Review of 'An evaluation of multi-fidelity methods for quantifying uncertainty in projections of ice-sheet mass-change' by Jakeman et al.

Doug Brinkerhoff

August 20, 2024

In this manuscript, Jakeman et al. present the application of multi-fidelity uncertainty quantification to accelerate – and hopefully provide more accurate – estimates of first- and second-order ensemble statistics. They begin by describing two approximations to the Stokes' equations, which trade solution expressivity for computational expense and serve as the basis for their multifidelity methods. They next describe a mechanism by which to characterize an approximate posterior distribution (based on a low-rank Laplace approximation) over basal traction conditioned on surface velocity observation, which serves as the source of samples for Monte Carlo sampling of ice volume evolution, the primary quantity of interest in this work. The primary methodological advance is the introduction of an adaptive control variate (ACV) estimator for the mean and variance of mass change after approximately a century of ice evolution. This estimator leverages correlations between so-called low- and high-fidelity models (which have different computational expense) to effectively reduce the error in Monte Carlo estimates of ice volume change relative to predictions made using a limited number of high-fidelity model predictions on its own. They present this method for a single high and low-fidelity model, and then extend the analysis to the case where there exists one high-fidelity model and a hierarchy of multiple low-fidelity models. Such methods require the establishment of statistical relationships between the low and high-fidelity models, and also careful selection of the number of samples evaluated for each consituent model: the authors carefully present strategies for these tasks within the framework of a fixed computational budget, and explore the implications of these strategies when they are better informed by empirical analysis than theory. The manuscript applies these methods to the Humboldt Glacier basin of Northwest Greenland and show that the present methods can be used to perform effective uncertainty quantification, at least over the limited subset of uncertain parameters that the authors' consider.

This work is an important and timely contribution to the growing effort towards robust uncertainty quantification in ice sheet modeling. I have no major objections to the scientific content of this work, which I find to be well-motivated and defensible. I do think that the work relies on language and a presentation style that will be challenging for many readers, particularly those without a specialized statistical background. As a general comment, I would encourage the authors to try to provide more intuition and plain-language summaries, particularly in Section 4. I provide more specific examples, alongside other detailed comments below.

L26 Here and elsewhere, 'effects' should be 'affects'

L97 Should $H(x, y, z)$ be $H(x, y, t)$?

L98 The MOLHO (or the Blatter-Pattyn approximation) doesn't neglect vertical velocity, it is just eliminated from the system of equations via mass conservation and the assumption of hydrostatic pressure. It can always be determined a posteriori from the horizontal velocity components.

L127 I don't think Dukowicz is the best reference here. Pattyn (2003) is more commonly cited, or if the preference is for something that clearly describes the hierarchy of approximations, Schoof and Hindmarsh (2009).

L137 I think that citet should be used rather than citep here.

L152 The discretization of the continuity equation is non-trivial and should be described here. How was it stabilized? How was positivity (even in the absence of negative forcing) ensured?

L172 Who specifically considers friction to be such a large source of uncertainty? Plenty of recent work has shown that forcing terms are the most important uncertainty sources, particularly at long time scales. I don't have a problem with focusing on traction here, but I think it is important to contextualize this choice a little bit more fully.

L180 I think that the community is using the term 'Gaussian process' with some frequency now, so it would be good to at least mention that as a name for what is going on here (and a reference to, say Rasmussen and Williams (2006)).

L212 For what it's worth, it's a stretch to call BedMachine 'data' – it is the result of a PDE-constrained optimization scheme that relies on assumptions of climate, smoothness, and a variety of other things. Again, nothing different needs to be done, but it is important to state that this inferred geometry is assumed error-free.

L232 This scaling term is less mysterious when reported with units 'number of observations per area'.

L263–267 Here and elsewhere, please be sure to use a consistent tense. This switches from present to past inside a sentence.

L298 Should these sums have N^{-1} in front of them?

L298 For the inner sum, please use a different index variable than n.

L298 Is the optimization of η mandatory or does this work with arbitrary η? What is the objective that is optimized?

L298 The reader would benefit from a description of what this equation means and some intuition of why this works. It appears to be that the low-fidelity terms yield a correction to the high fidelity statistic, but it is somewhat surprising that this doesn't need to include any explicitly quantified relationship between the two models. It would also be helpful to emphasize that the Θ_0 and Θ_1 can have different set sizes.

L323 The line about some samples being shared is vague. Please elaborate on what this means.

L328 'statistics' → 'statistic'.

Eq. 14 As before, are there alternatives to using this value for η?

Eq. 16 Split this into two equations, and add matrix sizes for each.

L352 I'm not sure I understand this sentence.

L356 I think it would be better to include more detail about how these expressions are used to compute Eq. 15 than just referencing Dixon (2023). Otherwise, it sort of feels like a lot of space gets used describing Eqs. 16 and 17, but they never really go anywhere.

Eq. 20 Why is it the case that minimizing the determinant of the covariance determines an optimial sampling strategy?

Fig. 4 and accompanying text I don't think that the text does a sufficient job of describing the principles behind these different sampling strategies. Figure 4 tells me that Θ_0 and Θ_1^* share their samples, and different schemes use entirely different or appended different samples for Θ_1, but I cannot grasp from the text why this is significant. This needs to be motivated fully or de-emphasized and more carefully referenced.

L394 'model' appears twice.

L433–439 Is there an argument that can be made here to reassure a reader that the observed changes are due to real climate/ice dynamic effects and not so-called 'transients' resulting from inconsistency between initial conditions, physics, and input fields? Fig. 7 (left) has some rather surprising high-frequency noise in the surface elevation change – it would be helpful to know where this comes from.

L453 Just to clarify, was Θ_{pilot} shared across models, or were the samples different for each model?

L458 'significant' is an unfortunately subjective term here - to my eye, the differences in the heights of the referenced bars seems rather insignificant. Is it possible to elaborate on the meaning of 'significant' here, and why it should be viewed as such?

L467 This sentence is a bit challenging, with 5(!) nested prepositions and two uses of 'variance' each describing different things. I recognize that it is challenging to compactly describe the statistics of statistics, but is it possible to relax this sentence a bit?

L476 This assertion is surprising to me, and a citation describing the assumption of pilot statistic exactness would be useful.

L494 Is there an interpretation of why some models appear to be more informative than others, or is this just random chance? I can't identify a mechanism for why some low-fidelity models were chosen more frequently, but understanding that (or being able to predict it a priori) would be exceptionally useful.

L496 I don't understand what is meant by a 'hierarchical relationship' here.

L525 I get where these numbers come from after some digging back through the other sections, but it would be helpful to remind the reader where each of the terms in the cost expression represent.

L536 Is this extreme asymmetry between the number of high- and low-fidelity model evaluations typical? I think that this is a significant and interesting result if the high-fidelity model is only really needed to, e.g. characterize the spatial variability of the mean solution, but the low-fidelity models are sufficient to characterize all of the uncertainty.

Table 1 I don't think this needs to be a table.

L589 The variance doesn't have the same units as the mean, so I'm not sure what numbers I'm looking at. Is 17.68 the standard deviation?