

**Response to R1 (second revision) on:**

***‘Antarctic ice sheet model comparison with uncurated geological constraints shows that higher spatial resolution improves deglacial reconstructions’***

---

**Below we respond to reviewer suggestions. We thank the reviewer again for their time and insights towards improving the clarity and utility of our work.**

I thank the authors for carefully addressing my comments on the manuscript. These revisions have made the paper much easier to read, which I think most readers will appreciate for a paper of this length. I have a few further comments and suggestions, but I consider the paper pretty much ready for publication.

All line numbers refer to the version with tracked changes.

L 213: Are ocean temperatures still taken from a single ocean depth of around 300–400m, as was previously the case with this model?

**Yes, specified.**

L236: As in my previous review, this statement is incorrect, or is at least ambiguous. DeConto et al. (2021) demonstrated convergence with respect to resolution below about 10km for a nested domain of Thwaites Glacier.

**This sentence has been added verbatim.**

Section 2.4: Are you using a full-factorial ensemble design? It could be useful to have a small table with the parameters and their values.

**No; we select only a subset of the full parameter space in order to simulate the widest possible variation in ice sheet behavior; our goal is to vary key parameters to produce a small but representative ensemble of model simulations in order to explore the application of our new methodology. Thus the focus on the specific parameters and their values is of less importance to our analysis, though we describe each parameter fully in ‘Section 2.4 Parameter variation’.**

Section 4: In the author’s response to reviewers, there is an explanation that really helped clear up my confusion during the first round of review: “We don’t combine these metrics together into one total score, because our metrics are aimed towards answering different questions about model data fit. The ‘optimal’ weights for combining metrics into one total score would differ based on the user’s interest.” I might just be missing it, but I don’t see a statement like this in the revised text. I think including something like this in the first few paragraphs of section 4 would

help readers understand the rest of that section.

**Added to the first paragraph in section 4.**

L560: sentence needs revising: “such as snow buildup patterns such as wind scoops”

**Adjusted.**

L680: Model resolution could lead to systematic offsets as well; for example, coarse resolution models are often more susceptible to MISI-style retreat than high-resolution models. Conversely, forcing uncertainty can be random rather than systematic.

**Agreed, we further toned down the sentence accordingly: “*On the other hand, if model-data offsets are randomly (rather than systematically) distributed among sites in close proximity, this could motivate a reassessment of potential issues with either model resolution or the interpretation of geologic data.*”**

Fig 12b: Does this include just the 40 km runs?

**Yes; now specified.**

**Response to R2 (second revision) on:**

***‘Antarctic ice sheet model comparison with uncurated geological constraints shows that higher spatial resolution improves deglacial reconstructions’***

---

**Below we respond to reviewer suggestions in detail; for typographical errors or small corrections, the \* symbol indicates that the correction was made as suggested in the revised manuscript.**

I have gone through the revised manuscript and must say it is a massive improvement from the first version in all manners: it is better structured, flows much better while reading, and the key concepts and metrics are much better described. I particularly liked the idea of adding a table that describes the different metrics, even if there's room for improvement (see suggestions below). In general, I find this work a significant step forwards in term of data-model comparisons at paleo scales in ice sheet modelling, and only offer relatively minor suggestions, mostly editorial and some others to clarify key missing points.

Once these are taken care of, I am happy to recommend this manuscript for publication in The Cryosphere. I look forward to seeing it published!

**We thank the reviewer again for their time and insights towards improving the clarity and utility of our work.**

L30: "Because the time period [...] is too short", as opposed to "are". \*

L156: I still find it confusing that this is called a "secondary set of cosmogenic nuclide model constraints (on the maximum ice thickness during the last glacial cycle)", considering it is essentially the same Ice-D dataset. It would be much clearer if the two datasets were explicitly split into something like "thinning constraints" and "maximum thickness constraints" (instead of calling one of them 'secondary'), stating that both were based on the Ice-D dataset. I know it is largely a naming issue, and perhaps a pet peeve, but I found it much easier to understand what was being done once I decided to call them like that myself.

**Reworded: “In addition to the exposure-age dataset of thinning constraints, we also compile an exposure-age dataset of maximum-thickness constraints, comprised of sites where exposure age measurements bracket the local last-glacial-cycle ice thickness change. This secondary dataset also leverages cosmogenic nuclide model measurements but applies them to constrain the maximum ice thickness achieved during the last glacial cycle.”**

L203-206: It is not yet clear to me what "linearly downscaling the continental simulation across the preceding 2,000 years" means. Do you average the model results between 32 and 30ka, and this is what you (bi?)linearly interpolate to the regional domain? This needs to be more clearly

described.

**Reworded: “At the beginning of each high-resolution nested simulation, initial conditions are generated by downscaling the continental simulation boundary conditions during a 32-30ka relaxation period, allowing the model to adjust to the high-resolution domain, before branching off the 30ka-0 ensemble experiments.”**

L244-247: I understand that it is nice and important to acknowledge previous studies that tried something similar, but I feel like this was already done in the introduction (for this specific study and others), and the justification for the choice of parameters, parameter space, and number of simulations is already given at the start of this section. This paragraph feels a bit redundant and unnecessary, so I'd be tempted to remove it. \*

L273: "this vertical window is  $\pm 47\text{m}$ ". \*

Table 1, "Role in the 'float scoring approach'" for M\_model: is it supposed to be "The 'model misfit' model score is computed by summing all 'float misfit' site scores"? This table is really helpful, but still needs more clear descriptions. Some variable names are the same for different variables, which adds to the confusion. Is there a way to improve the variable naming so each reference to another metric can be clearly related to its respective cell? For example, 'float misfit' and 'exceedance' are mentioned in the description cells, but never appear as variable names.

**I think we have now resolved this confusion by consistently calling all analyses at a site “site misfits” (e.g., site float misfit, site best-time-offset misfit, site exceedance misfit), and calling all continent- or region-wide model analyses (summing across sites) “misfit scores” (e.g., model float score, model best-time-offset float score, model exceedance score). So, we only use the word ‘misfit’ in the context of a sample or site misfit statistic; and we only use the word ‘score’ in the context of a model-wide misfit statistic. This is adjusted in Table 1 and throughout text.**

L485: There's an extra '('. \*

Figure 7: Panel (b) seems to be reversed, as the figure caption and panel (a) treat the applied negative offset as making the modelled deglaciation happen later (and thus shifted towards the left). The caption states "-2 kyr" as the offset, whereas in the figure it is +2 kyr. It also refers to a red dot, whereas the dot is (I believe) plotted in blue.

**The plot in panel b was plotted incorrectly; this has been corrected.**

L647: The mention of Fig. 16 comes before many other figures (I believe before Fig. 10). I understand that it is convenient to have it as such, but please double check if that is in accordance with editorial guidelines, or if that figure needs to be put further up/moved to the

supplement/appendix.

**I defer to the editor and journal copy-editors here.**

L677: correct C14 to 14C at the end of the line. \*

L715: Is H\_model computed as max(H\_model) because it is the maximum thickness change over the simulation time? If that is correct, please make it explicit. **Yes, corrected.**

Fig 12b Please change the x-axis direction so it is in the same order as the other figures, for consistency. \*

L811-813: I still do not agree with "A numerical simulation where the modeled timing of thinning has to be shifted by a different amount at each site to minimize model/data mismatch is a poor representation of the geologic record;". It is not reasonable to believe that offsets in time should be the same across the entire ice sheet, as the climate evolution is not homogenous or synchronous over the entire continent. I do, however, agree with the follow-up argument (L814-816), where it is explicitly stated "A simulation where model/data mismatch at each site can be minimized by shifting the modeled timing of thinning by the same value everywhere \*\*within a catchment\*\* is a good representation...". I think the same reservation (i.e., "within a catchment") applies to the former, and would suggest the authors adding it. **Agree; added!**

Fig 13a,b: As in Fig. 7b, please double check the offset signals (and axes directions) so all figures are consistent with each other.

**This is correct (the inconsistency was the plotting in Fig. 7b).**

L875-877: which lingering challenges? Please state them (or at least the main ones), otherwise this sentence feels too generic. \*

L885: There's a typo in the in-text citation style. \*

L885-886: This statement is fairly unspecific and general: "a symptom of systematic difficulties with model reconstructions of deglacial ice sheet behavior". This point is much better addressed in L996-1004, so I'd suggest either referencing it (as done before in the manuscript for other concepts/arguments), or expanding on this statement. \*

L900: I just noticed that "behaviour" here has the British spelling, whereas it has the American spelling throughout the rest of the manuscript. According to the TC guidelines, both styles are allowed but they need to be consistent throughout the manuscript. \*

L924-926: I find it hard to take this statement at face value considering that the x and y axis ranges in Fig. 14 are different in all panels. In absolute values, the exceedance scores are much better for EAIS and TAM, whereas the float score varies quite a lot in all plots. The distribution of all points in the plot would change if the axes limits were kept the same for all of them, so I wonder if it would make sense to normalise the scores across regions? Is there any better way to compare the data than how is presented here?

**After exploring various options for presenting these data, we eventually chose to vary the x and y axes across subplot in order to best show the tradeoffs between the two metrics; normalizing the scores was a second option, but we eventually decided against this route because the magnitude of both exceedance scores and float scores we felt was significant. For the exceedance scores, a model ice thickness exceedance of, say, 18km (Fig. 14b) is much more drastic (and thus important to reflect) compared to, say, 0.45km (Fig. 14d), which is not so compelling to me.**

**The statement in L924-926 therefore concerns just the *relative tradeoff* between metrics (“*the timing vs. thickness tradeoff... is prominent in East Antarctica (i.e., the Transantarctic Mountain region...)*”) rather than contrast the magnitudes or ranges in magnitudes, due to the differing axes shown in Fig. 14.**

Fig. 15: Please double check whether the sign of the offset agrees with the intended sign after checking Figs. 7 and 13.

**The sign of the offset is correct here (the inconsistency was the plotting in Fig. 7b).**

L974: Is it worth adding that despite persisting for longer, they do match the thinning curve better (i.e., lower float scores)? \*

Fig. 16g: Please add to the caption what is being plotted in greyscale. Is that ice thickness or basal topography? If so, at what time? \*