Review of Constantijn et al. 2021, GMD (round 2)

I would like to thank the authors for taking the time to respond to my and other reviewers' comments. I find that the manuscript has improved especially with the presentation of the experiments with the suggested resolutions.

Unfortunately, despite the additional information added to the manuscript, I cannot recommend this manuscript for publication at this time. Rather I would say that if my major concern is not addressed, I would have to reject it (I'll explain a bit more below). Maybe my first round of review was not clear enough, and I apologize about that.

I will start here with a few general points before stating my main concern and ending with other comments.

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.

My major concern:
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 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.

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).

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).

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.

Other comments

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.

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

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:

$$w\_b = \frac{1}{b_{max} - b\_min} \begin{cases} \max(0, b - b_{min},), \mathrm{b} < 0 \\ \min(b_{max} - b_{min}, \mathrm{b} - b_{min},), \mathrm{b} > 0 \end{cases}$$

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:

$$d\_w = \begin{cases} \max(b_{max}, \mathrm{z_{SL}} - \mathrm{b}), \mathrm{b} > 0 \\ \min(-b_{min}, \mathrm{z_{SL}} - \mathrm{b}), \mathrm{b} < 0 \end{cases}$$

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.

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.

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!

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.)

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

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.

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.

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, 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.)

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!

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.

P19, l6-l10: Do you have any evidence to present (experiments, calculations, …) to argue a lag of 3-4 years at the most between ABUM and ABUK? If not, I would stay away from it. The one thing you know is that the sea level rise results for ABUK should be greater than the one from ABUM.

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.

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.

P20, l6: replace "With a minimum of effort" by "With minimum effort".

P20, l7: a space is missing after ")".