We thank the reviewer for their insightful comments and constructive feedback. Below, we indicate the reviewer's comments in blue and our response in black.

R: The analysis of what feedbacks are important in the snowball Earth transition is not something completely new. For example, the contribution of Pierrehumbert et al. (2011) in the "Snowmip" activity is not mentioned at all (which is odd considering that the authors write about scientifc discussions with Raymond Pierrehumbert and other involved persons in their acknowledgements). Pierrehumbert et al. (2011) specifically discuss that snow and sea-ice albedos together with clouds and atmospheric circulation have a major control of the models behaviour during snowball Earth initiation. There is a short introduction to the topic in the very first paragraph, but I would strongly suggest to add more discussion of the existing literature here.

We added some discussion and references in the Introduction. It is however important to note that Pierrehumbert et al. (2011) summarises past research, notably on Snowball Earth modelling, such as the research of Voigt and Marotzke (2010) or Voigt et al. (2011) which are cited in our paper. Snowmip consists of two coupled models, one of them being ECHAM5/MPIOM which is a previous version of the model we are using and that has been extensively described in papers we cite.

R: l. 31: After the introduction, it is still not clear how the logic behind the ECS constraint works. I would suggest to elaborate on the approach, i.e. that you constrain ECS ultimately by fitting a line for the relationship between a models ECS and the simulated LGM temperature anomaly, and then assume that there is a fixed value of the anomaly at which a snowball Earth would be initiated (in all models). Since a snowball Earth did not happen, you have a new and independent limit on ECS. This reasoning is completely missing in the introduction, so the reader has to guess how this should work. I would suggest to add a bit more explanation in the abstract too, since this is the main point of the paper.

We added some explanation in the Introduction.

R: l. 47-50: How is the growing sea ice problem then treated in MPI-ESM? The reasoning behind having the limit is to avoid numerical artefacts or model crashes as the surface layer runs dry. After reading the rest of the manuscript, it seems like this is just ignored and only the 50 ppm-simulation crashes after 1848 years because of this problem. But why is this not a problem earlier and in the other simulations? This limiter was included, because even pre-industrial states sometimes got too thick ice and crashed. Why are your much colder simulations not crashing quickly?

The sea-ice is left free to grow, until the point where the model stops working, which we refer to as as numerical instability. This is explained in our Methods section: "The growth of thick sea-ice leads to numerical instabilities in the model (> 12 meters). We do not artificially limit sea-ice growth as in other studies (Voigt and Marotzke, 2010; Voigt et al., 2011; Voigt and Abbot, 2012), as it generates latent

heat at the base of sea-ice (Marotzke and Botzet, 2007) which changes the required CO2 forcing for snowball Earth initiation (Hörner et al., 2022)."
Limiting sea-ice thickness can impact the simulation as written above, and is irrelevant for our question as the snowball Earth instability will usually happen before the model reaches numerical instability. The abrupt50ppm run was stopped manually to save resources. It is the only run that does not reach instability; the other runs do crash because of a known behaviour of too much sea-ice in narrow basins, such as the Baltic Sea. This was written only in the caption of Table 1 so we added it in the Methods section as well:
"∗∗The runs were manually stopped and are expected to reach equilibrium in a cold non-snowball state. (Abrupt50ppm)"

R: Methods generally: The authors refer to other papers for their methods, but I think some more explanation should enter also here, to let the reader understand what is being done without having to read up in other papers.

We added a subsection in the Methods which describe the Gregory method (Gregory et al., 2004), and how to read the Gregory plots which are in our paper (Fig.3 and 5)

R: Table 1: It does not really become clear why all the different runs are being done. A bit more thourough explanation of why the individual sets of simulations are being done would make it easier to understand from the beginning, and not after reading the whole manuscrpt.

CO2 has a quasi-logarithmic forcing behaviour so all runs correspond to one half of the previous CO2 concentration, yielding a nearly linear change in radiative forcing. This is a typical approach in climate sensitivity and feedback studies as it simplifies the computation of climate sensitivity. This also allows an easy comparison to simulations at 0.5x, 2x, 4x and 8x pre-industrial CO2 concentrations which are part of CMIP and are often shown in the IPCC reports. We added a few clarifications in the Methods section.

R: When first looking at Fig. 1, I assumed all dots correspond to stable climate states, but later it became clear that these were taken out of transient simulations that are in the middle of approaching a snowball Earth. Over which time periods where they averaged and should these numbers really be produced from non-equlibrium climate states? There generally needs to be more explanation which goes together with the fact that the method section is quite short. There is a similar issue with Fig. 2. These maps are all from one simulation, so over which time periods were they created? A figure with some simple time series of the runs would be very helpful.

Each dot is a one-year average of the same transient simulation as the climate cools down and the top-of-atmosphere energetic imbalance is being reduced. This is a standard approach in the Gregory method as described in Gregory et al. (2004). As this question was common among the reviewers, we added a sub-section which describes the Gregory method and how to read Gregory plots of

Fig.3 and 5, as well was a schematic explaining the approach and how it compares to the real climate system. If the forcing is abruptly set at the beginning of a run, the Gregory plot already acts as a time series: the beginning will always be the right-most point, when temperature is the closest to pre-industrial and the TOA imbalance is equal to the forcing imposed, and the end of the simulation is the left-most point, where in our case the temperature is the coldest. For the maps, they are averaged over a non-fixed time period, as our feedback calculations are averaged over bins of data points spanning 5 degrees of cooling: this is explained for Fig.1 but this was not properly acknowledged for the map. 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.

R: l. 106-108: First, it is said that positive cloud feedbacks at the sea-ice edge decrease the temperature of instability, then it is said that cloud feedbacks substantially increase the threshold CO2 level, which are opposing statements. I can imagine this dicrepancy comes from comparing local to global effects of the cloud feedback or because neglecting the cloud feedback would decrease global mean temperatures at a reference CO2 level, but please explain in more detail what is the cause of the discrepancy here.

Cloud feedbacks are even more positive at stronger forcing in the initial degrees of cooling (Fig.1), which influences the cooling rate of tropical oceans and affect the temperature of stability. However, interactive cloud feedbacks also allow snowball Earth states to be triggered at much higher CO2 level than when cloud feedbacks are locked (zero). Those two statements are not contradictory as they both refer to local and global effects, as well as the influence of clouds and time and state. We acknowledge the confusion and lack of explanation of Fig.1 and we added details in the text.

R: Fig. 3 description: Terming some phrases "stable climate" even though some of the runs are in the middle of a transition towards a snowball Earth seems incorrect to me. Please elaborate and be more specific about what is actually meant here.

We removed stable/unstable and only kept positive/negative feedback phases.

R: l. 115: To me Fig. 3 does not really show that the different runs transit towards a snowball Earth at roughly the same (transient) global mean temperature. Again, a simple figure with time series of global mean temperature in all the runs would be helpful to support these kind of statements.

We added a subsection on the Gregory method and Gregory plots. A figure with a time series would show exactly the same global mean temperature, which is now shown by the dashed line in the schematic figure we added in the Methods section.

R: Fig. 4: Generally, schematics are nice, but this one is hardly telling a story. Could it be a bit more elaborate or just left out? If you want to keep this in, I am not going to oppose.

We would like to keep it.

R: Chapter 4: I am having a hard time believing that the temperature at which the climate transits to a snowball Earth state is similar across CO2 forcings, setups and even supposedly climate models (l. 118). First of all, what temperature is even meant here? These are all transient simulations, so I assume it is not the global mean temperature of the last stable pre-snowball equilibrium climate (which would surely be highly dependent on the climate model). Is it the (global mean) temperature at which the TOA radiative imbalance starts to grow over time again, as marked by the different phases in Figs. 3 and A1? But then, this is all but a clearly defined temperature range, as can be seen from the large ranges where the imbalance goes sideways in some runs, and even if taking the shown different colors as markers, there is still a range of ~10 K between the different runs. Does this count as "broadly similar"? Especially the statement that this temperature will still be the same even with other climate models seems dubios to me. Again, some of this issue could potentially be resolved by simply explaining more thourougly the procedure that was taken.

We added explanation on the transition temperature in the subsection regarding the Gregory method. The snowball Earth temperature is the point where the TOA imbalance is the least negative: any other year is more unstable than this point, therefore it is the last moment where a stable climate state can exist before transiting to a snowball Earth state. The reason why it is not a fixed value but a range of temperatures is explained in Section 4: “Nevertheless, the transition temperatures of each phase show a slight shift to lower values under stronger negative forcing. Therefore, the climate system deviates from pure state-dependent behaviour as the strength of the radiative cooling and the speed of transition to snowball Earth increases.” We also explain a bit further than fast simulations, which are used here to highlight those effects, are not preferable: “. All in all, we suggest slow, low forcing simulations are preferable when analysing snowball Earth transitions, as 1) fast transitions to snowball Earth are hardly realistic, as geological snowball states may form over millions of years (e.g. Schrag et al., 2002) and 2) fast transitions involves temporal effects which would depart from state-dependency”.

R: l. 127: Now it sounds like the "inception temperature" is actually the global mean temperature of the last stable pre-snowball climate state? A clear definition of what you mean by this temperature is highly desirable. Also, the final global mean temperature of the 50 ppm run (assuming it reached climatic equilibrium and that a further lowering of CO2 would drive the climate into a snowball state) IS probably the "inception temperature", when defined as above. I don't think that the other transient simulations can give any more reliable input. Hence, simply finding the last stable pre-snowball climate by iteratively changing the CO2 concentration of individual runs would give a more reliable and more precise estimate of the "inception temperature" for a given model setup.

We added details in the Gregory method subsection. Iteratively changing CO_2 concentrations to identify the transition temperature is indeed what we did in this paper. As we point out here: “When abruptly decreasing the CO_2 concentration to 50 ppm (around 1/4 of pre-industrial CO_2), we find hints of instability near the global mean temperature of 0°C”, we indeed believe the inception temperature is close to 0°C. This simulation would be however more than 2000 year long, so continuously iterating around that value would be extremely expensive, and also if our study only focused on values around 50 ppm we could not have been able to discuss state and time dependency over large ranges of CO_2 concentrations and thereby relate to other studies.

R: l. 142: I find it very problematic to make such statements. First of all, what is the uncertainty range here? It seems to be in the order of at least 5 K. Second, it should be made clear that if this would be a sound statement, it would only be valid for the modern arrangement of continents and not for the setup during the Neoproterozoic, where the snowball Earth actually occurred. This needs to be specified. Lastly, this temperature is in fact highly model dependent. From personal experience, I can say that parameterisations like the sea-ice and snow albedos or even small changes in parameters of sea-ice dynamics can shift the transition towards snowball Earth inception substantially.

We added the effect of uncertainty on Fig.6, where we show the impact on the constraint is actually minimal compared to previous estimates of the upper bound on ECS. Whether this argument is not valid for Neoproterozoic continents is irrelevant for our question: we actually need to have modern continents to constrain the upper bound of ECS from LGM simulations, as these use modern continents, and the message of this paper is not about the geological past of Earth. While the temperature response to a given forcing could be model-dependent, we believe state-dependency around the inception point could be similar across models, and published studies, including this one, show that it is indeed the case for MPI-ESM1.2, CESM1.2, and CESM2 (Eisenman and Armour, 2024)

l. 151: how does Fig. 5 show that the instability is around 0°C?

We added details on how to read a Gregory plot in the Methods section. Fig.5 shows that CESM1.2 has a similar behaviour and transits at almost the same temperature as MPI-ESM1.2: in the case of a simulation at around 2 ppm of CO_2, this is around -5°C. Since we show in Fig.3 that the time-dependency effects move the transition temperature from 0°C to -5°C and below, we expect CESM1.2 to also transits at 0°C under much lower forcing, particularly since it is the case for CESM2 (Eisenman and Armour, 2024). Unfortunately, such simulation would take months to run for CESM1.2 so we decided to save time and resources when running CESM1.2

l. 162: How are the critical ECS and the confidence interval computed? It seems like the values just come from the fit of the regression line in Fig. 6. But how does this account for the uncertainty in the "inception temperature", which surely has an uncertainty range of several degrees, maybe up to 5 K or more. This would substantially increase the uncertainty in the calculated upper limit of ECS.

We added the effect of uncertainty in Fig.6. Even with a large uncertainty of 5 K on the transition temperature, this affects the upper bound of ECS estimate uncertainty by a bit less than 2 K. We emphasise that we are constraining the upper bound of ECS here, not the best estimate value, which is necessarily lower. Much larger values on the upper bound are often given, and with some lines of evidence the upper bound is basically infinite. Therefore, even providing a value of 10 K brings valuable information to the community wide effort of constraining ECS.

l. 179: Point 3 is not really part of the recipe from the points above, but rather a general proposal, hence it doesn't really fit into the list.

Adjusted.

We applied corrections based on the following comments. For the case of 50 ppm written as “close to 1/4”, this is a choice we made to be consistent with our runs being halving/doubling of CO2 concentrations from pre-industrial values.

- l.15 and other locations: To my knowledge, it should generally be "sea ice" without the hyphen, but then "sea-ice albedo", i.e. including a hyphen when combined with a following noun.
- l. 18 "referred to as"
- l.38 bad punctuation around MPI-ESM1.2
- l. 69-70: Example of weak language, making it hard to follow the text. "... the highest value of the Earth's ECS that does not lead to an unstable LGM state represents an upper limit..."
- l. 93 "snow ball"?
- l. 125: 50 ppm is rather 1/5 to 1/6 and not 1/4 of PI CO2, why not be precise?
- l. 140: "involve"
- l. 150 "as MPI-ESM1.2"
- l. 156-158: bad punctuation or sentence structure
- l.176 "surface"
- l. 184 "...model Earth's climate sensitivity." What does this man?
- l. 290: The doi in the reference does not go to the actual article, but to an eossar link. Please link the actual article.