The manuscript, An evaluation of multi-fidelity methods for quantifying uncertainty in projections of ice-sheet mass change by Jakeman et al, uses a new computational approach to determining the posterior uncertainty of ice mass change in a glacier forecast conditioned on observational data and uncertainty. The main contribution is a Multi Fidelity Uncertainty Quantification (MFUQ) scheme which samples from a probability distribution (see below) and provides an inexpensive means of Monte Carlo variance reduction in the calculated statistics that requires far less simulation time. This is achieved through generating ensembles from models that are of lower fidelity (coarser resolution / longer time steps) whose dependencies on the input parameters are similar. The probability distribution which is sampled -- that of the sliding parameter conditioned on observations and model physics -- would be too expensive to find via Monte Carlo methods. Rather, a method introduced by others in the literature -- which approximates this posterior as Gaussian and finds a low rank approximation to the inverse covariance matrix to make the problem tractable -- is used.

The methodology introduced in the paper – the MFUQ scheme – is fairly well described and seems quite useful, and its results deserve to be shown.

However, there are a number of major issues I have with the manuscript. Aside from a number of writing issues, such as inconsistent statements and introducing of terms and symbols without definition or explanation (see specific comments), I feel that the messaging of the paper in the introduction is not in line with what the authors have actually done. Furthermore they have downplayed or overlooked recent works in the literature – works which, in some cases, bring the methodology of this study into question. I will highlight these in general comments below.

Finally I should point out first though that it monte carlo methods are not my area of expertise. I have some specific comments about certain things that looked as thought they might be typos or need more explanation. Largely however I do not have much to say about the actual MFUQ methodology and its presentation, and I hope that other referees can assess it better.

General Comments.

1. The paper sets out to deal with parametric uncertainty, which is the case. But the introduction is written in a way that makes it seem that MFUQ is used to solve the "full" problem – that is, quantifying the probability density of mass change conditioned on the model and observations, which can be termed $p\ (Q \mid m, U)$ where Q is the mass change, m is the model and U is the observations. But in truth a different method (Hessian-based) was used to find the posterior density of the frictional field θ, and then this was sampled from to find the posterior of Q i.e. $p\ (Q \mid m, \theta)\ p\ (\theta \mid m, U)$ -- and $p\ (Q \mid \theta)$ is the only component being determined by MFUQ.

I think this could be potentially very misleading and give the impression that MFUQ is capable of the "full" problem when from the results of the paper it definitely is not. This is very important: given the newness of the fields of ice-sheet modelling and ice-sheet uncertainty quantification there is extensive misunderstanding about which problems can be tackled by sampling methods and which require alternative methods. Although this is somewhat covered in lines 61-71 of the manuscript, the passage requires familiarity with the field and with both MC and Hessian-based UQ. It needs to be much more clear – with mathematical formality – which distribution is being quantified using MFUQ.

2. The manuscript is also misleading about contributions in this paper versus in the literature. Specific examples are given below, but the manuscript does not acknowledge previous authors' attempts to quantify the uncertainty of high-dimensional parametric uncertainty. In particular, a recent paper in The Cryosphere (Recinos et al, 2023, hereby shortened as BR23) has been overlooked. The authors can certainly be forgiven for this of course as the paper came out only last year, but it is extremely relevant to many of the assumptions and calculations within the manuscript (and is mentioned extensively in the specific comments below). Additionally, based on this paper there are several assumptions and/or approximations that give me serious reservations about this paper's results – these are easily identifiable in the specific comments where BR23 is mentioned.

3. The underlying premise of the paper is that, given a Hessian-based approximation of the posterior parameter density has *already* been carried out, "traditional" means of sampling from this posterior density is too expensive. But another such approach – using the *linearization* of the mass change model $f(\theta)$ (using either Automatic Differentiation or some other form of differentiation) to project the posterior uncertainty of θ onto the quantity of interest – exists, and is not at all mentioned. Playing devil's advocate, such an approach assumes near-linearity of $f(\theta)$, but linearity has already been assumed in the posterior calculation of θ. Moreover at least two prior papers – Isaac et al (2015) and BR23 – have used this method (see eq. 24 of Isaac et al 2015, or eq. 15 of BR23), and the latter comprehensively tested the linearity assumption. Given this, I would expect acknowledgement of this very relevant and related approach, its drawbacks and benefits, and fit (or lack thereof) to the current problem.

Specific Comments.

L23-25. This is a good outlay of the different sources of uncertainty. What is missing is a definitive statement that the only type of uncertainty being quantified in this paper is parametric uncertainty.

L26-27. "but the impact of discretization errors has not been explicitly considered with other sources of uncertainty". And it has not in this study either, right? As I understand it the MFUQ scheme is solely to estimate parameter uncertainty of the 1km, 9-day MOLHO model – it did not quantify disc. uncertainty despite using different discretizations.

L37-41. This is a good place to cite works such as Isaac et al 2015 (and various papers by Noemi Petra e.g. Petra et al 2013), and BR23.

L60-61. As noted above, quantifying the impact of a high-dimensional parameterization of basal friction on long-term projections is not novel (cf. BR23 – unless you are distinctly saying that 40 years is not long-term and 80 years is!)

L62-64. As noted above, Isaac et al, whose methodology you cite and use, arguably did this.

L66. Im not sure why you include Isaac 2015 in a list of papers using low-dimensional parameterisations – they used $O(10^6)$ parameters in their basal sliding parameterization.

Fig 2, 3, 5, 6, 7, and 13: you need to show the coordinate axes in all visualisations of the model domain – and there should be one figure showing the placement of Humboldt in Greenland.

L181: "covariance" – prior or posterior?

L190-193. I have deep concerns about your parameter choices. Firstly, what is the pointwise variance? Secondly, how did you arrive at this correlation length as suitable – on what basis? I do not see any physical reasoning leading to it. You are saying that the data essentially does not need to constrain variability on a scale smaller than this, which I don't think is an accurate statement. BR23 chose far smaller autocorrelations (~3km) using some degree of physical inference, and moreover showed that it was necessary to give reasonable values of posterior uncertainty (see comment on TABLE 1 regarding this assessment), and it is possible that in choosing such large numbers you are making the posterior uncertainty artificially small by choosing an overly-informative prior. This may be why you only needed < 1000 eigenvalues to represent the posterior as shown in the appendix. (see BR23 for details.)

L205. How did you generate mass balance? Did you run a regional climate model that incorporates firn and snow processes? If so, say so. Did you use a parameterization? If so, state it and the source.

L231. On what basis do you assume they are uncorrelated? The fact that the products are not posted with spatial correlations of error is not a reason – this is simply too difficult for them to calculate. Please highlight this, and state what the consequences of such an assumption could be for estimating posterior uncertainty.

Section 4: in general I think this section should be read over very carefully to look for typos and variables introduced without definition. Ill mention several below but these sections (the ones that I read closely) seem to have been written hastily.

L263, mean Q^{μ}: mean of what?? And what is Q? and are these "true" statistics or estimators since they have no subscript?

L265. Try to be consistent with tense throughout, and definitely within a sentence: "The second step simulate**s** the model at each realization … and comput**ed** the mass change.."

L271: "Any MC estimator Q" – do you mean Q_{α}^{μ} or $Q_{\alpha}^{\sigma\wedge 2}$, or both or neither?

Eq 11 – can you show how this is derived? At first glance it looked similar to the identity $E[(X-E[X])^2] = E[X^2]-(E[X])^2$ but I could not derive it using similar reasoning.

L279 did you mean MSE (II), rather than MSE (11)?

L279: I don't believe that all of these sources of uncertainty go into the bias term. My interpretation is that, for the purpose of your MFUQ, you are given a density of θ arising from the Isaac methodology. You then have a deterministic function $f_{\alpha}(X)$ which is given by your high fidelity model and its discretization, and is therefore deterministic. You are seeking properties of the probability distribution *induced by* f_{α} and the only actual uncertainty is how fast the MC converges. Model uncertainty and discretization uncertainty, while very real, are not accounted for in such a calculation.

L280 what does MSE (10) mean?

L280 ensures, for any set of model input samples,

First eq in 4.2.1 (not numbered) – is the 2nd term in brackets not divided by N1?

L316 – QoI not defined previously.

L324 – for the union of these sets to be null, both need to be null. Should it be an intersection symbol?

Eq 18. You seem to be estimating these statistics using straightforward (Naïve) MC. Why is this OK given the whole thrust of your study is that MC is too expensive to apply to the statistics of the ice model?

L418-422. State # of elements In models

L424 in the 1st para of 2.4 you state you use FEniCS. MALI is a C++ model with Fortran libraries and not, to my knowledge, written with fenics. **Which model(s) did you use???**

Fig 10 – I might be misunderstanding the methods but shouldn't there be units??

Table 1. This value is presented without validation. It is possible to do a "sanity check". BR23 use 2 essentially independent measurements of velocity (ITS_LIVE and MEaSUREs) to invert for parameters and simulate mass loss. If the difference seen is of almost negligible probability under the calculated posterior for mass loss – then there must be an issue with the calculated posterior. You are capable of doing this as well…

L550. "the SSA model was not.." can you provide an example or evidence of this?

L565. Im confused – I thought that the MFUQ was needed as you are sampling from a distribution of ~600 dimensions (the number of Eigenvals retained in the Hessian based UQ). If you have only 10 dimensions can you not use standard (naïve) MC?

L567: Appendix B

References

Recinos, B., Goldberg, D., Maddison, J. R., and Todd, J.: A framework for time-dependent ice sheet uncertainty quantification, applied to three West Antarctic ice streams, The Cryosphere, 17, 4241–4266, https://doi.org/10.5194/tc-17-4241-2023, 2023.

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

Petra, Noemi, et al. "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 (2014): A1525-A1555.