We thank the reviewers for their comments, in particular Prof. Maria Rugenstein for joining the review and providing insightful comments. Below are our answers in blue.

Reviewer #2:

I suggest to reject the paper for two reasons, each of which alone is sufficient for rejection.

First, the scientific essence of the paper relies on the emergent constraint established in Fig. 7. I have come to the conclusion that there is no linear relationship between LGM TS anomaly and ECS across models unless either CESM2 or the other CESM models are added (for further details see my last review comments). It is not scientifically sound to claim a robust relationship when in fact all of it stems from one model (family). I have given the authors ample opportunity to convince me of the opposite, but they have not. I have received a lot of comments about different PMIP generations, differences between global and tropical SST, and more. What remains is the fact that Fig. 7 shows a linear relationship, which entirely falls apart as soon as the upper left point is left out. The authors' main argument to refute this is that one could replace the CESM2 point by other points from the CESM family, which however does not change the fact that the relationship rests on one model (family). As a side note to this: I reject the notion raised in the authors' last response that "the method is not invalid". Of course, the method of emergent constraints is not invalid in general. But it is invalid in this particular application because there is no sufficiently robust inter-model relationship.

First, we would like to remind the editor that we still did not get access to the evidence described by the reviewer. As previously mentioned in our last answer to the reviewer, the document where the reviewer described their approach was not included in the comment, which means our response is based on a guess.

The scientific essence of the paper does not rely only on the emergent constraint: it relies on climate feedback estimates, and how they could be used within a hypothetical emergent constraint framework. It is physically reasonable to assert that models with higher ECS will cool more under LGM conditions, and therefore be closer to the snowball Earth inception point. Whether the constraint can be used as such, or how can it be improved (e.g. by inviting more models to perform these simulations) is an open question, which we deem useful when discussing the design of simulations for PMIP.

On several times, the reviewer specifies "one model (family)". There is a substantial difference between the relationship holding on a single model, or on a family of model: many models already share similarities in components or physics, yet it does not prohibit emergent constraint from existing, nor does it prohibit them from using single-model ensembles or family ensembles. Models within the same family (e.g. CESM1.0.5, 1.2, 1.3, 2) can display substantial differences in physics, to the point that some models from different modelling centres share more similarities than models developed within the same centre but on different generations (Knutti et al., 2013). We added this aspect in the paper, as to emphasise that emergent constraint relationships, in particular in paleoclimate settings, have to deal with limitations that force to use models that share similarities. This can notably be seen with CCSM4 and NorESM-L in Hargreaves and Annan (2016) or different versions of CCSM4 in Haywood et al. (2020). In the previous answer to the reviewer, we proved that the relationship is robust by notably including the CESM model family, which is not an outlier. Therefore rejecting the relationship because it holds on one model family is incorrect.

As mentioned in our last response to the reviewer and in the manuscript, the CESM family is exchangeable with CESM2. The only reason we ended up using CESM2 alone, is because we do not have access to the surface temperature field for the CESM family ensemble, but only SSTs, which would have led to confusion for the reader, since we discuss our results based on surface temperature values. Surface temperatures is the usual variable used in top-of-atmosphere variation plots ("Gregory plots", such as Fig.1). We simply verified that CESM2 or the CESM-family ensemble provide similar statistical parameters for the relationship (slope, intercept, significance). In the first submitted version of the manuscript, a similar figure as Fig.7 was included, where the ECS was reported versus SSTs, which allowed to include the CESM model family AND CESM2. The result is that the upper bound on ECS is 5.5 K (4.4 - 6.6 K, 90% confidence interval) using a threshold at which the Earth begins to initiate a snowball Earth transition from SSTs of -7.5°C. We showed that, when including the CESM family, the absence or presence of CESM2 affects only slightly the relationship and the median estimate of climate sensitivity (as it is also published in Renoult et al., 2023). We decide now to include again a version of Figure 6 of the first manuscript into the new manuscript, under the name Fig.7-B), and adding in the original text: 1) the temperature threshold value as seen from the SST perspective, which was originally in the first version of the manuscript, and now also in the shape of a Gregory-plot in Appendix, and 2) a quantification of ECS estimates when CESM2 is not included, which decreases the median value by roughly 0.5 K.

Secondly, on a higher level, I find the authors' response to my comments unsatisfactory. To make my point clear, this is the question I asked in the second round of revision: "I wonder how much the relationship would change if the upper left point (CESM2) were left out. Specifically, what are the central values and uncertainty intervals for the upper limit of ECS (assuming inception temperatures of 0 degC and 5 degC, respectively) if this model is discarded?" This question has a very clear answer that the authors could have provided by repeating an analysis with just one data point less, a task I later completed with minimal effort. Instead, I received a long answer asserting that "The relationship is not sensitive to the inclusion of CESM2" and that "the inclusion (or non-inclusion) of CESM2 and/or the CESM model family leads to the same slope and intercept for the emergent constraint relationship". I then showed both of these statements to be false. Excluding CESM2 and the CESM model family does lead to a different slope and intercept. I want to stress that this was not a side point, but the core of my very specific question. When I reiterated this point, the authors responded in the third round by stating that they "wished to describe that CESM2 and the CESM-family ensemble are exchangeable, meaning the relationship is significant as long as one of the two is present". This response appears inconsistent with their previous statements, which suggested otherwise. I am concerned about the shifting nature of the responses. In particular, had I not questioned the statements made in the second round of the revision and simply accepted the authors' assertions, the problem I raised in the first paragraph of this review would have remained unnoticed.

No, the relationship does not hold without CESM2, but it holds with the CESM-model family. The CESM-model family is not an outlier: it is an ensemble of 6 simulations with climate sensitivity close to the rest of the PMIP4 ensemble, which strengthens the relationship. Statistically, it is also exchangeable with CESM2, as published in Renoult et al. (2023). If we were to remove AWI-ESM-1-1-LR and AWI-ESM-2-1-LR on the basis that they are in a similar family, both models with a climate sensitivity of 3.6 K and SST anomalies around -1.6°C, the relationship would also most likely strengthen: however, we cannot arbitrarily remove models that share similarities because it is convenient or not for the relationship, unless they are visually a very strong outlier, such as the case of CESM2.

In the sentence "the inclusion (or non-inclusion) of CESM2 and/or the CESM model family leads to the same slope and intercept for the emergent constraint relationship", the reviewer picks the case where both CESM2 AND the CESM model family are not included, which we believe is more a lack of clarity on the grammar of the sentence in the manuscript than a scientific mistake. We have now adjusted the grammar of these sentences in the text to emphasise that there is no case where one or the other should be ignored.

There is indeed no reason to not include CESM2 or the CESM model family: the CESM model family exists, is not an outlier, and as discussed above, may display stronger dependencies with other models in the pool than between themselves. It is also a convenient and common method to circumvent the lack of models in PMIP ensembles. The fact that the CESM family and CESM2 are exchangeable is published in Renoult et al. (2023), in the Appendix of the paper and in our last answer to the reviewer, and is only used to facilitate the reading of the figures. As aforementioned, the original manuscript included a similar figure using SSTs, which we deemed confusing as the manuscript uses surface temperatures otherwise. When using SSTs, there is no reason to exclude CESM2 and the CESM-model family, as their combined effect strengthen the relationship and cannot be considered as an outlier effect. It is physically understandable that a high ECS model also cools more under LGM conditions.

On the contrary as what the reviewer suggests, the inclusion of CESM2, and the CESM-model family, and its consequences on the strengthening of the constraint should be seen as a strong motivation for modelling centres with high ECS models to also contribute to LGM simulations, rather than a reason for rejecting the method.

Prof. Maria Rugenstein:

I was asked to look over the reasoning for reviewer #2 to reject the paper. They raise two arguments: 1) the constraint in Fig.6 (former Fig. 7) is not robust and 2) the answers to the issues brought up before are not satisfactory.

I do not think the paper has to be rejected because of the potential non-robustness of Fig.6, as the rest of the paper seems valid and adds to an understudied problem and presents the findings well. Even if the relationship in Fig.6 was not robust, laying out the reasoning of the transition temperature being able to constrain ECS is worthwhile and the recipe the authors provided, might motivate some modeling centers to invest into paleo or LGM simulations in general. It's not the authors' fault that not many climate models take part in the LGM experiments and many paleo studies rely on relationships between even fewer models and even relying on outliers from one model family is (unfortunately) common for emergent constraints.

However, the reviewer has a point in that this needs to be discussed more clearly. Currently, around line 75 there is a tight mentioning on the subject of "model families" and Fig.6 is also barely explained. I suggest to add *one paragraph* around Fig. 6 (on page 9, 10 or 11) which states a) the issue of having only few simulations available, as is already mentioned in paragraph 165, b) the regression slope sensitivity to the CESM model family and other models, *quantitatively*, as reviewer #2 demands; leave-one-out seems meaningful to me.

We appreciate the reviewers positive assessment of our study. We have added a comment on "model families". Notably, models coming from the same modelling centres should not preclude their use in emergent constraints, as some models actually share more similarities with models from other modelling centres than with other model generations from the same institutes. As the reviewer here comments, there is limited data available, and that is common in paleo-based emergent constraint studies. We assume the reviewer here refers to Fig. 7 when mentioning Fig. 6, following the earlier comment of the reviewer "1) the constraint in Fig.6 (former Fig. 7)…". We have expanded the methods section to emphasise the limited paleo simulations and data that can affect Fig.7.

We have also added a figure (Fig.7-B)) and a quantitative analysis of the ECS estimates when the official PMIP version of CESM2 is not included (but the CESM family is, i.e. 1.2, 1.3 and low ECS versions of CESM2). This was originally part of the first submitted version of the manuscript, but we removed it during review as the figures used SSTs (since we do not have surface temperatures for the CESM family), and we thought it would be confusing for the reader since the snowball Earth transition threshold is computed from surface temperatures. Since the addition of the new figure required to also introduce the threshold temperature for the snowball Earth initiation, we also included a Gregory-type plot in the Appendix.

The reviewer #2 pointed out a case where both CESM2 and CESM-model family are excluded together, which is a grammatical error on our side: there is no scientific reason to exclude both together, because as published in Renoult et al. (2023), as shown in the manuscript and in the review, CESM2 does not act as an outlier when the CESM-model family is included. There is also no reason to test the relationship from surface temperatures without CESM2 (and without the CESM model family), because a statistically significant relationship does not exist: this was published in Renoult et al. (2023). We have expanded the discussion on the exchangeability of CESM2 and the CESM-family, and provided the different sensitivity tests under different configurations of the PMIP4 ensemble.