Referee comment on ‘Numerical issues in modeling ice sheet instabilities such as binge-purge type cyclic ice stream surging’

March 2023

Summary

In this manuscript, the authors explore the effect that a wide variety of factors, both numerical and physical, have on the time-dynamics of a synthetic thermomechanically coupled ice sheet constructed as a proxy for the Laurentide ice sheet during the last glacial period. In particular, they seek to understand which modelling choices lead to effects that rise above the well-documented sensitivity of such models - when a dependence of sliding on temperature is included - to symmetry breaking and chaotic dynamics from discretization.

I think that this manuscript has the potential to be a useful contribution to the (slowly) growing body of literature on the relationship between thermodynamics and sliding. However, I think that this manuscript needs quite a bit of work to be publishable. My criticism is not generally aimed at the science, but rather at the presentation which reads more like a set of internal notes than a manuscript. In particular the organization does not proceed in a logical manner, and I struggle to find a narrative that relates adjacent sections. In addition, the manuscript includes difficult-to-remember jargon and sometimes dubious interpretations of referenced material. There are far too many tables; it is very difficult to make sense of tabular numbers appearing in columns that reference one of many difficult to differentiate experiments. Perhaps most critically, some of the most important figures - many of which are referenced frequently and are essential to the paper’s conclusions - are relegated to supplemental material or appear in one of the authors’ previous papers. I do not suggest any new experiments or analysis here. However, I do suggest that the authors revise this manuscript with a keen eye for the fact that it will, ultimately and hopefully, be read by others.

Line by line comments

Title The title should be shortened to something like ‘Numerical issues in modeling binge-purge behavior in ice streams.’ No other instabilities are addressed and 'binge-purge', 'cyclic', and 'surging' are all redundant.

L1 Delete 'as in any environmental system'.

L20 I think it is strange to lead with validation as a means of motivating the current manuscript, when no validation occurs here. Validation is not the same as sensitivity testing.

L26 What is 'numerical noise'? Something random? Pseudo-randomness in a chaotic system? Numerical *error*? This is a critical consideration but it's not really clear what it means in this paper here and elsewhere.

L38 Define 'meaningful'

L44 The quote from Soucek and Martinec is relevant, but it misses the fact that there are a great many approximations that appear in, for example, the solution of the Stokes' equations. Why the emphasis on numerical rather than model error?

L60–72 While I recognize that this paper focuses on the ISMIP-HEINO setup, it would be worthwhile to try to contextualize this work with respect to the EISMINT-F experiments as well. There is a great deal of insight there regarding thermal sliding instabilities and the circumstances under which they appear.

Sec 1.1 I find the organization of the paper according to research question to be quite challenging to follow, perhaps mostly because there are so many (11) research questions. I think it would be better to group these into open questions rather than yes/no, and this might make for more comprehensible themes. For example, Q1,11 could be grouped as 'what aspects of simulated surges are due to numerical considerations?', while Q2–6,9 could be grouped as 'what modeling and solution choices influence surging?' and Q7,8,10 as 'what parameterizations of basal physics leads to the most robust conclusions?' or something similar.

L166 and elsewhere I don't find it helpful to reference a manuscript that is 'in preparation' because such manuscripts are not readable and sometimes fail to ever get published. Is there some source code that could be referenced instead? An instruction manual? An older manuscript from which the ideas in the in prep manuscript are adapted?

L198 I am deeply skeptical of a model that 'activates' stress terms based on a heuristic that in turn depends on whether the bed is soft or hard (whatever that means). Does it not seem that such an obviously non-physical choice could lead to just as much variability in surging behavior as any of these other mechanisms? Validation is mentioned in the introduction, but what about verification? How does the reader know (especially given that there is no current reference to the model description) that this model converges to the true solution of *some* physically and mathematically justified system of equations under discretization refinement?

L215 I don't know what 'legacy' means here.

Eq. 5 Is this supposed to be $F_{T_{ramp}}$? Otherwise F_{warm} is defined twice. Also, I think it's really awkward (ignoring subscrips) to have F depend on T, which depends on a different F. Maybe consider different notation?

Sec. 2.1.2 I generally find the phrase 'vector' to be unhelpful here. I think it would be better to describe how the ensemble is created (i.e. by selecting different values for each of eight parameters) and then referring to different members of the ensemble as, well, 'ensemble members'.

Sec. 2.1.2 It takes quite a bit of flipping around to understand why we're talking about ensembles at all. I think this section could use a clear explanation of the fact that you're running each subsequent experiment with multiple different parameter settings.

Sec. 2.1.3 I think that 'reference simulation' might be more clear than 'base setup'.

Sec. 2.1.3 I think that the very frequent referencing to future sections is not very helpful.

Sec. 2.1.3 and elsewhere There are far too many references to supplementary information in this manuscript. SI is intended for things that either cannot appear in the manuscript itself due to medium (e.g. code or videos) or that offer additional insight or detail into some aspect of the work but that is not essential to the results. In this case, the mass balance forcing (the single most important thing in determining long term ice sheet extent) is relegated to the supplement, but really should be in the main text.

2.2.1 It's strange to imply that PISM is *not* also optimized for computational speed. It's the *parallel* ice sheet model, after all.

L303–310 I think it would be helpful not to mix units of measurement (m/a and m/d). What is a 'stable solution of the numerical matrix solver'? Is 'observed range' the heuristic from Cuffey and Paterson?

L318 'event' and 'HE' seem to be used interchangeably in the manuscript. I think it would be better to just use 'HE'.

L330 This is another circumstance where including the supplemental figure in the main manuscript would be very helpful.

L341–342 'ice-free when no surge occurs.' I'm not sure it's possible to ascribe a date to when something doesn't happen.

3.1.1 What about PISM? Can surges be understood similarly to those in GSM?

3.1.2 I'm not sure it's reasonable to try to state a specific justification for why the time scales of this highly idealized and not-observationally-constrained experiment are dissimilar to geological records: many different factors may be in play here (some improving the fit, some to its detriment) and (as an example) saying that the period mismatch is just the result of domain size seems likes its missing a broader set of possibilities.

3.1.3 I again struggle with the notion of 'numerical noise'; the paper would be well served by having a much more in-depth description of what is meant by this and where it comes from. With respect to the latter, in most modelling exercises, the numerical error is something that can be well quantified through comparison to exact solutions or by a theoretical analysis of the interpolation properties of numerical method. Yet here, what we're effectively measuring is the system's sensitive dependence to small perturbations. In that sense, it doesn't necessarily follow that stricter convergence criteria (in the case of GSM) would necessarily lead to any less 'noise'. Could an equivalent result be achieved by just adding white noise to the initial conditions?

With PISM, it's also somewhat concerning that the number of cores used is a source of noise. Why is parallelism leading to different numerical solutions? I suppose it probably doesn't matter, but why not use the same mechanism of tolerance change to get modified runs as in GSM?

3.1.5 This section on implicit coupling is so vague as to be useless. What even is 'implicit coupling' in this context? Is this the same as implicit time stepping, i.e. Backward-Euler?

3.2.3 I can't figure out what TpmTrans, TpmInt, or any other Tpm thing are. If they are described earlier, such a description needs to be here instead or also. If they are not, they need to be defined (and not in the supplement).

3.2.4 It's not my preference, but if you prefer to have the actual equations describing GSM in the supplement, I suppose that's fine. However we need at least a little bit of a qualitative description of what these different 'weights' imply. What is the context for understanding why these different choices should yield different surging behaviors?

Fig. 7 I honestly can't figure out what this figure is trying to convey. Part of this is that I also can't figure out what the part of the text that references it is trying to convey either (L534–540). Please try to make this a little bit more clear.

3.3 Brinkerhoff and Johnson, 2015 suggest that the inclusion of membrane stresses leads to convergence under spatial grid refinement, whereas without them, the SIA does not lead to convergence. Can you place those results in the context of this section? Are the relatively weak convergence results here a result of GSM and PISM velocity solvers being insufficiently 'higher-order'?

Sec. 4 If you maintain these research questions as an organizing principle, I would like to see them revisited as they are resolved by the experiments rather than all at once at the end.

L738–746 I think that this section is kind of weird: none of the results presented in this work actually refute the resolution dependency conclusions of Hindmarsh (2009) or Brinkerhoff and Johnson (2015), yet the paragraph is written as if they did. As mentioned before, it seems just as reasonable to assert that those works saw more robust numerical convergence due to the use of more consistent membrane stress resolution schemes rather than because they fortuitously (or nefariously) made parameter choices that suppressed resolution effects.