Author's response to Anonymous Referee 2 Comment 2

August 8, 2023

1 General Comment

[This is an updated version of the manuscript with new title "Numerical issues in modeling thermally and hydraulically driven ice stream surge cycling" by Hank et al. It is greatly improved based on revisions responding to prior reviews. I think it has more logical flow and clearer messages about the importance of various modeling choices in simulating ice stream surging.

In some ways I think the revision does not quite go far enough in paring down the level of detail in the manuscript and the amount of repetition in describing the results (mainly in the amount of recapping of results in section 4 right after they are described in a similar level of detail in section 3). However, I also recognize that stylistic suggestions regarding the structure of papers are ultimately up to the authors. As someone who has worked on this problem and is predisposed to being very interested in the topic, I found it hard to get through the 40+ pages of painstaking detail on this suite of modeling experiments as a linear reading experience. It may be that most readers will use this paper as a reference guide, without reading it in its entirety. Even in such a form, it will be a useful contribution to the field. Ultimately it is up to the authors to decide whether it is worth substantially reducing the level of detail in sections 2 and 4. In section 2, there are places where modeling detail not pertinent to the questions defined in this study could be moved to a supplement or even just cited out to a previous paper. In section 4, there are many descriptions of results that are also given in section 3 which could be removed. Section 1.3 could be greatly shortened by just stating the questions you set out to answer. The first 20 lines of section 5 recap results right after a section that is largely recapping of results. Overall, I think it is possible for this manuscript to be 25 % shorter without losing much of the important detail of interest to readers.

One other major concern that remains from the first review is that it does not seem like a parameterization should be explicitly made to be resolution-dependent. It is unclear that the "optimal" temperature ramp parameters for coarse resolution would be the same between GSM and other models, and its unclear what is gained by designing the parameterization this way since it will undoutedly affect many others metrics beyond those which are the focus of this study. Additionally, it appears that the result of "compensation" of temperature ramp parameters for resolution difference are not even definitive, so why not just exclude this aspect of the discussion?]

We thank the referee for their constructive comments. We have addressed all the referee's comments and revised the text accordingly. A point-by-point reply is reported below, with referee comments in orange and our replies in black.

To reduce the length of the paper, we removed the repetitive parts in Section 4. Furthermore, we removed all details in Section 2 that are not directly related to one of the experiments in this manuscript. The maximum time step size experiment was removed entirely, as it did not add anything new to the literature. Additionally, we removed the first part of the conclusions (Section 5), as this information is also given in Section 4. We refrain from just stating the research questions in Section 1.3 because we prefer to have at least some context of why these questions are important and why we are trying to answer them. However, we shortened the description to the essential information. Overall, a reduction of 25 % (or 10 pages) is not possible without removing essential information or entire experiments, especially since the reduction was partly offset by some of the requested revisions (e.g., replacement of Figure 12). Nevertheless, we were able to reduce the length of the paper by 4 pages.

The arguments regarding the necessity for a basal temperature ramp and its resolution dependency are clearly stated in the text and would apply to any discretized glaciological ice sheet model. However, we agree that, e.g., different ice dynamics might lead to slightly different "optimal" basal temperature ramp parameters. We make this clear now with *However, given potential dependencies on the particular ice sheet model, we recommend resolution testing to determine the optimal basal temperature ramp.*. Since this is discussed in Section 4, we removed it from the conclusions.

2 Detailed comments

Responses to comments requiring a detailed response are listed below. Otherwise we have fully implemented the specific suggestions of the reviewer and do not list them below.

[Throughout: the term parameter "vector" is more typically called a parameter "set"] Since "parameter vector" is also common usage and technically accurate, we favor this.

[Ln 8: high → fine] We replaced all instances of *high horizontal grid resolutions* with *fine horizontal grid resolutions*.

[Ln 17: delete sentence one] and [Ln 17: The relevance of ice sheet modeling...of numerical implementations.] The first sentence was included to emphasis that despite the frequent use of ice sheet models within the glaciological community, the numerical aspects are often a *black box* for their users. To keep this emphasis, we decided to stick to the current version: *The use of ice sheet models has grown at least an order of magnitude over the last two decades. The relevance of such modeling studies to the actual physical system can be unclear without careful consideration and testing of numerical aspects and implementations.*.

[Ln 21: significant → realistic] While it is difficult to assess whether model results are physically realistic (especially in a paleo context), this is not the point we want to raise here. Furthermore, the results can be within a physically realistic range but still be a consequence of model-specific numerical choices. Therefore, we keep *physically significant*.

[Ln 31: work → heat] and [Ln 34: work → heating] Ln 31 is correct as stated *heat from geothermal and deformation work sources*. For Ln 34, we now state *increase of heat from deformation work*.

[Ln 34: delete "or to"] Changed: *can warm the surrounding ice close to or to the pressure melting point*

to: *can warm the surrounding ice (close) to the pressure melting point.*

[Ln 44: delete "quasi"] Since the periodicity is often irregular, we prefer to keep *quasi*. However, to clarify this, we updated the reference from [Robel et al., 2013] to [Souček and Martinec, 2011].

[Ln 84: model equations are not a numerical choice, they are a component of the system this is being modeled.] We agree that the wording was too vague. Normally, anything required just because of discretization (in time and space) is not a model equation. To make this clearer, we updated: *In this study we seek to disentangle the effects of numerical choices (both in terms of model components and in terms of their implementation) on ice sheet surges.*

to: *Herein, we disentangle the effects of numerical choices (e.g., grid size) and physical system processes (e.g., sub-temperate basal sliding) on ice sheet surges via numerical experiments.*

[Ln 92: delete "high variance"] The "high variance" is a key aspect of our experimental design, as we try to *partly address potential non-linear dependencies of surge cycling on model parameters*. Therefore, we prefer to keep this phrase.

[Ln 213: shelf-shallow ice] Changed to: *hybrid shallow shelf/ice physics*.

[Ln 309 (and throughout): the term "one-at-a-time" is not really necessary (it doesn't matter whether you sample in serial or in parallel)] To avoid the apparent confusion, we've change *one-at-a-time* to *one-factor-at-a-time.*

[Ln 439: do the surge characterisitics change with the threshold for activating SSA? I was surprised that this numerical choice was not tested, given how exhaustive this study is otherwise.] The effect of different SSA activation thresholds was tested and is briefly decribed in Sec. 2.1.1: *The hybrid SIA/SSA ice dynamics are activated for grid cells with a SIA velocity exceeding* $30\ \mathrm{m\ yr^{-1}}$. *Changing these activation velocities (*$20\ \mathrm{m\ yr^{-1}}$ *and* $40\ \mathrm{m\ yr^{-1}}$*) has no significant effect on the surge characteristics (Table S1). Activating the SSA everywhere leads to more, shorter, and weaker surges because no threshold velocity needs to be overcome to initiate basal sliding (Sec. S1.2). Note that we set an upper limit of* $40\ \mathrm{km\ yr^{-1}}$ *for the SSA velocity to ensure that sliding velocities stay within a physically reasonable range.*

The exact results can be found in Sec. S1.2 of the supplement.

[Ln 457: "numerical induced bifurcation" is not the right terminology. This behavior is typically called "spontaneous symmetry breaking" and it was previously identified in the context of ice stream in Sayag and Tziperman, 2011)] Thanks for bringing this up. We have updated: *Since the GSM setup and climate forcing are symmetric about the horizontal axis in the middle of the pseudo-Hudson Strait (*$y = 250$ km *in Fig. 1), we interpret the induced asymmetry as a numerical induced bifurcation.*

to: *Since the GSM setup and climate forcing are symmetric about the horizontal axis in the middle of the pseudo-Hudson Strait (*$y = 250$ km *in Fig. 1), we interpret the induced asymmetry as 'spontaneous symmetry breaking' similar to the results described in Sayag and Tziperman [2011].*

[Ln 473: early on you say that you won't be comparing the GSM to the PISM results directly, but then you do here?] To avoid debate over what *comparison* means/entails, we've changed the original sentences from: *As the two model setups and physics are somewhat different (see Table 2 for details), we do not intend to compare model results directly. Instead, our aim is to increase confidence in model results by showing that the same conclusions can be drawn from two different models.*

to: *As the two model setups and physics are somewhat different (see Table 2 for details), this permits more confident conclusions that are not model specific.*

[Ln 502: there should be an explanation in the main text as to why the number of cores affects the solution. Many readers while find this to be concerning.] Agreed. We have added the following sentence to the revised draft: *These minor differences can be caused by, for example, a different order of floating point arithmetic operations and the processor-number-dependent preconditioner used in PISM [PISM 2.0.6 documentation, 2023].*

[Ln 509: given that both models have tolerances that can be changed, why not use the same criteria for defining MNEEs in both models? This difference seems to make things unnecessarily complicated.] As mentioned in your comment above, many readers will expect the same or at least similar results for different numbers of cores. We decided to use the number of cores for two reasons. 1) To emphasize that in a highly non-linear system such as the one examined here, even the smallest differences can lead to substantial differences in surge characteristics. 2) To show the potential numerical sensitivity of the default PISM setup, likely blindly used by many ice sheet modellers, to prompt the community to pay more attention to numerical issues.

[Ln 515: is this just noise in the initial condition or at every time step?] It is noise at every climatic time step. We updated: *Adding low levels of uniformly distributed surface temperature noise (*$\pm 0.1^\circ C$ *and* $\pm 0.5^\circ C$*) to the climate forcing does not significantly affect the surge characteristics for the GSM (Table S3).*

to: *Adding low levels of uniformly distributed surface temperature noise (maximum amplitude of* $\pm 0.1^\circ C$ *and* $\pm 0.5^\circ C$*) to the climate forcing (updated every* 100 yr*) does not significantly affect the surge characteristics for the GSM (Table S3).*

[Ln 527: but is the implicit coupling scheme more accurate (for example with respect to grid spacing)] A priori, the implicit scheme will be more accurate. But given the highly increased computational cost (increase in the run time of ~ 265 %), a repeat of the horizontal grid resolution convergence study with the implicit coupling scheme is beyond the bounds of this paper.

[Ln 634-651: the discussion here is confusing and could be shortened and clarified] To make this part less confusing, we changed it from: *We complement the above analysis by upscaling the* 3.125 km *reference runs. For example, a* 25*x*25 km *grid cell contains a patch of* 64 3.125*x*3.125 km *grid cells. The scatter plot (e.g., Fig. 10) of the warm-based fraction (basal temperature with respect to the pressure melting point at* $0\,^\circ C$*) and the mean basal temperature with respect to the pressure melting point of the patch can be used to estimate the parameters* T_{ramp} *and* T_{exp} *of the basal temperature ramp (Eq. (8)). However, this does not account for the connectivity between the faces of, e.g., a* 25 km *grid cell. Without a continuous warm-based channel from one grid cell interface to another, there should be effectively no basal sliding across the grid cell, even when the average basal temperature is close to the pressure melting point. Consequently, this estimate for the basal temperature ramp should be a lower bound to the points in the scatter plot. Furthermore, the upscaling results depend on the bed properties (soft sediment vs. hard bedrock) and the specific scenario (surge vs. quiescent phase). As such, the upscaling statistics only consider grid cells within the pseudo-Hudson Strait area during surges. Due to the limited storage capacity for the* 10 yr *output fields, only the first* 10 kyr *after the first surge are used for the upscaling experiments.*

to: *A more physically-based approach to determining an appropriate scale-compensating temperature ramp stems from our motivation for research question Q5 above. We bundle all* 3.125*x*3.125 km *grid cells of our reference runs into patches of, e.g.,* 64 *grid cells. Each patch represents a coarser, e.g.,* 25*x*25 km *grid cell. We then determine the warm-based fraction (basal temperature at the pressure melting point) and the mean basal temperature with respect to the pressure melting point of each patch. We can then estimate the parameters* T_{ramp} *and* T_{exp} *of the basal temperature ramp (Eq. (8)) by plotting the warm-based fraction against the mean basal temperature for all patches (e.g., Fig. 10) and fitting a basal temperature ramp with the preliminary assumption that a corresponding coarse grid cell should have an ice streaming fraction proportional to the sub-grid warm-based area.*

However, this upscaling analysis does not account for the connectivity between the faces of, e.g., a 25 km *grid cell. Without a continuous warm-based channel from one grid cell interface to another, there should be effectively no basal sliding across the grid cell, even when the average basal temperature is close to the pressure melting point. Consequently, the best estimate for the two parameters of the basal temperature ramp should be a lower bound to the points in the scatter plot.*

Furthermore, the upscaling results depend on the bed properties (soft sediment vs. hard bedrock) and the specific scenario (surge vs. quiescent phase). Therefore, we only consider patches within the pseudo-Hudson Strait area during surges. Due to the limited storage capacity for the 10 yr *output fields, only the first* 10 kyr *after the first surge are used for the upscaling experiments.*

However, due to this clarification, we were unable to shorten this part.

[Ln 741: the correct way to describe these results is that the solution exhibits "convergence", but is not "converged"] We changed it from: *In general, both models show convergence, but the discrepancies between different horizontal grid resolutions are significant.*

to *In general, both models exhibit convergence under systematic horizontal grid refinement for the overall ice volume (mean bias, Table S19+S23 and S26), but the solution is not fully converged at the finest resolutions tested.*

[Figure 12: It would be clearer to plot the surge characteristics/metrics with respect to grid spacing. This is the typical plot made in convergence studies.] Agreed. We moved Figure 12 into the supplement and replaced it in the main text with a plot similar to Figures 6 and 7.

[Section 3.4.2: it is difficult to take anything away from this convergence study since half of the simulations did not finish. Would make more sense to just delete this section an add a sentence at end of previous section saying that PISM is too computationally intensive to conduct a robust convergence study.] We agree that the limited number of runs can somewhat skew the statistics, especially when comparing the results of the convergence study to those of other PISM experiments. However, the remaining runs still provide useful insight. Furthermore, the GSM convergence study (and all other experiments) is only based on one additional run (5 GSM vs. 4 PISM runs). Considering that runs crashed for some GSM experiments, excluding the PISM convergence study based on these grounds would also mean excluding some GSM experiments. Since this section is already quite short, we prefer to keep it its current form.

[Ln 788-794: I struggled with this paragraph because it is not clear that is possible to compare the magnitude of sensitivity to very different kinds of changes to parameters and other numerical choices. It would be better to just whether changes were greater or less than MNEEs. Isn't that the point of defining these?] Agreed. We removed this part from the revised draft.

[Ln 811-812: why do you expect this? Seems to be quite a big claim.] Upon rereading, our intended message (without checking the actual numerical sensitivity, one should not assume that MNEEs can be ignored) was lost. We have updated this paragraph from: *We expect other ice sheet models with a comparable experimental design and ice dynamics to show similar levels of MNEEs. To minimize the possibility of interpreting numerical errors as a physical response to a change in model setup, it is crucial to determine MNEEs (or a comparable metric).*

to: *Given the nonlinearities in the SSA (or higher approximation) ice sheet system, there is no a priori reason to confidently assume other ice sheet models will have ignorable MNEEs for unstable contexts such as surge cycling and grounding line response. Therefore, it is crucial to determine MNEEs (or a comparable metric) to minimize the possibility of interpreting numerical errors as a physical response to a change in model setup or forcing.*

[Ln 880-881: I'm not sure that this can be generalized since this hydrological model is quite simple and doesn't simulate many things other subglacial hydrology models do, like the increase in effective pressure with the development of channelized hydrology.] Thanks for bringing this up. While we have not tested different hydrology models for the GSM in this study, the simple hydrology model has been shown to be adequate for this context given the large parametric uncertainties in more complete basal hydrology models [Drew and Tarasov, 2022, under review]. We have added: *In general, this also holds for subglacial hydrology models with higher complexity [Drew and Tarasov, 2022, under review].*

[Ln 964: given the difference in configurations should we even expect to get temperature spokes in these simulations? fast flow is confined to a narrow strip which is different from the EISMINT simulations.] We agree that we do not necessarily expect temperature spokes in our simulations. This comparison was added by request of the first reviewer. Since it does not add key information to the understanding of this manuscript, we removed it from the revised draft.

References

M. Drew and L. Tarasov. Surging of a hudson strait scale ice stream: Subglacial hydrology matters but the process details don't. *The Cryosphere Discussions*, 2022:1–41, 2022. doi: 10.5194/tc-2022-226. URL https://tc.copernicus.org/preprints/tc-2022-226/.

PISM 2.0.6 documentation. Petsc options for pism users, August 2023. URL https://www.pism.io/docs/manual/practical-usage/petsc-options.html.

A. A. Robel, E. Degiuli, C. Schoof, and E. Tziperman. Dynamics of ice stream temporal variability: Modes, scales, and hysteresis. *Journal of Geophysical Research: Earth Surface*, 118(2):925–936, 2013. ISSN 21699011. doi: 10.1002/jgrf.20072.

Roiy Sayag and Eli Tziperman. Interaction and variability of ice streams under a triple-valued sliding law and non-Newtonian rheology. *Journal of Geophysical Research: Earth Surface*, 116(1), 2011. ISSN 21699011. doi: 10.1029/2010JF001839.

Ondřej Souček and Zdenek Martinec. ISMIP-HEINO experiment revisited: Effect of higher-order approximation and sensitivity study. *Journal of Glaciology*, 57(206):1158–1170, 2011. ISSN 00221430. doi: 10.3189/002214311798843278.