Response to the review of Tijn Berends (Referee #3)

Dear Referee,
Thank you for reviewing our manuscript and for your effort and time spent on this. Your constructive feedback will help to improve the final revised version of this manuscript.

Based on this feedback, we will substantially revise the manuscript with the following main changes:

- we will reduce the complexity of the used model to rely on fewer parameters
- The benthic δ^{18}O curve (probstack with removed trend) from Clark et al. (2025) will be used as a new main target. Berends and Rohling SL curves will only be shown in SI
- ORB, ABR and GRAD models will be removed to focus on the novel RAMP model. This will make the manuscript more concise and highlight the novelities compared to the Legrain et al. (2023) model

In the following, we respond to your comments and propose several changes to the manuscript, motivated by your suggestions.

Author comments
Referee comments

Major comments

My main issue with this manuscript is that it is a continuation of the work by Legrain et al. (2023), but it is difficult to understand from the manuscript how the new model and its results differ from those by Legrain et al. Right now, a reader who is not familiar with Legrain et al. could be given the false impression that the model and the experiments are entirely novel. (In fact, even for me it is difficult to spot the differences, and I actually reviewed the Legrain et al. paper!). To prevent this, when presenting the model equations, it should be made clear how and why they are different from those already published. Similarly, since three of the four experiments (ORB, ABR, and GRAD) were also performed by Legrain et al., the results of those experiments with the earlier model version should be shown for comparison, to illustrate the effect of the changes to the model.

We agree. One major improvement compared to the Legrain et al. model is of technical nature regarding the improved tuning strategy and the improved speed of the code. However, this was barely mentioned in the manuscript, since we wanted to focus on the model output rather than on these technical details. The model speed was significantly increased by translating the code into JAX (a python library which allows for accelerated computation and just-in-time compilation). Furthermore, we added three additional solver, namely ptemcee (parallel-tempered MCMC), dynesty (dynamic nested sampling) and PyMC. With the improved tuning procedure and with the increase in speed, which allows a more rigorous

tuning, we could already reduce the model data mismatch compared to the original Legrain et al. model, without any changes in the parameterisations. We propose to add a paragraph in the model section of the revised version of the manuscript, where we highlight the increase in speed and the newly added solver to the model.
A reason for repeating the ORB, ABR and GRAD models was to show that despite changes in the model, the results obtained in the Legrain paper remained the same, i.e. the ORB model failed in reconstructing the MPT, while the GRAD model yielded better results than the ABR model. However, since the new RAMP model can converge to a constant v_0 parameter, an abrupt-like change, a gradual trend over the entire Quaternary, or something in between, it already tests the ORB, ABR and GRAD scenarios and yields the most likely forcing scenario. Therefore, we agree that there is no longer a need to include the ORB, ABR and GRAD models. They will be removed in the revised manuscript and the focus will be solely on the RAMP model. This will make the revised manuscript more concise and better highlight the novelities compared to the Legrain et al. (2023) paper.

My second issue regards the way the performance of the different model versions is assessed, namely by comparing to the sea-level curve from my 2021 paper (Berends et al., 2021a). I hesitate to call it a reconstruction (despite the rather optimistic title of that publication!), because, as I cautioned readers in that paper, that curve is based on a model-based deconvolution of the LR04 benthic d18O stack, and includes major uncertainties. For example, at the last glacial maximum, I simulated a sea-level drop of only 100 m, well below the accepted value of 130 m. As I stated in the discussion section of that paper, the results presented there "...should not be interpreted as a realistic reconstruction of what the world looked like in terms of global climate, ice-sheet geometry, sea level, and CO2 during these periods of geological history. Rather, ... they should be viewed as scenarios which can help in interpreting an expected new ice-core record." Considering the magnitude of the errors and uncertainties in this curve (i.e. up to 30 m), I do not think it can be used to (in)validate the results of the conceptual models presented here, which all lie well within this uncertainty (i.e. RMSE's of less than 20 m). Given that other sea-level curves (like the different ones from Rohling et al., which are also cited in the manuscript) suffer from similar uncertainties, in my view this means that the currently available sea-level data is insufficient to choose one of the author's models over another. This implies that the conclusion of the manuscript should be that, "until more accurate observations become available, both abrupt and gradual changes in the Earth system can explain the available (inaccurate) observations equally well". Of course, that does not mean it is not worthwhile to study the implications of abrupt vs. gradual changes, or of introducing non-linearities such as the ice-volume feedback in the RAMP-l model. But at present, such studies are, by necessity, exploratory in nature.

We agree. We will no longer use the Berends or Rohling sea level curves as our main tuning targets in the revised manuscript, but only show them in the SI. To discuss the limitations of these model-based sea level reconstructions, we propose to add the following paragraph to our discussion:

"Available global mean sea level data exhibit large uncertainties. Between 50 and 30 ka, geological and geochemical reconstructions of the GMSL vary up to 60 m and above (Farmer et al., 2023). Model-based deconvolutions of global δ^{18}O into GMSL, like for the Berends et al. (2021) and Rohling et al. (2022) sea level reconstructions, exhibit similar large uncertainties. While the Berends curve shows an LGM lowstand of around 100 m, the Rohling curve gives around 108 m. Both values are well below observational-based reconstructions of around 130 - 135 m for the LGM (Austermann et al., 2013; Lambeck et al., 2014; Yokoyama et al., 2000). These large uncertainties limit the usability of sea level reconstructions as targets for conceptual models, as presented in this study."

Instead of using the Berends sea level curve, we will follow the advice from Referee 1 and use the benthic δ^{18}O record (probstack with trend removed) from Clark et al. (2025) as our main target. Furthermore, we will discuss the model performance on the newly available seawater $\delta^{18}O_{sw}$ record from Clark et al. (2025), based on a benthic δ^{18}O deconvolution motivated by an ocean temperature data compilation. It gives a contrasting view compared to the Berends and Rohling deconvolutions by exhibiting a rise in $\delta^{18}O_{sw}$ during the MPT, interrupting a previous period of increasing values and showing larger glacial interglacial amplitudes in the pre-MPT cycles. However, a final reconstruction of this $\delta^{18}O_{sw}$ record into sea level is not yet available. That's why we will mainly focus on the benthic record. Following the advice of Referee 2, we reduce the complexity of the RAMP model. The results of this reduced RAMP model for the three different tuning targets (Berends SL, Clark $\delta^{18}O_b$, Clark $\delta^{18}O_{sw}$) can be seen in Fig. 1. Our reduced RAMP model yields similar results for this target, i.e. a long-lasting trend for the ramp, which started over 2 Ma. While the model performs well on this target, it has more difficulties with the seawater δ^{18}O record from Clark et al. (2025) (Fig. 1c), as can be seen by comparing the R values. In general, the model cannot accurately reproduce the larger pre-MPT amplitudes of the glacial-interglacial cycles and the decreasing trend of $\delta^{18}O_{sw}$ during the MPT. The RAMP formulation seems to be less suited for this target, as can bee seen by the identified ramp period, which lies now in the interval $\sim$ 400 - 200 ka. For the $\delta^{18}O_{sw}$ target, the model also cannot reproduce the shift towards the 100-kyr periodicity correctly, since it only shows a 100-kyr signal for the last 200 ka.

The authors also apply their different model versions to the future. I do not immediately see the relevance of such experiments, given that present-day CO2 concentration far exceeds the Pleistocene range. The results presented here should, as far as I understand, be interpreted as "possible future patterns of glaciation in the absence of any anthropogenic CO2 emissions" – i.e., a thought experiment. Since such a thought experiment lies quite far from the scope of the rest of the manuscript, I think it might be best to leave it out altogether.

While neglected in most other conceptual models, we believe that extrapolating at least the next glacial cycle is an important tool of model evaluation, since this represents the only time interval in which a model cannot be tuned or fitted onto some existing paleoclimatic target curve. Of course, the extrapolated curves do not resemble nature, as they lack anthropogenic CO_2 emissions. Therefore, as stated by the referee, they must be interpreted as purely academic thought experiments. More complex models, coupled to the carbon cycle, project, even without any anthropogenic influence, an unprecedentedly long interglacial, lasting for another 50 kyr (Ganopolski et al., 2016; Talento and Ganopolski, 2021). In contrast to this, all of our models, no matter of which simulation period, project that the next glacial cycle has just started or is about to start. In their recent study, investigating the

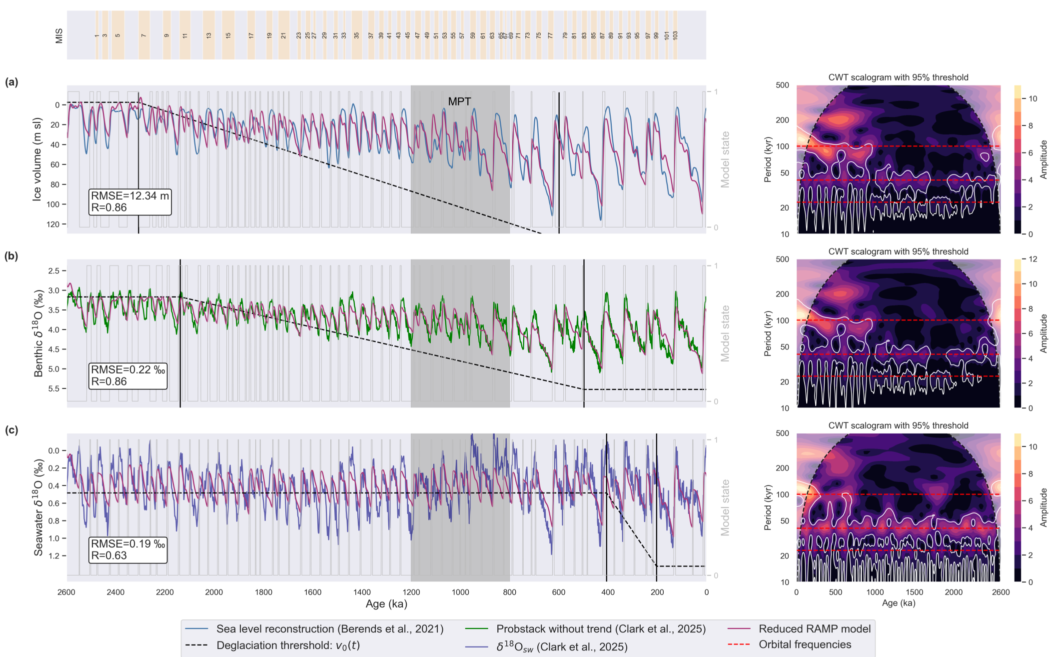

Figure 1: Reduced RAMP model tuned on three different tergets: a) Berends et al. (2021) sea level curve, b) Clark et al. (2025) benthic $\delta^{18}O_b$ (probstack with trend removed) and c) Clark et al. (2025) sea water $\delta^{18}O_{sw}$.

influence of precession and obliquity on glacial interglacial cycles, Barker et al. (2025) estimate that the current interglacial conditions would last for 11 kyr, when obliquity reaches its next minimum and the succeeding glacial would be interrupted in around 66 kyr (again neglecting anthropogenic effects). Other conceptual models similarly predict a contrasting view for the next glaciation: Calder (1974) projected a start 5 ka and an end in around 119 kyr (see his Fig. 2), Imbrie and Imbrie (1980) estimated a start 6 ka (see their Fig. 7) and Fig. 5 in Paillard (2015) reveals that only a large enough glaciation threshold can lead to a prolonged Holocene in the Paillard (1998) model, while a lower threshold would have terminated the Holocene already a few kyr ago. Therefore, we believe that the failure in simulating a prolonged Holocene in our model, as well as in other conceptual models or studies, is a striking feature to mention, which might be due to the missing carbon feedback in the model or a glaciation threshold which only depends on orbital forcing or parameters.

To account for the Referee's comment, we propose to add some parts of the above paragraph into the discussion and put more focus on the failure in simulating a prolonged Holocene in the model.

Lastly, I believe that, with thirty pages of body (plus ten pages of bibliography and appendices) the manuscript is too long, especially considering the rather simple model and limited number of experiments. There is a lot of repetition going on, with e.g. the nature of the four different experiments being described in the experimental set-up, in the results, and then again in the discussion. A lot of text could, and therefore should, be removed without losing information or coherence.

We agree. By removing the ORB, ABR and GRAD models, the revised version will be significantly shorter. Furthermore, we follow the advice from Referee 1 and shift the appendices into the SI. This will help to make the revised manuscript more concise.

Minor comments

Sect. 2.1: Please add a few lines explaining why it is physically justified to assign different weights to the different orbital signals (eccentricity, obliquity, and precession) in the different experiments. A flake of snow lying on top of an ice sheet cannot know the eccentricity of Earth's orbit or the tilt of its axis; it can only feel how much solar heat it is receiving at that moment (which is a simple mathematical result of those orbital parameters, with no room for "optimization"). Of course, it has to do with the asymmetry in the sensitivity of the mass balance to summer vs. winter changes; that fact deserves to be mentioned.

The main idea behind using a linear combination of these three orbital parameters instead of relying on a single metric, as has been done in many other models (e.g. Paillard (1998); Ganopolski (2024); Leloup and Paillard (2022)), is to reduce a potential source of model bias. We propose to include the following lines to the manuscript:

"Leloup and Paillard (2022) showed that the model outcome depends on the chosen insolation forcing (e.g. summer solstice, caloric season, ISI above some threshold at 65° N). This poses a bias to the model by selecting a specific insolation metric. Instead, we use a linear combination of precession, co-precession and obliquity, which can represent insolation at most latitudes and seasons (Imbrie et al., 2011). This approach has been used successfully in other models before (Imbrie et al., 2011; Legrain et al., 2023)."

Fig. 2: it looks like the ORB and ABR models produce higher-than-present sea levels between 2.6 and 2.2 Myr. Should I interpret this as retreat of the Antarctic ice sheet? How does your model handle negative values for the ice volume?

Negative values are possible in the model and handled normally. For extended runs longer than 2.6 Myr, they are needed to account for higher sea levels prior to the onset of Northern Hemisphere glaciations and are also visible in reconstructions by Rohling et al. (2022) and Berends et al. (2021). Due to the poor performance of the ORB model, there is no point in interpreting these negative results, since the model yields an unrealistic output. The ABR model does show some negative values around 2.4 Ma. Since the model has no spatial resolution, it does not allow for localising the melting location. However, it is most likely due to a combined melting of ice sheets from Antarctica, Greenland and glacier retreat compared to present-day conditions.

Fig. 3 and other places: I'd like to suggest again (as I did to Legrain et al. before) to use a wavelet transform (or something similar) to show the temporal changes in frequency content. This is much more intuitive and informative than showing time-averaged power spectra for two different blocks of time.

We agree. Wavelet scalograms, as in Fig. 1, will be used in the revised manuscript.

Table 2: Please explain why running the models over different lengths of time (2.0 Myr, 2.6 Myr, or 3.6 Myr) can yield such different results. Are the models also optimised separately?

Yes, the models were always optimised for the corresponding simulation period. The initial ice volume v_i is a tunable parameter and has to be adjusted for each simulation period separately. Furthermore, the deglaciation parameter v_0 denotes for the ABR, GRAD and RAMP models the initial value of $v_0(t)$. This parameter must be different for the considered time periods as well. If the same set of parameters is used from a shorter to a longer simulation period, the obtained ice volume reconstruction would tend to have a positive offset during the start of the simulation, since the global ice volume gradually increased over time (visible in the Berends and Rohling reconstructions). To avoid this and to test how consistent the model output is, we always retuned the models for the specific time period. However, if the model output significantly differs from one time period to another, this hints towards the existence of multiple local minima in the parameter space or model deficiencies. The ORB model differed for all three time periods, since the solver could not identify a single best set of parameters for all three time periods. In combination with the failure to reproduce the MPT in any of these simulations, this shows the model deficiencies of the ORB model. To a smaller extent, this can be seen for the ABR model as well. Since it relies on a single constant v_0 parameter prior to the MPT, it struggles in reproducing the shift from smaller global ice volumes (at 3.6 Ma or 2.6 Ma), compared to the larger ones right before the MPT.

Only the GRAD and RAMP modes yielded similar results for all three simulation periods, showing their consistency, hence being the more favourable models.
Additionally, the reported RMSEs in Table 2 cannot be compared for different time periods, since the magnitude of the RMSE depends on the target data. Between 3.6 Ma and 2.8 Ma, the Berends reconstruction shows very small values with low variation, leading to lower RMSEs compared to the 2 Myr simulations.

References

J. Austermann, J. X. Mitrovica, K. Latychev, and G. A. Milne. Barbados-based estimate of ice volume at Last Glacial Maximum affected by subducted plate. *Nature Geoscience*, 6(7):553–557, July 2013. ISSN 1752-0908. doi: 10.1038/ngeo1859. URL `https://www.nature.com/articles/ngeo1859`. Publisher: Nature Publishing Group.

S. Barker, L. E. Lisiecki, G. Knorr, S. Nuber, and P. C. Tzedakis. Distinct roles for precession, obliquity, and eccentricity in Pleistocene 100-kyr glacial cycles. *Science*, 387(6737):eadp3491, Feb. 2025. doi: 10.1126/science.adp3491. URL `https://www.science.org/doi/10.1126/science.adp3491`. Publisher: American Association for the Advancement of Science.

C. J. Berends, B. de Boer, and R. S. W. van de Wal. Reconstructing the evolution of ice sheets, sea level, and atmospheric CO_2 during the past 3.6 million years. *Climate of the Past*, 17(1):361–377, Feb. 2021. ISSN 1814-9324. doi: 10.5194/cp-17-361-2021. URL `https://cp.copernicus.org/articles/17/361/2021/`. Publisher: Copernicus GmbH.

N. Calder. Arithmetic of ice ages. *Nature*, 252(5480):216–218, Nov. 1974. ISSN 1476-4687. doi: 10.1038/252216a0. URL `https://www.nature.com/articles/252216a0`. Publisher: Nature Publishing Group.

P. U. Clark, J. D. Shakun, Y. Rosenthal, C. Zhu, P. J. Bartlein, J. M. Gregory, P. Köhler, Z. Liu, and D. P. Schrag. Mean ocean temperature change and decomposition of the benthic ^{18}O record over the past 4.5 million years. *Climate of the Past*, 21(6):973–1000, June 2025. ISSN 1814-9324. doi: 10.5194/cp-21-973-2025. URL `https://cp.copernicus.org/articles/21/973/2025/`. Publisher: Copernicus GmbH.

J. R. Farmer, T. Pico, O. M. Underwood, R. Cleveland Stout, J. Granger, T. M. Cronin, F. Fripiat, A. Martínez-García, G. H. Haug, and D. M. Sigman. The Bering Strait was flooded 10,000 years before the Last Glacial Maximum. *Proceedings of the National Academy of Sciences*, 120(1): e2206742119, Jan. 2023. doi: 10.1073/pnas.2206742119. URL `https://www.pnas.org/doi/abs/10.1073/pnas.2206742119`. Publisher: Proceedings of the National Academy of Sciences.

A. Ganopolski. Toward generalized Milankovitch theory (GMT). *Climate of the Past*, 20(1):151–185, Jan. 2024. ISSN 1814-9324. doi: 10.5194/cp-20-151-2024. URL `https://cp.copernicus.org/articles/20/151/2024/`. Publisher: Copernicus GmbH.

A. Ganopolski, R. Winkelmann, and H. J. Schellnhuber. Critical insolation–CO2 relation for diagnosing past and future glacial inception. *Nature*, 529(7585):200–203, Jan. 2016. ISSN 1476-4687. doi: 10.1038/nature16494. URL `https://www.nature.com/articles/nature16494`. Publisher: Nature Publishing Group.

J. Imbrie and J. Z. Imbrie. Modeling the Climatic Response to Orbital Variations. *Science*, 207 (4434):943–953, Feb. 1980. doi: 10.1126/science.207.4434.943. URL https://www.science.org/doi/10.1126/science.207.4434.943. Publisher: American Association for the Advancement of Science.

J. Z. Imbrie, A. Imbrie-Moore, and L. E. Lisiecki. A phase-space model for Pleistocene ice volume. *Earth and Planetary Science Letters*, 307(1-2):94–102, July 2011. ISSN 0012821X. doi: 10.1016/j.epsl.2011.04.018. URL http://arxiv.org/abs/1104.3610. arXiv:1104.3610 [astro-ph, physics:physics].

K. Lambeck, H. Rouby, A. Purcell, Y. Sun, and M. Sambridge. Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. *Proceedings of the National Academy of Sciences*, 111(43):15296–15303, Oct. 2014. doi: 10.1073/pnas.1411762111. URL https://www.pnas.org/doi/10.1073/pnas.1411762111. Publisher: Proceedings of the National Academy of Sciences.

E. Legrain, F. Parrenin, and E. Capron. A gradual change is more likely to have caused the Mid-Pleistocene Transition than an abrupt event. *Communications Earth & Environment*, 4(1):1–10, Mar. 2023. ISSN 2662-4435. doi: 10.1038/s43247-023-00754-0. URL https://www.nature.com/articles/s43247-023-00754-0. Number: 1 Publisher: Nature Publishing Group.

G. Leloup and D. Paillard. Influence of the choice of insolation forcing on the results of a conceptual glacial cycle model. *Climate of the Past*, 18(3):547–558, Mar. 2022. ISSN 1814-9324. doi: 10.5194/cp-18-547-2022. URL https://cp.copernicus.org/articles/18/547/2022/. Publisher: Copernicus GmbH.

D. Paillard. The timing of Pleistocene glaciations from a simple multiple-state climate model. *Nature*, 391(6665):378–381, Jan. 1998. ISSN 1476-4687. doi: 10.1038/34891. URL https://www.nature.com/articles/34891. Publisher: Nature Publishing Group.

D. Paillard. Quaternary glaciations: from observations to theories. *Quaternary Science Reviews*, 107:11–24, Jan. 2015. ISSN 0277-3791. doi: 10.1016/j.quascirev.2014.10.002. URL https://www.sciencedirect.com/science/article/pii/S0277379114003898.

E. J. Rohling, G. L. Foster, T. M. Gernon, K. M. Grant, D. Heslop, F. D. Hibbert, A. P. Roberts, and J. Yu. Comparison and Synthesis of Sea-Level and Deep-Sea Temperature Variations Over the Past 40 Million Years. *Reviews of Geophysics*, 60(4):e2022RG000775, 2022. ISSN 1944-9208. doi: 10.1029/2022RG000775. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2022RG000775. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022RG000775.

S. Talento and A. Ganopolski. Reduced-complexity model for the impact of anthropogenic CO_2 emissions on future glacial cycles. *Earth System Dynamics*, 12(4):1275–1293, Nov. 2021. ISSN 2190-4979. doi: 10.5194/esd-12-1275-2021. URL https://esd.copernicus.org/articles/12/1275/2021/. Publisher: Copernicus GmbH.

Y. Yokoyama, K. Lambeck, P. De Deckker, P. Johnston, and L. K. Fifield. Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, 406(6797):713–716, Aug. 2000. ISSN 1476-4687. doi: 10.1038/35021035. URL https://www.nature.com/articles/35021035. Publisher: Nature Publishing Group.