We appreciate the reviewer's thorough evaluation and recommendations. Below, we indicate the reviewer's comments in blue and our response in black.

L15 The authors state "During the Neoproterozoic (>635 million years ago) …", here the Neoproterozoic ran from ca. 1000 - 538 Mya, and the Cryogenian stage (encompassed the two episodes of the Snowball Earth) ran from ca. 720 – 635 Mya. Your phrasing just needs to be tightened as this would get picked up by someone working in the Precambrian.

To emphasise that our paper is only about system dynamics and not specific events in the geological past of Earth, we removed the Neoproterozoic statement and only refers to "Earth history" in the introduction.

What defines your snowball Earth State? The definition of Snowball Earth generally falls into two camps, the Hard Snowball Earth in which the ocean is fully glaciated with marine terminating tropical glaciers, and a Soft Snowball Earth scenario in which sea ice reached into and perhaps beyond the sub-tropics (extending to approx. 10 - 25 degree latitude, sometimes referred to as waterbelt or slushbelt). Within the manuscript is snowball Earth inception when the tropical ocean has >0% sea ice fraction (as shown in Figure B3). My apologies if I have missed your definition of snowball Earth inception.

In our case the snowball Earth is an entirely frozen world ocean. However, this specification is irrelevant, as the main point of the manuscript is around the moment the Earth enters instability towards that state.

L25-29 If some PMIP4 models do indeed attain a snowball Earth state in the modelling of the LGM, is this more likely due to deficiencies in model parameterisations? If so, are we really deriving insight into the Earth's climate sensitivity? If my thinking here is logical, this needs a brief discussion in the text.

We are unsure what the reviewer means by "deficiencies". The behaviour of the model is anything but unphysical. A model entering a snowball Earth state as a response to a large negative forcing is completely physical, and it was already shown by Budyko (1969) as an expected behaviour of any numerical model which has a reasonable sea-ice albedo feedback. We believe the reviewer might have meant "unprecedented in geological history". Whether some models experience snowball Earth state due to logical behaviour or some deficiencies within PMIP4 is unfortunately unclear because these runs have never been published. This is what we are promoting in our paper. The only published model we know of that has shown signs of a runaway within PMIP4 is CESM2, and that is also a logical behaviour due to its very high ECS and temperature response.

L28 what models are you referring to when you say these models start transitioning to a snowball state (the PMIP4 models that fail to model the LGM?)

In this sentence we refer mainly to CESM2. We know some models failed to contribute to PMIP4 despite attempting to run the LGM (IPSL model, EC-EARTH model). Those models have high ECS therefore it could also be due to runaway. Unfortunately those runs are often unpublished.

L38-40. You state that you are using an ESM. It would be beneficial to the reader if you briefly described which sub-models are included (e.g., is there an interactive Ice Sheet Model incorporated? If so is it a global model or just GIS & AIS domains) or are you referring to an atmosphere-ocean coupled model. I assume the latter. I would also write a sentence or two on the sea ice model as the physics and parameterisations would be important to a reader looking to understand and compare your modelling results with their own. It is also relevant to discuss the snow-albedo parametrisation within the model (e.g., deep snow albedo), I am assuming in these simulations that year-round persistent terrestrial snow cover advances equatorward, so a discussion of snow cover within the model is pertinent.

We cite Mauritsen et al. (2019) in our Methods section, which is an extremely detailed description of the model we used. We did not perform any other modification. There is no dynamic ice sheet model. The sea-ice and snow-albedo parametrisation is essentially the same as Voigt and Abbot (2012), except we have melt ponds, which are described in Mauritsen et al. (2019). The model simulates equatorward advancing sea ice, but the snow cover is relatively thin or absent due to the very cold and dry conditions.

L41-43 You state that you use the “coarse” T31 truncated model for PI and the “low resolution” T63 model for LGM. I would be more consistent with your description of resolution, as T63 is higher resolution that T31. Or drop the terms “coarse resolution” and “low resolution” and use “low resolution” and “higher resolution”.

Coarse and Low are terms which belong to the nomenclature of the models. MPI-ESM1.2-CR is T31, coarse resolution and MPI-ESM1.2-LR is T63 low resolution. Therefore we will not change it in the text. T63 was standard for the CMIP6 simulations conducted with MPI-ESM1.2 for the latest IPCC report.

Are there any differences in other relevant parameterisations between MPI-ESM1.2-CR (T31 31L) and MPI-ESM1.2-LR (T63 47L) that could impact the outcomes of the study? For example, do both models share the same ocean model (in terms of spatial resolution and sea ice model parameters)?

There are some differences in tuning parameters of the model clouds between the CR and LR models. These are intended to compensate for different model biases caused by the coarse resolution. After reviewing the differences, we have identified that some of the parameter settings are likely to alter slightly the cloud feedbacks. Nevertheless, we note that total feedback and equilibrium climate sensitivity is very similar between CR and LR, and when comparing Fig.3 and Appendix A1, this does not affect our results on a global scale, but could slightly impact local climate feedbacks.

It would be useful for the study if the same run (of Table 1.) was conducted for both CR and LR versions of the model. I understand from Table 1 that some of this is an opportunistic ensemble, and so I understand if this isn't possible in a reasonable time frame.

MPI-ESM1.2-LR is drastically more expensive to run than CR, and that is without including the PRP module which already doubles simulation cost. MPI-ESM1.2-LR is also more sensitive to numerical instabilities and we were not able to run it with the large changes of CO2 concentrations that were used in our study.

L47 A bit unclear here for the reader – are you saying that the act of limiting sea ice growth generates latent heat or does thick sea ice itself generate latent heat?

Sea-ice growth creates latent heat, so if thick sea-ice is being artificially removed then latent heat would be generated in the ocean out of nowhere.

Table 1. 1/2056x CO2 should be 1/2048x CO2?

Fixed.

L56 Why does the length of the run vary with forcing? Could you have run each simulation long enough so that you didn't have to use (linear?) regression of TOA radiation imbalance.

The length of the run is proportional to the initial forcing, as larger forcing will imply a larger cooling rate. All simulations eventually reach a numerical instability due to thick sea-ice, therefore larger forcing will result in shorter runs. It is not possible to run the simulation to a stable snowball Earth state with this version of the model.

L55–57 This is a little bit unclear so needs clarifying- are you computing the climate feedback by determining the change in TOA imbalance over a 5 C temperature change?

Yes, we calculate the slope of TOA imbalance versus surface cooling by taking bins of year spanning 5°C of temperatures. Because the simulations are relatively long, but all of the different length, bins of 5 degrees contain enough years to filter out random patterns and allow comparisons between runs over a relevant range of temperatures.

L59-60 It would be helpful to the reader if you could succinctly describe how these are locked (1-2 sentences) L60 mentions that these are imposed on the radiation parameterization, but this is still not that clear to someone not experienced in locking feedbacks.

This is already described in the Methods section: "These locked-feedback transient simulations read the pre-industrial control albedo, clouds, temperature and humidity and impose them on the radiation parameterization regardless of the changes the system is experiencing, such as the increasing extent of sea-ice." A large number of other papers use this methodology (e.g. Wetherald and Manage (1988), Mauritsen et al. (2013)). It is not clear what other details the reviewer is expecting.

L64-69 So are you saying that you are comparing the snowball transition state to the modelled LGM state to derive the Equilibrium climate state?

We are using the transition temperature leading to a snowball state as the maximum temperature the Earth can cool down in an LGM state. Since we know a relationship between simulated LGM temperatures, ECS, and we know that during the LGM the Earth did not enter snowball state, we can infer the upper bound on ECS.

L81 be slightly more specific here and say these all relate to global mean surface air temperature.
L91 “southern sea ice edge” I think you mean something like “equatorward sea ice edge” unless you are just considering the northern hemisphere sea ice extent here.

Fixed.

L114 It is not entirely clear from Fig 3. how the climate transition is broadly similar. What feature am I looking at in Figure 3? (is this where TOA radiative imbalance is most +ve?). Do you need to show on Fig 3 the global temperature in which your criteria for snowball Earth is met?

We added an explanation of the Gregory method in the Methods section. The transition is always the point where the imbalance is the least negative, as any other year in the simulation would be in a more negative imbalance than this one.

Figures 3 and A1 have the 2X CO_2 in the upper right quadrant extending beyond the numbered parts of the y-axis. I would adjust the x- and y-axis numbering so that they extend into the +ve values.

Fig.3 and A1 have ticks which we believe are helpful enough for a reader to understand the values.

L115 Appendix A should this instead be Figure A1

Appendix A contains Figure A1 and a discussion of that figure.

L125 50 ppm is closer to 1/6 of pre-industrial than ¼

This is a choice we made to be consistent with our runs being halving/doubling of CO_2 concentrations from pre-industrial values.

L144 I believe this statement depends upon how tightly you are defining the transition / inception into a snowball Earth state, definition of this transition would clarify this.

We provided a better definition in the methods section.

L161. I am a little uncertain how this 5.5 K (4.4 -6.6, 90% confidence interval) statement relates to the data within Figure 6 – I guess this where your linear fit intercepts with the upper bounds of your instability threshold. Could this be identified within Figure 6.

We added lines on Figure 6 to emphasise that the value and the interval is where our inception temperature crosses with the relationship and its confidence interval.

L301, You state that your model is numerically unstable at low CO_2 concentrations and so compare solar-forced snowball Earth transitions instead. Given that you are running at low CO_2 levels, has this numerical instability impacted your work?

We focused our work on the transition temperature, which happens at much higher temperatures than the ones close to a numerical instability. Numerical instability due to low temperatures also occur in response to solar forcing.

In the text, be consistent with the term "snow ball" or "snowball"
Some instances of informal scientific language used. 36 "notoriously difficult" or L145 "notoriously performed" are not scientific terms (notorious often means something bad or unfavourable that is somewhat common knowledge). Modify also L46 "few simulations" and L88 "All in all".

After reviewing the text we removed "notoriously" at L36.

Equilibrium climate sensitivity (ECS) is the temperature change to a doubling of CO_2 once the atmosphere and ocean (including deep ocean) has equilibrated to that change in energy. I believe here you are in fact just considering only atmospheric changes. I think you therefore need to reconcile this ECS with your definition of climate sensitivity. Here also the definition of your snowball Earth is important (sea ice at the equator? I assume you are holding terrestrial boundary conditions fixed – no expansion of terrestrial glaciation).

No, our ECS follows the same definition as provided in most ECS studies. It includes changes in atmosphere, ocean, vegetation. What is omitted are changes related to methane and other chemical and aerosol feedbacks, however those changes are omitted already in most LGM runs and ECS estimates, therefore the impact can be neglected (Renoult et al., 2023). According to IPCC AR6, the ECS should be defined without changes to ice sheets.