Reviewers' comments are in blue. **Responses are in black**. New additions to the text are reported in *Italics*.

RC2

This manuscript presents a well-written and comprehensive overview of the next generation of GeoMIP experiments, and the authors have (mostly) done a good job with producing what will likely be an important reference documents for years to come. However, there are a few major and several minor points that should be addressed prior to publication, as follows.

There is very little attention to model evaluation in the manuscript, and assessment of whether models are fit for purpose is largely ignored. For the SRM methods considered (SAI, MCB, MCT) I would as a minimum expect a couple of sentences for each about what variables or benchmarking cases models should simulated well for us to have confidence in them (as now briefly mentioned for MCT, where cloud phase is stated to be a key variable)

We thank Dr. Storelvmo for her comments and feedback. The point on validation is particularly well taken, and we have made sure to address it in the revised manuscript, by mentioning in Section 5 specific documentation and validation needs: "*It will also be important for modeling centers to ensure their documentation for the models' versions used for GeoMIP experiments includes information related to components of the model necessary to understand the response to specific SRM methods (for instance, cloud microphysical schemes for MCB, CCT and MCT, details of the treatment of stratospheric aerosols for SAI), as part of the documentation required for the CMIP Rapid Evaluation Framework (REF) (Hoffman et al., 2025), as well as specific diagnostics that would enable a better validation of SRM-specific aspects (for instance, diagnostics from historical simulations with interactive volcanoes to enable comparison with observations, Quaglia et al. (2023)) in order to make systematic benchmarking of the models used simpler (Hassler et al., 2026).*"

The choice of the ML scenario as the only baseline scenario in the Tier 1 experiments needs better justification. Arguably, with its very steep drop in emissions starting almost immediately, this scenario appears on the optimistic side given that global CO2 emissions are still going UP, not down. As such, it could have been valuable to place some more emphasis on the M scenario. I was very surprised that when a second baseline scenario was proposed in the Tier 2 experiments, it was the LN scenario (which I don't think anyone really thinks is going to materialize), and not M, which from my reading of Visioni et al. (2025) was the alternative scenario being considered by the community.

This is a fair observation, and one we should have discussed more in the paper, considering it was a large topic of debate during the Tokyo GeoMIP meeting. We added the following paragraph to Section 3:
"*During the Fifteenth and Sixteenth GeoMIP workshops, there was extended and lively debate over the selection of the ML scenario over the M one: while neither are entirely compatible with SSP2-4.5, ML is certainly on the more "optimistic" side, with emission reductions starting sooner and far more extensively as compared to M, where emissions plateau throughout the century. While full consensus was elusive, a*

***clear majority of participants preferred ML on the specific ground that simulating SRM methods under an emission scenario with strong mitigation would strengthen the message that SRM should not be considered as a substitute to mitigation, and that a stronger physical signal with more cooling could be obtained by simulating a scenario with more cooling like G7-0.5K-SAI (see Section 4.4)*"**

The paper discusses preliminary model results, which is very nice, but many of the related figure panels are never referred to. An example is Figure 3, which is never referred to in its entirety. There are references to Figs. 3b, d and f but there are no figures 3b, d and f (Figure 3 lacks sub-panel labeling). Likewise, Figure 5 is not referred to in its entirety, only sub-panel 5a (which does at least exist). This is sloppy - either figures should be referred to or they should not be in the paper.

Thank you for this note. We tried to balance providing some level of results without making that the focus of the paper, and we should have tried better! In the revised version, we made sure to discuss some figures more in depth, and moved some others to the supplementary following the suggestion from Reviewer 1.

We add the following paragraph on the MCB figure:
"Figure \ref{fig:gm6cb} shows results from the G6-1.5K-MCB experiment across four ESMs. The global-mean surface temperature is maintained at the 2020-2039 levels of the SSP2-4.5 simulations (panel c), by injecting sea-salt aerosols in the five midlatitude regions shown in panel f. While the warming pattern from SSP2-4.5 (panel e) is largely offset by the cooling pattern from MCB (panel g), some residual differences remain (panel f) with overcooling over the oceans and residual Arctic warming. Models differ on the interhemispheric temperature asymmetry under both SSP2-4.5 and G6-1.5K-MCB (panel d). The mass required to maintain temperatures at 2020-2039 levels differs substantially between models (panel a), but a lot of that difference is due to discrepancies in the size of the emitted sea-salt aerosol such that the number emission rate differs less between models (panel b)."

Cirrus Cloud Thinning is completely ignored in the manuscript, after being mentioned alongside MCT in the introduction. I am not necessarily calling for a CCT experiment (although there appears to be some renewed interest in this method), but a few sentences about why there are no experiments exploring CCT further this time around would be good.

Thank you for catching this! We have added the following phrase to the Introduction: "*Experiments described here include SAI, MCB and MCT. An idealized experiment involving CCT was proposed in Kravitz et al. (2015) (G7cirrus) but its results from four ESMs were only recently analyzed in Smith et al. (2026), and the community has not expressed specific interest in updating this experiment or proposing new ones for CMIP7; this doesn't preclude that renewed interest might result in future GeoMIP CCT experiments.*"

Line 58-59: climate modelers are not distinct from climate researchers, but rather a subset of climate researchers

Edited to “2) offering a standardized framework of simulations for the purpose of broader assessments of SRM, making sure decisions around experimental details are taken by a broad representative community of climate scientists, which includes modelers and those interested in potential downstream impacts of SRM (including on agriculture, society, infrastructure and ecological systems) that is capable of assessing and documenting different needs and opinions;”

Line 376: This is an interesting point regarding the non-linearity of MCB. This sentence refers to how efficiency may go down as sea salt emissions go up, but in the baseline scenario anthropogenic aerosol emissions also go down quite rapidly, could this matter?

Response:
That’s a good point, we have not quantified how the MCB forcing may change under different levels of anthropogenic aerosol emissions. But we add the following: “This makes it challenging to estimate a full MCB emission time series prior to running a scenario simulation (also known as a "feedforward" estimate), because the model sensitivity derived from Inj-seasalt-midlat will be an underestimate early on in the simulation, when emissions are still low, and an overestimate late in the simulation, when clouds become saturated with aerosol. *We also note that anthropogenic aerosol emissions are projected to decrease in future scenarios which would make clouds more susceptible to brightening by MCB, but this has not been quantified.* Feedback controllers can partially account for such nonlinearities, as they can adjust emissions in response to these deviations.”

Line 534: These INP perturbations would be expected to over-seed Arctic mixed-phase clouds. Why such high concentrations?

Response:

The perturbations of 10^4, 10^5, and 10^6 L^{-1} were only used in the initial tests. For MCT-L1-Seeding, the proposed perturbation strengths were instead chosen as 10^2, 10^4, and 10^6 L^{-1} (Table 7) to capture weaker perturbations.

Although high perturbation strengths, such as 10^6 L^{-1}, were demonstrated to over-seed (i.e., increase the total water path) in Arctic mixed-phase clouds, it remains uncertain as to how robust this signal is across models. We therefore still propose including such high perturbation strengths in MCT-MIP to determine at which seeding levels different models begin to exhibit over-seeding, as observed for E3SMv3 in our initial tests.

Table 6: It would be very relevant to add information about whether models include secondary ice production, if possible

Response:

For all four Level-0 models, Hallett–Mossop (rime splintering) is the only secondary ice production (SIP) mechanism considered. The table was updated.

Line 605: ODS abbreviation is never introduced

Since we only use it once, we now explicitly spell it out as ozone depleting substances.

Line 660: This should be «timing», not «timeliness», I think? Generally, there are a few typos here and there that I didn't bother pointing out, but an additional round of proofreading before re-submission would be beneficial

Thank you, we have indeed found some that we have now fixed.