable support from JM. DG and JM developed the mathematical framework implemented in `fenics_ice`. DG wrote the manuscript and ran and analysed the experiments.

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

Acknowledgements. The authors acknowledge NERC Standard Grants NE/M003590/1 and NE/T001607/1 (QUoRUM).

References

Alexanderian, A., Petra, N., Stadler, G., and Ghattas, O.: A-Optimal Design of Experiments for Infinite-Dimensional Bayesian Linear Inverse Problems with Regularized ℓ_0-Sparsification, SIAM Journal on Scientific Computing, 36, A2122–A2148, https://doi.org/10.1137/130933381, 2014.

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.

Alnæs, M. S., Logg, A., Ølgaard, K. B., Rognes, M. E., and Wells, G. N.: Unified Form Language: A Domain-Specific Language for Weak Formulations of Partial Differential Equations, ACM Trans. Math. Softw., 40, https://doi.org/10.1145/2566630, 2014.

Arthern, R. J., Hindmarsh, R. C. A., and Williams, C. R.: Flow speed within the Antarctic ice sheet and its controls inferred from satellite observations, Journal of Geophysical Research: Earth Surface, 120, 1171–1188, http://dx.doi.org/10.1002/2014JF003239, 2015.

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 Discussions, 2020, 1–33, https://doi.org/10.5194/tc-2020-229, https://tc.copernicus.org/preprints/tc-2020-229/, 2020.

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.

Chen, P.: Hessian Matrix vs. Gauss-Newton Hessian Matrix, SIAM J. Numer. Anal., 49, 1417–1435, https://doi.org/10.1137/100799988, 2011.

Cornford, S. L., Martin, D. F., Graves, D. T., Ranken, D. F., Le Brocq, A. M., Gladstone, R. M., Payne, A. J., Ng, E. G., and Lipscomb, W. H.: Adaptive Mesh, Finite Volume Modeling of Marine Ice Sheets, J. Comput. Phys., 232, 529–549, http://dx.doi.org/10.1016/j.jcp.2012.08.037, 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, https://www.the-cryosphere.net/9/1579/2015/, 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, https://tc.copernicus.org/articles/14/2283/2020/, 2020.

Cuffey, K. and Paterson, W. S. B.: The Physics of Glaciers, Butterworth Heinemann, Oxford, 4th edn., 2010.

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.

Dukowicz, J. K., Price, S. F., and Lipscomp, W. H.: Consistent approximations and boundary conditions for ice-sheet dynamics from a principle of least action, Journal of Glaciology, 56, 480–496, 2010.

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-9-1427-2015, http://www.the-cryosphere.net/9/1427/2015/, 2015.

Gagliardini, O., Durand, G., Zwinger, T., Hindmarsh, R. C. A., and Meur, E. L.: Coupling of ice shelf melting and buttressing is a key process in ice sheet dynamics, Geophys. Res. Lett., https://doi.org/10.1029/2010GL043334, 2010.

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.

Gladstone, R. M., Lee, V., Rougier, J., Payne, A. J., Hellmer, H., Le Brocq, A., Shepherd, A., Edwards, T. L., Gregory, J., and Cornford, S. L.: Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model, Earth and Planetary Science Letters, 333, 191–199, https://doi.org/10.1016/j.epsl.2012.04.022, 2012.

Glen, J. W.: The creep of polycrystalline ice, Proc. Royal Soc. of London, Series A, 228, 519–538, 1955.

Goldberg, D. N., Heimbach, P., Joughin, I., and Smith, B.: Committed retreat of Smith, Pope, and Kohler Glaciers over the next 30 years inferred by transient model calibration, The Cryosphere, 9, 2429–2446, https://doi.org/10.5194/tc-9-2429-2015, https://www.the-cryosphere.net/9/2429/2015/, 2015.

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

Hernandez, V., Roman, J. E., and Vidal, V.: SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems, ACM Trans. Math. Software, 31, 351–362, 2005.

Higham, N. J.: Functions of Matrices: Theory and Computation, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

Hindmarsh, R. C. A. and Payne, A. J.: Time-step limits for stable solutions of the ice-sheet equation, Annals of Glaciology, 23, 74–85, https://doi.org/10.1017/S0260305500013288, 1996.

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/http://dx.doi.org/10.1016/j.jcp.2015.04.047, http://www.sciencedirect.com/science/article/pii/S0021999115003046, 2015.

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.

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.

Kaminski, T., Kauker, F., Eicken, H., and Karcher, M.: Exploring the utility of quantitative network design in evaluating Arctic sea ice thickness sampling strategies, The Cryosphere, 9, 1721–1733, https://doi.org/10.5194/tc-9-1721-2015, 2015.

Keuthen, M. and Ulbrich, M.: Moreau–Yosida regularization in shape optimization with geometric constraints, Computational Optimization and Applications, 62, 181–216, 2015.

Khodabakhshi, P., Willcox, K. E., and Gunzburger, M.: A multifidelity method for a nonlocal diffusion model, Applied Mathematics Letters, 121, 107 361, 2021.

Larour, E., Rignot, E., Joughin, I., and Aubry, D.: Rheology of the Ronne Ice Shelf, Antarctica, inferred from satellite radar interferometry data using an inverse control method, Geophys. Res. Lett., 32, L05 503, https://doi.org/10.1029/2004GL021693, 2005.

Larour, E., Utke, J., Csatho, B., Schenk, A., Seroussi, H., Morlighem, M., Rignot, E., Schlegel, N., and Khazendar, A.: Inferred basal friction and surface mass balance of the Northeast Greenland Ice Stream using data assimilation of ICESat (Ice Cloud and land Elevation

Satellite) surface altimetry and ISSM (Ice Sheet System Model), The Cryosphere, 8, 2335–2351, https://doi.org/10.5194/tc-8-2335-2014, http://www.the-cryosphere.net/8/2335/2014/, 2014.

Logg, A., Mardal, K.-A., and Wells, G.: Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book, Springer Publishing Company, Incorporated, 2012.

Loose, N., Heimbach, P., Pillar, H., and Nisancioglu, K.: Quantifying Dynamical Proxy Potential through Oceanic Teleconnections in the North Atlantic, Earth and Space Science Open Archive, https://doi.org/10.1002/essoar.10502065.1, https://www.essoar.org/doi/10.1002/essoar.10502065.1, publisher: Earth and Space Science Open Archive, 2020.

MacAyeal, D. R.: Large-scale ice flow over a viscous basal sediment: Theory and application to Ice Stream B, Antarctica, Journal of Geophysical Research-Solid Earth and Planets, 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.

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, 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.

Nias, I. J., Cornford, S. L., and Paybe, A. J.: Contrasting the modelled sensitivity of the Amundsen Sea Embayment ice streams, Journal of Glaciology, 62, 552–562, https://doi.org/10.1017/jog.2016.40, 2016.

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.

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

Rignot, E., Mouginot, J., and Scheuchl., B.: MEaSUREs InSAR-Based Antarctica Ice Velocity Map, Version 2, https://doi.org/10.5067/D7GK8F5J8M8R, accessed Jan 20 2021, 2017.

Ritz, C., Edwards, T. L., Durand, G., Payne, A. J., Peyaud, V., and Hindmarsh, R. C. A.: Potential sea-level rise from Antarctic ice-sheet instability constrained by observations, Nature, 528, 115–118, https://doi.org/10.1038/nature16147, 2015.

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, https://www.pnas.org/content/116/30/14887, 2019.

Rommelaere, V.: Large-scale rheology of the Ross Ice Shelf, Antarctica, computed by a control method, Journal of Glaciology, 24, 694–712, 1997.

Schoof, C.: A variational approach to ice stream flow, J. Fluid Mech., 556, 227–251, 2006.

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 Glaciol., 49, 91–95, 2008.

Shapero, D., Badgeley, J., Hoffmann, A., and Joughin, I.: *icepack*: a new glacier flow modeling package in Python, version 1.0, Geoscientific Model Development Discussions, 2021, 1–34, https://doi.org/10.5194/gmd-2020-419, https://gmd.copernicus.org/preprints/gmd-2020-419/, 2021.

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

Tierney, L.: Markov Chains for Exploring Posterior Distributions, Ann. Statist., 22, 1701–1728, https://doi.org/10.1214/aos/1176325750, https://doi.org/10.1214/aos/1176325750, 1994.

Vieli, A. and Payne, A. J.: Application of control methods for modelling the flow of Pine Island Glacier, West Antarctica, Annals of Glaciol., 36, 197–204, 2003.

Villa, U., Petra, N., and Ghattas, O.: hIPPYlib: an Extensible Software Framework for Large-scale Deterministic and Bayesian Inverse Problems, Journal of Open Source Software, 3, https://doi.org/10.21105/joss.00940, 2018.

Waddington, E., Neumann, T., Koutnik, M., Marshall, H., and Morse, D.: Inference of accumulation-rate patterns from deep layers in glaciers and ice sheets, Journal of Glaciology, 53, 694–712, 2007.

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, https://doi.org/10.1145/279232.279236, 1997.

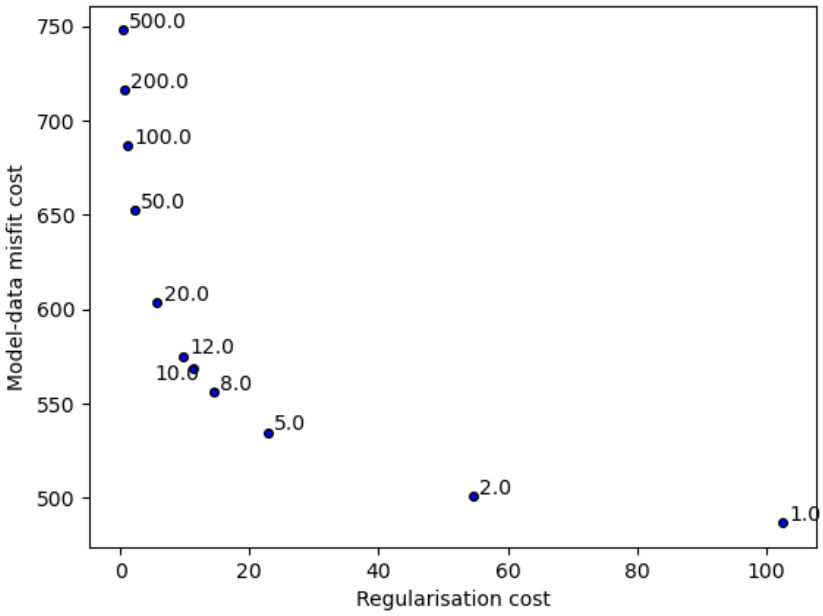

Figure 1. An L-curve showing the tradeoff between model misfit (first and second terms of Eq. 32) and the regularisation cost (the second terms of Eq. 32 divided by γ). Associated values of the regularisation parameter γ are shown. In all optimizations, δ is equal to 10^{-5} and observational points occur at intervals of 2 km.

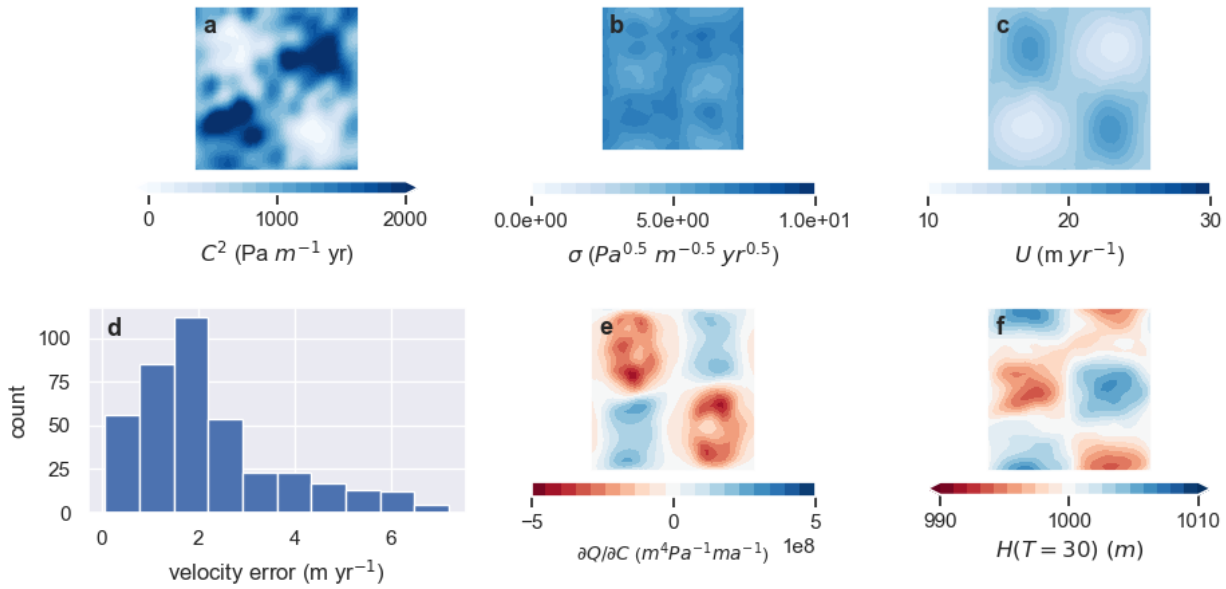

Figure 2. Results of the control-method inversion with $\gamma = 50$. (a) The recovered basal traction C^2. (b) The point-wise standard deviation of the sliding parameter C. (c) The surface speed associated with the inverted C. (d) Histogram of model-data velocity misfit (where misfit is the 2-norm of the difference of observed and modelled velocity). (e) The sensitivity of Q_{30}^{IS} to the sliding parameter. (f) Thickness after 30 years of time stepping.

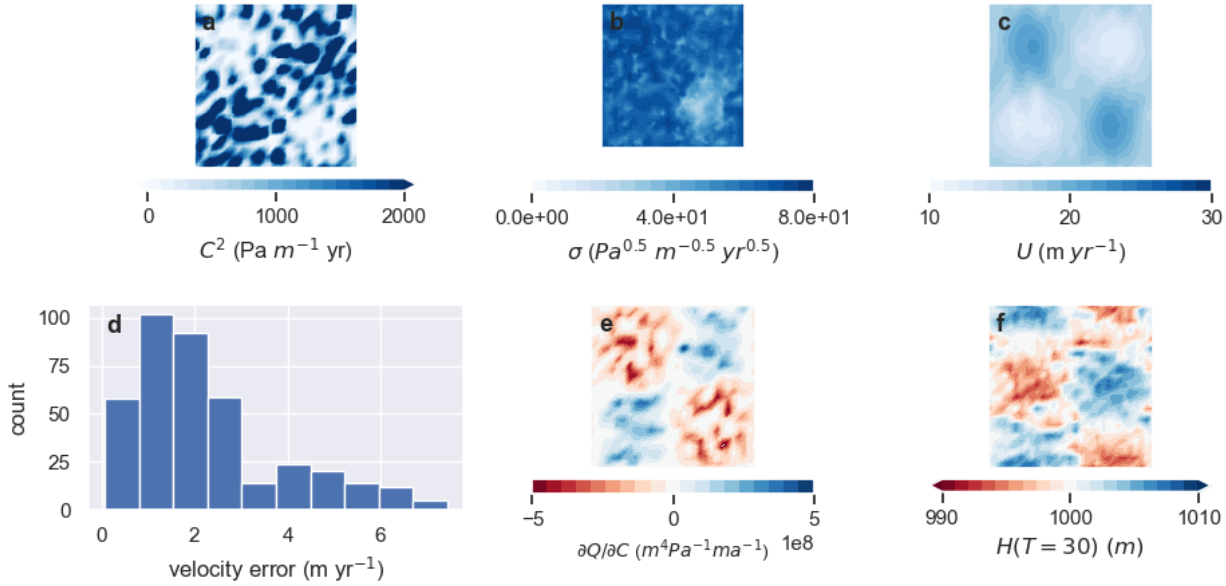

Figure 3. Similar to Fig. 2 but with $\gamma = 1$. Note the difference in colormap with Fig. Fig. 2(b).

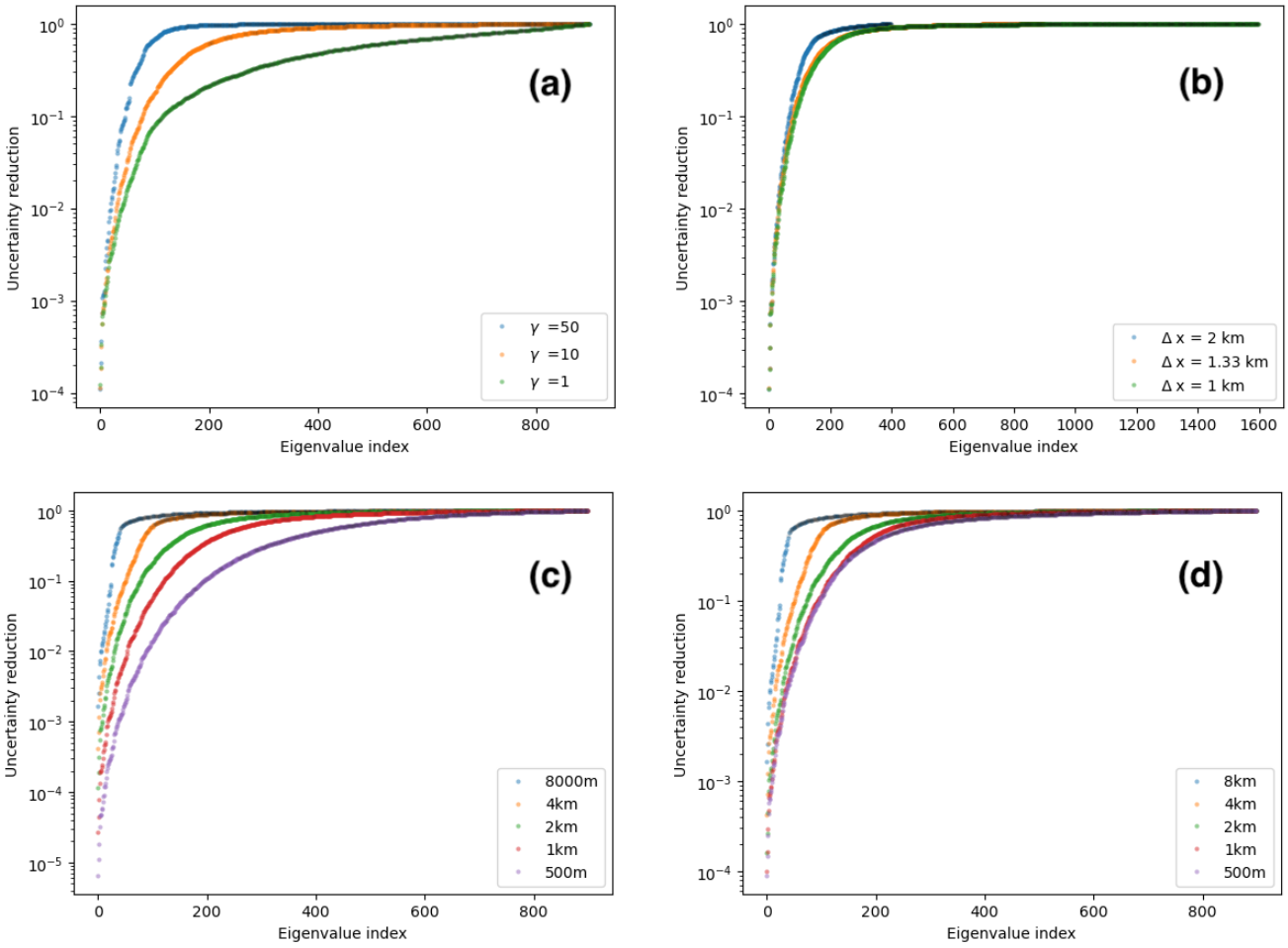

Figure 4. Uncertainty reduction factor $1/(1+\lambda_k)$ versus eigenvalue index k for a range of ~~experimnents~~experiments. (a) Dependence of reduction spectra on regularisation parameter γ_α. (b) Dependence of reduction spectra on model resolution. (c) Dependence of reduction spectra on density of observational sample points. (d) Dependence of reduction spectra on density of observational sample points with nonzero observational covariance.

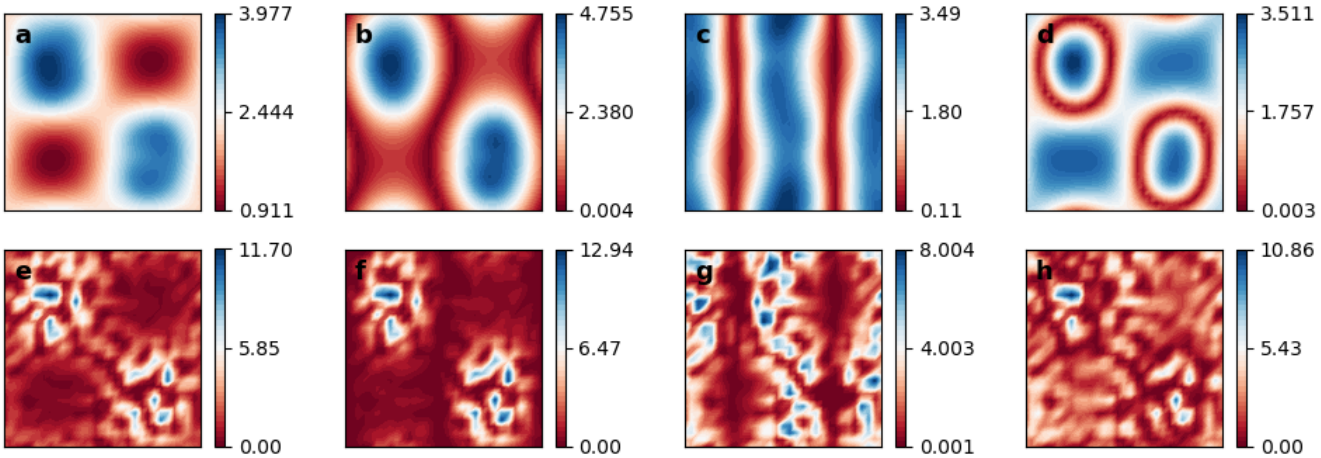

Figure 5. (a)-(d): Leading four eigenvectors of C in γ_{50} experiment. (e)-(g): Leading four eigenvectors of C in γ_1 experiment.

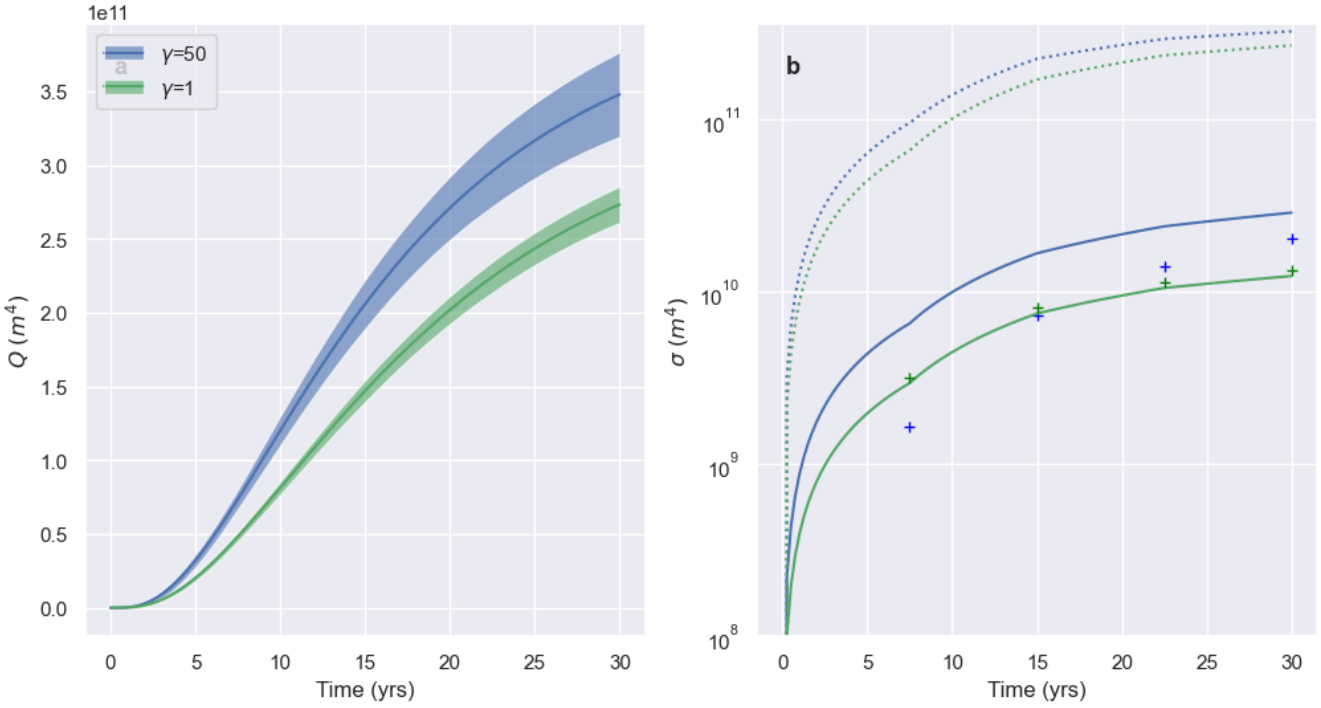

Figure 6. (a) Paths of Q_T^{IS} in γ_{50} experiment (blue) and γ_1 experiment (green). Shading shows the 1-σ uncertainty interval for each trajectory calculated by projecting the Hessian-based (posterior) uncertainty along the linearised trajectory. (b) Uncertainties in the time-dependent experiments over time. Dashed lines: prior uncertainties projected along the linearised Q_T^{IS} trajectory. Solid lines: Hessian-based posterior uncertainties projected along the linearised trajectory. Markers: Standard deviation of Q_T^{IS} from sampling the posterior density. In both panels, green corresponds to γ_{50}, and blue to γ_1.

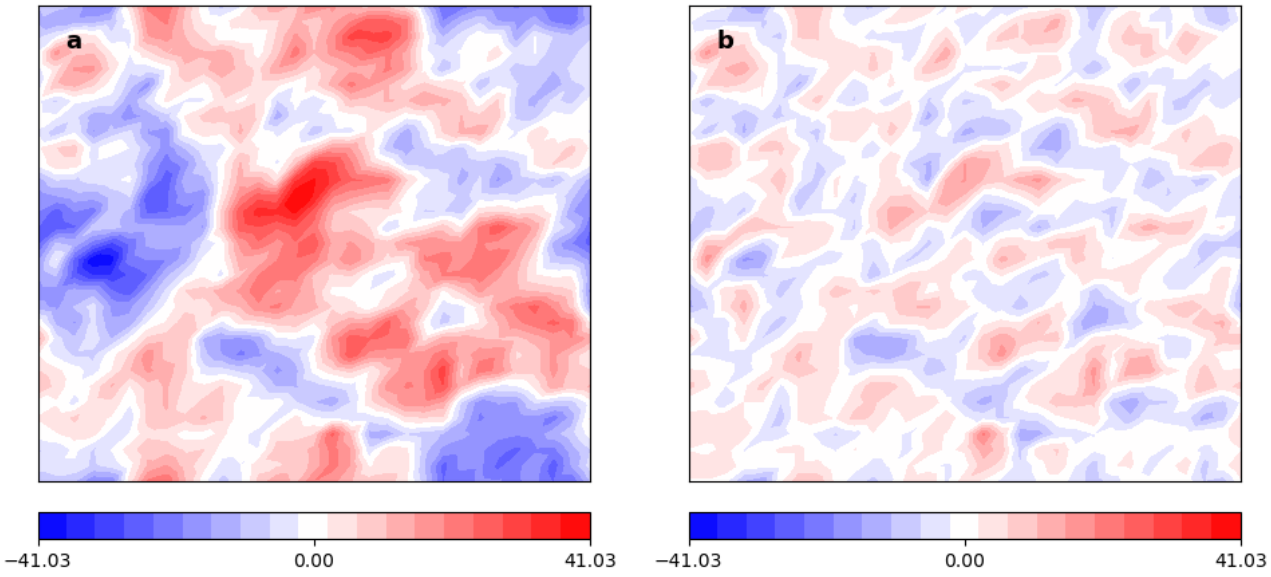

Figure 7. (a) A realisation of the prior density of C for the γ_{50} experiment. (b) A realisation of the posterior density of C for the γ_{50} experiment with mean $\overline{c}_{MAP}$ removed.

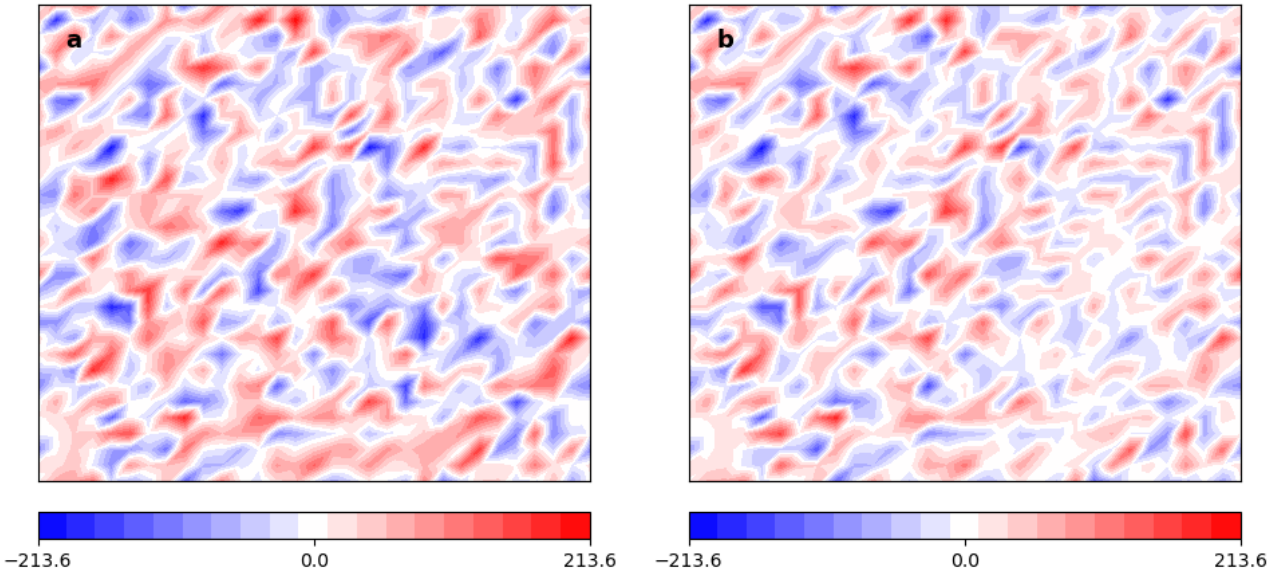

Figure 8. (a) A realisation of the prior density of C for the γ_1 experiment. (b) A realisation of the posterior density of C for the γ_1 experiment with mean $\bar{c}_{MAP}$ removed. Note the difference in colormap with Fig. 7.

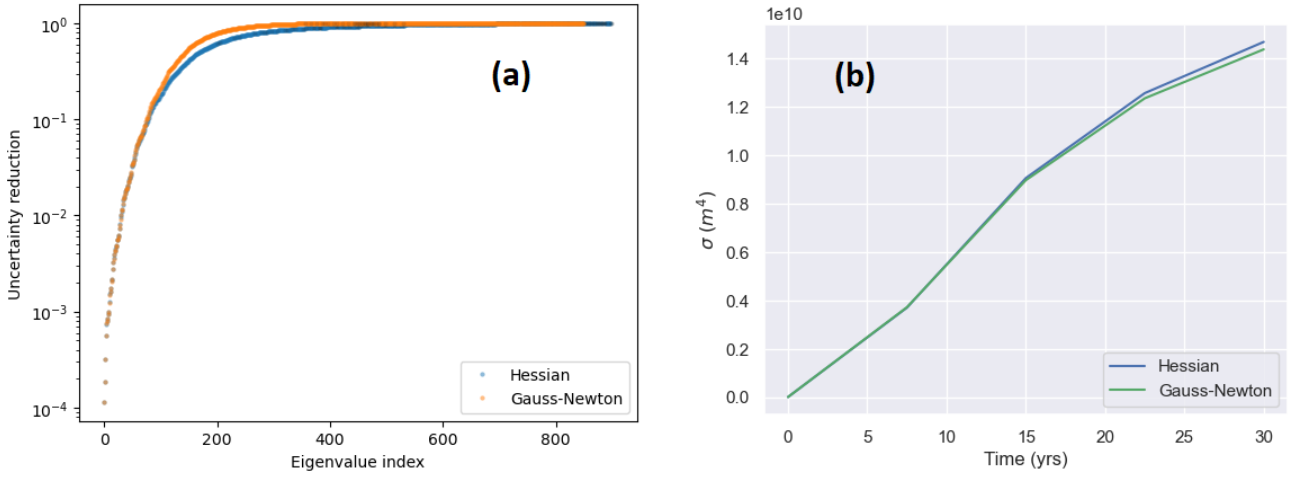

Figure 9. (a) Uncertainty reduction factor $1/(1+\lambda_k)$ versus eigenvalue index k for $\gamma = 10$ and observational spacing of 2 km, found with the full Hessian and Gauss-Newton approximation. (b) Hessian-based posterior uncertainties of Q_T^{LS} over time based on the full Hessian and Gauss-Newton approximations.