Rebuttal to the review by Johannes Feldmann

We thank the reviewer for their insightful and constructive comments on our manuscript. We'd hereby like to address their concerns, and propose revisions to our manuscript to alleviate them. Reviewer's comments are displayed in boldface, replies in regular type.

P4,Eq.(1): The meaning of the x and y indices and the bar above the variables is still not given in the text. Please add.

We will do so.

P6,L6-8: I suggested to remove "the square of" because otherwise from that sentenced one would deduce that Feldmann et al, 2014 and Leguy et al, 2021 also use the the square of the grounded fraction. Just wanted to point to this again but would leave it to the authors to change the statement or not.

We agree that the current phrasing is confusing regarding the way the grounded fractions are used in PISM and CISM; we will follow the reviewer's suggestion.

P17,L15: The authors announced but missed to include the reference Robin (1955).

We will include this reference.

P21,L9: investigating → investigation

We will change this.

I could not find the clarifications mentioned by the authors. If they are somewhere else, please indicate line numbers.

My point:

P4,Sec.2.2: The introduction mentions the advantages of the DIVA approach compared to the hybrid SIA/SSA approach and briefly mentions which stress terms the DIVA approach covers. Sec. 2.2, that includes the mathematical equations of the stress balance would be suited to refer to these stress terms. I suggest to name which of the shown

equations/terms correspond to which stress terms (SIA, SSA and additional stresses that are not captured by the SIA/SSA). That would give a lot more clarity on what the actual difference between DIVA and hybrid SIA/SSA is.

Author response:
We will add these clarifications to the text immediately after Eq. 1.

I could not find the clarifications mentioned by the authors. If they are somewhere else, please indicate line numbers.

My point:
16,L14: I would be interested in more details on the simplicity of the mentioned rheology, damage and subglacial hydrology. I recommend to discuss them here or to present details in the section 2.

Author response:
We will briefly mention the treatment of rheology and damage in section 2. The (lack of) treatment of subglacial hydrology is already included in the description of basal sliding; pore water pressure is calculated solely based on bedrock elevation, following Martin et al. (2011).

I was not able to find the announced additions to Sec. 2 regarding the ice rheology and damage. Please indicate line numbers on where the changes were made.

We apologise for these oversights. The changes we mentioned in the previous rebuttal were included in the manuscript, but apparently they were accidentally reverted at some point in the different versions of our document. We apologise for this oversight, and we will make sure to include them in the next revision.

Rebuttal to the review by an anonymous reviewer

We thank the reviewer for their insightful and constructive comments on our manuscript. We'd hereby like to address their concerns, and propose revisions to our manuscript to alleviate them. Reviewer's comments are displayed in boldface, replies in regular type.

We have grouped the comments about the ABUMIP experiment, the MISMIP+ experiment, and the model resolution together. Otherwise comments are addressed in order.

1 Comments about ABUMIP

I thank you for adding the control experiment to the ABUMIP set of plots, it is very informative. I am concerned you thought it would be good enough to show a control that is drifting without forcing by 1.2 m of sea level rise in 500 years. At the end of that time, both Ronne and Ross ice shelves are almost gone. This is worrisome. And the linear trend is not plateauing based on your figure.

…

Before this manuscript can be published: please perform the Antarctic spin-ups at the different resolutions you are presenting that are long enough to prevent your Antarctic ice sheet to drift so much during the control experiment. Doing so would give the reader confidence you can set up your model for paleo or future Antarctic simulations. In addition, please complete your analysis by showing a difference of your end of spin-up runs with your thickness target for your run at 40km and 16km resolutions (since you plan on using both resolutions for Antarctica for paleo and future runs respectfully). Redo the ABUM and ABUK simulations starting from these new spin-ups. Redo Figure 8 accordingly.

…

P17, l14: why did you choose a spin-up of only 500 yr? Clearly you have not reached a steady state. Please continue your spin-ups until you do so. (See main comments from earlier.)

…

P18, l5: Yes, your drift in ABUC is quite large for a steady state, especially at 40km. This alone indicates that your simulation is not ready. See earlier comments.

…

P18, l7: I disagree, this should not be improved in future work and should be investigated now. This paper is a benchmark for your future scientific studies, and for this reason it is reasonable to expect that you can perform and

show one example of an "acceptable" spin-up. (IMAU-ICE v1.0 could do so according to their ABUC results.) All along you are claiming that you can, and it is reasonable to perform simulations at 40km. Clearly, with this result, it isn't! Arguably, it is not as well at 20km resolution. After 500 year of control experiments, both the Ross and Ronne ice shelves are already almost gone!

Right now, you are showing that IMAU-ICE is not ready to perform sensible Antarctic simulations. And if you cannot produce a good initial steady state for your ABUMIP experiment then you should take out the ABUMIP experiment from this paper and replace it with something else and revise your text accordingly.

…

Now that I have witnessed your way of performing Antarctic spin-up, I would highly encourage you to show the result of a Greenland spin-up at 20km resolution, since you plan on using this resolution for your future studies. (I am simply asking about a spin-up here, not a transient run of any kind). Such a result would strengthen the proof of concept of your new numerical capabilities (I am thinking about DIVA here).

…

P4, l16: in this paper you have not shown any results of any of the continental ice sheets you plan on running with except one for Antarctica (which is quite unfortunate). Showing an initial thickness from ice sheets you experimented with could add great value to your manuscript. In addition, you specifically refer to future Greenland ice sheet (GrIS) simulations here which brings me to one of my general remarks (see above) that showing a steady state simulation of the GrIS would be a great addition to this manuscript.

We agree that, before our model can be used for actual future projections, we need to set up a better spin-up which has a lower model drift. The reason we chose to include the ABUMIP experiment, is that the ABUMIP experimental protocol does not require the model to be in a steady state at the start of the experiment; several of the other models in the ensemble by Sun et al. (2020) show a drift of well over half a meter of sea-level rise/fall in ABUC. Even with a model that perfectly resolves the stress balance, achieving an initial state that both matches the present-day observed geometry and also has no model drift, is very likely not possible without inverting for either basal roughness and/or basal melt (which of course implies the assumption that the present-day ice sheet is indeed in a steady-state – which is very debatable!). Since we only want to benchmark the ice-dynamical component of our model in this manuscript (verification, not validation), setting up such an initialisation is beyond the scope of this publication. We will reflect these thoughts in the manuscript.

As a compromise, we have improved the initialisation for our ABUMIP simulations. In addition to the mechanical relaxation (now shortened from 500 yr to 100 yr), we have added a 240-kyr thermal spin-up (two complete glacial cycles) to include the "thermal memory" of the glacial maximum in the viscosity, followed by a 100-kyr mechanical relaxation with fixed shelves, to allow the grounded ice to equilibrate. This means that our spun-up geometry now deviates further from the

present-day observations than before (particularly in East Antarctica, where the ice sheet tends to “flatten out”, thinning in the interior and thickening at the margins), but has negligible drift. Achieving a proper steady state in the ABUC experiment requires prescribing appropriate basal melt rates, which is admittedly difficult with the melt parameterisation described in the current manuscript. Since the basal melt formulation does not affect the ABUM and ABUK experiments, we opted instead to derive melt rates for ABUC using a simple geometry-based inversion (increasing melt rates when the shelf thickens, and vice versa; following Bernales et al., 2017). This results in a model drift of -0.1 to 0.1 m.s.l.e. after 500 years in ABUC. With this new spin-up procedure, sea-level rise in ABUM and ABUK is now smaller than in the previous set of results: 7.5 – 7.8 m for ABUM, and 7.6 – 7.9 m for ABUK. Note that, as before, the resolution dependence is still very small.
We will adapt the manuscript to reflect these changes.

One small side note: the Filchner-Ronne and Ross ice shelves did not collapse in our previous ABUC results, but they were not visible in the figure due to an unfortunate choice of colormap. The lowest bin of elevation values was shown in white, which was not visible on the white background – which escaped our notice. Thank you for pointing this out, we will fix this.

2 Comments about MISMIP+

In this manuscript, you are showing some numerical capabilities about your model which is one side of the story. The other side of the story is you try to convince the reader that your model is suitable for continental scale simulations as well. (You do show the different domains of possible computation in Fig.1 after all.) So far, for this latter point, you are convincing me that your model is not ready for continental scale simulations, specifically with your ABUMIP experiments. In our first round of comment, me and another reviewer did recommend you adding some MISMIP+ experiments to show the robustness of your choice of resolution and melt parameterization especially for marine ice sheet configuration. You chose not to do so arguing that it is material for another publication. Doing so puts you at a risk that readers will not believe that your choice of model configuration for ABUMIP is suitable. It does so with me. You mentioned the length of your manuscript for not showing the MISMIP+ experiments here. I will say that this should never be the sole excuse for not adding a scientific result to a manuscript. The gmd journal is actually a good place where I can expect papers to be longer because authors are trying to show development of their models and convince the reader of their scientific capabilities (just what you are trying to do).

…

I understand you have a paper underway (hopefully) about MISMIP+ experiments to complement this one. For the time being, please add a quick highlight of your MISMIP+ experiment results supporting your default of using the FCMP parameterization from Leguy et al. (2021). (Unless you want to add these experiments to this paper to show your results at 40km down to 10km).

We will add a paragraph about the MISMIP+ experiment. We will show the "default" simulations, which exactly follow the protocol from Asay-Davis et al. (2016). Our results for these experiments agree well with the ensemble results of Cornford et al. (2020), lying close to the ensemble mean, and well within the ensemble range. The upcoming paper about the more detailed investigation of this experiments includes simulations with a different resolution, stress balance, sliding law, sub-grid melt scheme, and basal melt parameterisation. This new paper is almost finished, and we expect to submit this within a few weeks.

3 Comments about model resolution

I feel a bit uneasy with a current tendency of ice sheet modeling papers stating it is fine to use coarse resolution (here 40km) and allow the model to have greater error because it will be used for paleoclimate simulations. I respectfully disagree with this mindset, and the computational expense of long paleoclimate simulations should not be an excuse to (gravely) misrepresent ice sheet behavior. In this paper you also mention that you plan on using the model for future projection study as well (using 16km resolution) meaning this "benchmark" paper should also be used to convince the reader it is acceptable to do so with this model, which it isn't.

In my first round of review, I did request for the sections presenting the experiments to be a bit more quantitative in their discussions; sometimes adding tables is also a good idea. I appreciate an effort was made in adding convergence figures (e.g., figure 7b) and I would like to see more discussion on the relative errors whether with respect to the analytical solutions or with the highest resolution used in these experiments. Simply stating something like "look, we do see convergence with resolution" is over simplistic. In Sec3.2, Fig.3, it is not only difficult to get a sense of the error with respect to the analytical solution (the y axis is not precise enough) but also you state (in your discussion) that your error at 40km is a reliable result for paleoclimate studies. Two arguments here: 1. What is an acceptable/reliable error? You never define this concept (and it is hard to do actually). 2. You seem to argue that an error of 350m/y with respect to the analytical solution is acceptable. I would beg to differ! And this is where a more detailed discussion either here or in the discussion needs to take place explaining why this magnitude of error is good enough for paleoclimate simulations. Based on the results on this section, I would feel way more comfortable if you chose to use a resolution of 20km.

…

P17, l6: This argument alone is insufficient to justify an acceptable result. (If I run an ice sheet model with a resolution of 200km and see an error of 195km, does it make it acceptable to use such a resolution?) Also, this part of your sentence is a repetition of what you already said 2 sentences prior. At 40km, the hysteresis corresponds to about 25% of the grounding line displacement. Here you are benchmarking your model meaning you will refer to it to explain results of

future science experiments. I would advise to acknowledge this large error, and in the discussion, give an example of a situation for coupled climate models for which this error could be of small importance.

Deeming an error to be acceptable is a tricky business and at the end modelers will run with whatever they feel comfortable with to justify their science.

What I can see in these experiments is that, again, running with a resolution of 20km leads to well improved results compared to 40 with again a superlinear convergence. Any idea what is happening with your 16km resolution results? This might be of importance since you plan on using this resolution to run continental scale future scenarios. Based on this experiment, it seems that 20km is better suited.

…

P19, l32: I strongly disagree with this statement, see previous comments. Define "reliable results"? This section is perhaps a good place to spend time on arguing why you think this resolution is adequate and cite references to back up your statements. Just saying it does not make it true.

It is important to keep in mind that palaeo-ice-sheet simulations generally are interested in large-scale ice-sheet evolution. When studying glacial cycles, we don't want to know at exactly what rate the grounding line of one particular outlet glacier is retreating during a century; we want to know how many tens of metres of sea-level change we can expect after ten thousand years. This is not just because that is the kind of quantity we can actually compare to proxy data, but also because the uncertainties in the forcing (palaeoclimate, paleogeography) are so large. Comparing modelled small-scale ice-sheet features to data becomes meaningless when the large-scale forcing is so uncertain, and so we generally tolerate such small features becoming obscured by a low resolution. In all of our experiments that concern large-scale dynamic evolution (Halfar dome, Bueler dome, MISMIP, ABUMIP), our model produces robust results even at 40 km resolution. The only experiment where we do indeed see some resolution dependence is the SSA ice-stream experiment mentioned by the reviewer, which does not concern dynamic evolution, but only instantaneous velocities. The most realistic and representative experiment is ABUMIP, and there we see no appreciable difference between the different resolutions (not in the previous set of results, and not in the new). This gives us confidence that, even at 40 km resolution, our model can simulate large-scale ice-sheet geometry and (rates of) sea-level change accurately enough that the model errors are negligibly small compared to the forcing errors typical of palaeo simulations. We will clarify these thoughts in the manuscript.

Regarding the reviewer's statement about computational expense: the harsh reality is that computational resources are limited, no matter how much we'd like them not to be. Every doubling of the resolution increases the computation time by about a factor 10. For a few one-off experiments that could be achievable, but palaeo-modelling usually involves sensitivity analysis based on ensemble simulations. We therefore partly agree with the reviewer that we should aim to always include a few higher-resolution runs in our ensemble studies, but it is not something we can feasibly do for every experiment (especially since the importance of accounting for the long paleo histories of the Greenland and Antarctic ice sheets in projections of their future retreat is becoming increasingly clear, e.g. Yang et al., 2022: Impact of paleoclimate on present and future evolution of the Greenland Ice Sheet, PLoS ONE 17, https://doi.org/10.1371/journal. pone.0259816). We do not believe that means we should not do any palaeo modelling at all until we come up with a faster model or a bigger computer; it just means that we have to accept that our results are not as accurate as they might at some future time become. We will reflect these thoughts in the manuscript.

4 Other comments

P7, l5: I suggest rewriting "The way the stress balance is discretize" by "The stress balance discretization".

We will do so.

P7, eq10 and 12: my previous comment on these equations might have been misunderstood. I appreciate you adding the bounds 0<=w_b<=1 and 0<=\lambda_w<=1 within the equations for clarity but it is still confusing. What I would like to see specifically written in the text is:

1. that for equation 10, b is bounded between b_min and b_max and you achieve this by writing something like:

w_b = ! ! max(0, *b* − *b*%&',), b < 0 "!"##"_%&' min(*b*%() − *b*%&', b − *b*%&',), b > 0

2. that d_w is bounded between 0 and 1000 so that the bounds for d_w are satisfied. Right now, you write d_w = z_SL -b. So if z_SL=0 (say see level reference set to today's value) and b=- 1100, then d_w=1100 and lambda_w=1.1. So clearly, you are taking the minimum between 1 and 1.1 here. Conversely, if b is above sea level, d_w<0 and in this case, you are taking the maximum between d_w and 0. So please, for clarity write d_w using something like:

d_w=!max(*b*%(),z*+ −b),b>0 min(−*b*%&',z*+ −b),b<0

3. After defining your equations, in the text (and you deleted it), mention that both w_b and d_w are capped between 0 and 1.

The scaling coefficients w_b and lambda_w are limited between 0 and 1 after they are calculated, e.g. if d_w = 1100, then lambda_w = 1. We will clarify this in the text and adjust the equations accordingly.

P10, Table 3: None of the parameters listed in this table are defined in the text. Please do so either in the table and/or in the text.

We will add a "description" column to Table 3, similar to Table 2.

P10, l13: you deleted the sentence stating the possible melt parameterization options available in IMAU-ICE v2.0. I believe it is good to keep them in the text and stating the default configuration for your version of the model. In the previous version you did mention these options but did not discuss the effect of them on your model output. If you do have results comparing them in the context of Antarctic simulation, please, add this discussion!

We will restore the description of the optional PMP and NMP sub-grid melt schemes, and mention the effects these have on the results of the MISMIP+ and ABUMIP experiments in some preliminary tests.

P12, l3: Can you be a bit more quantitative? Also, the convergence with resolution seems to be superlinear (at least at the center of the domain) and there is a net gain in using a resolution of 20km as opposed to 40km in these simulations. Why do you think the error at 40km resolution is acceptable? (See previous comments.) (I believe the new figure 3 is more informative compared to the older one; we learn something more about the model itself. Thank you for redoing it.)

We will add a panel to the figure showing the convergence, similar to the ones already included for the Halfar & Bueler domes and the MISMIP experiments. The order of convergence is about 1.7, which is indeed more than linear, and close to the value of 1.9 reported by Bueler and Brown (2009).

As mentioned before, we do not believe this experiment is very informative about the performance of the model in simulating the dynamic evolution of an ice sheet, as it only concerns instantaneous velocities in a confined setting. Our conclusion that 40 km is satisfactory is supported by the other experiments we present.

P12, Table 4: please, align your table headers (Parameter, description,...) on one line.

We will fix this.

P16, l8: Please rephrase this sentence. The ice flow factor is decreased (increased) as a step function after which you run your model forward in time for 15 kyr (not 25 kyr based on Fig. 7A) to a steady state. Please, add the values for the ice flow factors in your text.

We will do so.

P18, Fig.8: The caption indicates that the results from IMAU-ICE v1.0 are shown by dashed black line. I can only see a plain line. Also, which resolution are you displaying on the left? (And thank you for redoing this figure, it is much easier to read and more informative compared to the previous version.)

The results om IMAU-ICE v1.0 are shown by a thick grey line; we will correct the caption. The panels on the left show the 10 km results; we will mention this in the caption.

P19, l4: Why should this be as expected? I don't think so. Here you are comparing your results with ELMER/ice for the ABUM experiment. In the next sentence you argue that the difference between ABUM and ABUK should be small for a given model within 3-4 years of transient simulation (I'll get back to this later). The ELMER/ice result for ABUK is almost 8m lower than the one of ABUM after 500 years. Why do you think there ABUM results is correct and their ABUK result is wrong? I would argue the opposite!
I am not saying your results are wrong per se, I am saying that your argument for comparison is not sound. Your model configurations are different compared to all the models in Sun et al. (2020). And this is where your melt parameterization analysis could help you out.

We did not mean to argue that our results were correct because they were similar to ELMER; we will remove the reference to ELMER in the text. What we do mean to argue is that, purely from the physics, we do not expect to see much of a difference between the ABUK and ABUM experiments (although we do acknowledge that there is some room for discussion; we will phrase it less strongly in the manuscript). The fact that some other models in the Sun et al. ensemble, including ELMER, show substantial differences is certainly interesting, but explaining why this is the case is beyond the scope of our work. The fact that our own model does not show much of a difference is, in our view, supportive of our model's performance.

P19, l27: Some of your results are close to analytical solutions but they do not match! Also, you only compared to ELMER/ice and otherwise your ABUMIP results are outliers compared to other ice sheet models. Please rephrase

We will find another phrase to indicate that our results closely approximate the analytical solutions.

P20, l6: replace "With a minimum of effort" by "With minimum effort".
We will do so.

P20, l7: a space is missing after ")"
There is not; parentheses around an optional prefix should not have a space between the closing parenthesis and the word being prefixed.