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

RC1

Visioni et al. describe the plans for the contribution of a geoengineering MIP to CMIP7. This is a very laudable community effort, and the modelling community evidently organised the process of defining the simulations and simulation protocols in a very open and transparent way. The paper states that much of the definitions is based on the outcomes of a workshop to which the community was invited.
I have a few somewhat overarching suggestions for improvements, and then a number of smaller suggestions.

Some of the propositions/requests in my view are too vague to be a meaningful MIP contribution. Of course each modelling group can (and should) do their own simulations, and smaller groups can always come up with own ideas. The very meaning of a MIP is that it is a strict protocol and that a meaningfully large number of groups submit simulations that exactly follow this protocol. So my suggestion is to omit everything that is only vaguely defined. There could be for example an appendix to the paper in which such suggestions – potentially these are minority opinions from the workshop that didn't make it to the Tier-1 MIP proposal? – can be listed as examples.

We thank the reviewer for their supportive comments and for their suggestions. While true that MIPs have somewhat strict protocols, we believe there is a difference between strict experimental protocols and *mandatory* experiments - not that modeling centers can be mandated to do what they don't want to do (for instance, looking at the DAMIP paper (Gillet et al., 2025), there are plenty of cases of experiments included in DAMIP but only run by 1-2 models). In any case, our hope with the inclusion of the Tier 2 experiments is to make sure similarly strict protocols can be maintained also for experiments that only a subset of modeling centers will do, in order to facilitate intercomparisons later on and to avoid duplications of similar efforts. We strongly believe the Tier 2 list of experiments should be maintained in the main text.

A bit related comment: The real MIP proposal is the Tier-1 simulations. However, they are impossible to do without some of the Preparatory experiments. These Preparatory experiments, in turn, are not understandable really before one understood the Tier-1 experiments. To the reader of the paper, a much more logical structure would be to describe the Tier-1 experiments, and include in these the Preparatory experiments (except for the third block, Inj-Sulf-15N etc., which seem to be of rather little use except for the zoo of Tier-2 experiments). Of course I acknowledge it might be difficult to change such things after the workshop, though. In the Tier-2 experiments, to me only the experiments listed in Table 3 are well-enough defined to be part of CMIP7. The other experiments are very vague and would better be done and consolidated within the community. They could be described here in an Appendix making clear such ideas were discussed but then not included as being well-defined enough for a large MIP effort as part of CMIP7. Again I acknowledge that the

community may have discussed such arguments in detail and the harm done in listing such ideas here is probably minor since groups will carefully select what they contribute anyway.

We really thank the reviewer for their comment, and for acknowledging the challenge in reporting the results of community consultations while taking in valuable feedback. In response to this, we have updated Table 1 to make it clearer which preparatory experiments are required for Tier 1 experiments, and which are not.

The integration of GeoMIP into CMIP7 should be discussed. In particular, what are the synergies with AerChemMIP and DAMIP?

Thank you for the suggestion. We have added an extended reference to DAMIP in Section 2.1 "*Additionally, it is worth noting here that, as part of the Detection and Attribution MIP (DAMIP, Gillett et al., 2025), there are CMIP7 experiments aimed at isolating the historical volcanic forcing (hist-volc), and that CMIP7 forcings will also include volcanic SO2 emissions (Aubry et al., 2026): the use of these simulations could be rather important to evaluate the SAI response in models in light of their response to historical volcanism.*" and mentions to DAMIP and AerChemMIP2 in Section 5.

Comment: There are very important aspects of the Earth system models that may have a large influence on the outcomes (e.g. the cloud microphysical schemes for MCB and CCT). The protocol should define a way to document these particularly relevant ESM aspects in an easily accessible and transparent manner.

This is a great suggestion. In general, for CMIP7, modeling centers have already been asked to provide documentation for their CMIP7-ready version, and all of it will (should) be available through the Rapid Evaluation Framework. But we added a paragraph to Section 5 to make sure modeling teams are reminded of this. "*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 (Hoffman et al., 2025), in order to make systematic benchmarking of the models used simpler (Hassler et al., 2026).*"

Comment: Several figures show results but these rather distract from the MIP description. Fig. 3, 5, 7, 8 in my view contribute little to the paper but are rather confusing. They could go to an appendix or to a separate paper.

Thank you for the feedback. We have moved Figures 2 (which was not amongst the suggested figures from the reviewer, but we deemed that 2 was of somewhat less relevance than 3) and 8 to the Appendix, but we believe the other two figures can be quite useful for a reader to understand some of the broad responses from the currently available models, especially based on feedback from some of the GeoMIP users and from past GeoMIP description papers. We have also made sure to update captions and text to make the usefulness of the figures more evident.

L12 – “that all participating models are encouraged to run”: the corollary is that models are discouraged to run tier-II simulations

Fair! We changed it to “should”.

L15 – “propose their own experiments” links to my main comment above: this is not a contribution to a MIP but the bread-and-butter of every modelling team. It should be omitted from this paper or at best be moved to an appendix

We disagree, as explained above. While true that single modeling teams always do run their own simulations, finding a way to connect those targeted experiments to the broader purpose of a MIP could be useful, especially in helping users find those simulations and encouraging others models to join in - again, the DAMIP paper Fig. 1 is actually a good example of experiments run by only one models and sparsely cited: we would like to improve on that!

L26 – “increase of Earth’s albedo through _increased_ cloud brightness”

Fixed!

L26 – what could be “other cloud changes”? cloud coverage?

Added “*such as changes in cloud coverage*”

L47 – the “potential asymmetries” should be explained

Added “*in deployment across hemispheres*”

L111 – “this could include metrics”: I would expect from a MIP definition a definition of specific metrics. Maybe the text could point here to Section 5 (Data request) and take up this point in that section 5.

Good point. We went more in detail in the section, adding “*This could include variables that enable the evaluation of changes in monsoon rainfall (daily precipitation, sea surface temperatures), Arctic sea ice (extend and volume), and the AMOC (through ocean circulation diagnostics), as well as variables that enable estimates of hydrological cycle changes (see for instance the analyses of freshwater availability in Rezaei et al., 2026).*”

L125 – “Models can also perform…” this seems to be a redundant statement, of course they could. If there is a reason to state this here, this reason should be made transparent (e.g. “The GeoMIP community has also discussed simulations at other latitudes… but decided not to include those in the Tier-1 request.” or so).

Agreed. We simply removed the phrase from there and discussed the benefits of SPIs elsewhere in the section.

L141 – Only later the emission height and size distribution are defined, it would be useful to point to this here.

We edit as follows : “This simulation should use present-day SST and emission fields if possible and we recommend an initial emission rate of 100~Tg~yr^{-1} of NaCl (see below for the recommended aerosol size distribution) at the lowest atmospheric level. This falls roughly within the emission rates found in Alterskjaer et al, 2013 and Hirasawa et al, 2026.”

l147 – The community uses the term “aerosol-cloud interactions” since AR5, “aerosol indirect effects” is outdated.

We edit as suggested.

l155 – It should be clearly defined how this careful documentation should look like.

Good point! We add the following : “This documentation should include the assumed dry aerosol size distribution (mode radius and geometric standard deviation, or bin/mode boundaries for sectional schemes) and the assumed particle density.”

L160 – The Table said it is run for 20 years (not 40 years)

Thank you for catching that! We edit to “for a minimum of 20 years starting in 2035”.

L202 – Only one ensemble member run to 2150, was stated in the table.

The sentence as written implies that all ensemble members should run until 2150, which is what we intended.

L225 – “as little as possible” is a difficult statement, maybe simply “little”?

We edit to “avoid ITCZ disruption”.

L232 – I don’t understand the “Legendre polynomials” explanations. A clear definition of who T1 and T2 are computed should be provided.

We apologize for the confusion. We explained T_n in a very complicated way, now we replaced that with a clearer explanation including the way they are calculated.

L248 – For the MCB experiment, a definition for “maintain the PI+1.5°C temperature target” is provided, in line 332, at ±0.2°C. This seems meaningful.

Agreed. It was implicit, as it was also the indication for the old G6sulfur experiment but now we explicitly report it. We chose ±0.1 because for SAI the AOD to temperature response is more linear and thus more controllable, and modeling centers managed quite well for G6sulfur.

L254 – “it is possible” rather than saying what is possible the protocol should describe how it should be done exactly.

The sentences that follow describe how modifying their original algorithm for similar 3-DOF and 1-DOF experiments is done. We edit as follows: “it is possible to modify their original algorithm for similar 3-DOF and 1-DOF experiments with the following steps.”

L258 – “can be computed” similar point, it should be said what should be computed and how.

Thank you for pointing this out. We added the proper references for the computation.

L267 – “can request a dataset” – how exactly, to whom to address?

We added that the GeoMIP co-chairs can be addressed for this.

L327 – “lowest model layer” seems to be too vague. Models have vastly different vertical resolutions. The request should clarify a height range instead.

In general, natural sea salt aerosols are emitted into the lowest model layer. We added this specification to clarify that sea salt aerosol for MCB should also be treated in the same manner. While this might affect models’ response, it is not yet clear what height MCB aerosol will be emitted in an actual deployment, so we rely on the existing model choice for natural sea salt aerosol here. Also, it is much simpler for models to specify a surface emission than one at a specific height. We added “at the same vertical level as natural sea salt emissions (typically in the lowest level of the atmospheric grid)”

L329 – I would find a map useful.

The red boxes in Fig. 4 (new) show those ranges, so we added a pointer to that here.

L334 – Only one ensemble member to 2150, right?

No, as clarified above, we are recommending all ensemble members to go to 2150.

L363 – Also the number is evenly distributed since the size distribution is given.

Indeed. We added this for clarity.

L411/Figure 6 – The “intervention magnitude” curve for G7-0.5K-SAI seems to be wrong to me. The magnitude should be much larger very early after 2035.

Great catch. We modified the figure to ensure it reflects this.

L421 – If “ramped up linearly”, the temperature evolution would look very different from the one depicted in Fig. 6.

Yes, this is now taken care of in the revised figure.

L424 – But here no overshoot is considered?

That depends on what one considers "overshoot". The LN scenario is defined in the new ScenarioMIP paper as one with "[...]a higher overshoot of the 1.5 °C goal, followed by stringent climate policies resulting in net-negative greenhouse gas emissions to return to lower warming levels, thus supporting research into the reversibility of climate outcomes and their impacts." (Section 2.3, VanVuuren et al., 2026)

L435 – Same question, there is no overshoot in this scenario.

Same response as above - to clarify, we added the word "temperature" overshoot (i.e. any scenario where emissions shoot above a specific temperature target, or, if SAI is applied, emissions *would have* overshot if SAI had not been applied). It's a loose term (and hopefully the new IPCC report will have better guidelines on how to use it).

L517/Section 4.3 – This entire section deviates a lot from the other Tier-2 descriptions. As stated above, in my view best would be to move these very vaguely defined simulations out of the actual MIP description (e.g. to an appendix). If the authors choose they want to include these, the description of the mixed-phase cloud thinning experiments should be harmonised with the description of the other Tier-2 simulations.

There is a large difference between this experiment and the others. In the MCT case, there are really different details that need to be reported - because they are not to be found elsewhere for now - in more detail. So we think this is a good platform for these details, that might otherwise be unavailable to modeling teams in time for CMIP7 simulations.

L567 – specifically this section should be incorporated into Section 5.

This has been moved and a reference to the new Section 5.1 is left in Section 4.3.

L660 – "technique is" or "techniques are"

Fixed.

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

CC1

Comment: There is one point of contention I have with the paper, which is the use of the term "preindustrial" in declaring the target of these runs to be "PI+1.5degC". The paper specifies that PI is defined here as the 2020-2039 GMST average. Given that in 2025 temperatures have already warmed by ~1.2degC since the mid 1800's the 2030 average is likely to be even higher than that, so the target is really closer to 2.7+degC above the true pre-industrial. While the definition of "preindustrial" being used here is specified, the reality is that results are likely to be conveyed in a way that this detail is lost. Given that the effects of SRM depend on background climate I think it is important to use terminology that is not misleading. As such, I strongly recommend that these runs are not referred to as targeting "preindustrial + 1.5degC" but rather something like "present-day + 1.5degC".

We thank Dr. Doherty for her comments and feedback. After her comment, we had some time to discuss this with her in person and clarified in depth the misunderstanding. We do clarify now when the "preindustrial + 1.5degC" is first presented that it is intended to be "*a time period as close to ``present day'' as possible*".

Comment: In Section 3.4 it is specified that "MCB is applied by increasing injected sea salt aerosol emissions". While the size of sea salt to be injected is mentioned elsewhere, it would be good to again note here that this sea salt would be of a different size than the naturally emitted sea salt. This is not the same as just increasing sea salt emissions, and "increasing injected sea salt aerosol emissions" could be read to mean just increase the model's natural sea salt emissions. Given the strong size-dependence of cloud responses on aerosol size, increasing natural sea salt emissions will have very different effects than adding sea salt aerosol that is mostly or all in the coarse mode, as is the case for natural sea salt emissions.

Good point! We add 'We recommend that the injected sea salt aerosols have approximate effective dry radii of 30 nm to 50 nm if possible, though this may vary depending on the activation parameterization and natural sea salt aerosol representation, as discussed in section 2.2.'

Comment: lines 341-342: The Rasch et al. (2025) paper should be cited here in and the sentences that follow, as their runs were also focused on low latitude MCB.

Good point! We assume you mean Rasch et al. (2024) 'A protocol for model intercomparison of impacts of MCB'. We have added the citation in L.341-2, and it was already in the following sentence.

Comment: lines 457-458: I found this statement odd, as it implies that the equator-to-pole gradient could only be controlled by deploying MCB at polar latitudes, and this may not be the case. Unless there's a basis for this assumption, I'd suggest rewording to a more general

statement that more research would be needed to determine how well the equator-to-pole gradient could be controlled through different MCB deployment patterns.

Fair point! We rephrase as follows: 'While a 3DOF case aligning with SAI may enable clearer inter-method comparison, more research would be required to determine how well the equator-to-pole gradient could be controlled through different MCB deployment patterns.'

Comment: line 531: In the SH much of the area south of 60S is land, as is some of the area north of 60N. Are clouds over land also perturbed in these runs, or only clouds over sea ice?

Response: We agree that this needs clarification.
As stated in the text, "Perturbations were done over the polar ocean domain (including sea-ice) poleward of 60°N and 60°S." We now clarify in the text that clouds over land were not perturbed, to remain consistent with the perturbation strategy of Villanueva et al. (2022), where it was shown that perturbing marine clouds leads to stronger cooling effects.

Comment: line 532: Regarding the specification of winter as "(Nov, Dec, Jan and Feb)": presumably this is May, Jun, Jul, Aug for the SH winter.

Response: Yes, we thank the reviewer for pointing this out. In the Southern Hemisphere, the perturbed season (winter) corresponds to May, June, July, and August. We have clarified this in the revised manuscript by explicitly referring to winter in each hemisphere.

Comment: Finally I'll note that I disagree with the Anonymous Reviewer #1 that Figures 3, 5 7 and 8 should all be moved to a supplemental section. I agree that Figure 3 is rather confusing as to what it's trying to convey but I would encourage perhaps revising that figure rather than (re-)moving it. Figure 5 seems to be a pretty important visualization of the G6-1.5K-MCB protocol. Figure 7 similarly is a helpful demonstration of the MCT protocol -- though Figure 8, I'll agree, is less essential given that this is not a Tier 1 (or even Tier 2) protocol, so could be omitted or moved to a supplemental section but I don't think it necessarily should be.

Thank you for the feedback! We indeed moved only Fig. 3 and 8 over to the Appendix, to streamline the reading.

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

CC2

Comment: For the model experiments in preparation for the G7-1.5K-MCB ("Inj-seasalt-midlat-SST" and "Inj-seasalt-midlat"), the protocols include the minimum number of years (5 and 20, respectively) but are not clear on a start date. For Inj-seasalt-midlat, a starting year of 2035 is implied based on what is stated in lines 157-160 but is not explicitly clear in Table 1 like it is for the SAI experiments. Should all experiments including the preparation ones branch off and begin MCB emissions in 2035?

Thank you for this. We agree that this ought to be clarified. We will add 2035 as a preferred starting point for inj-seasalt-midlat. We leave the start date for inj-seasalt-midlat-SST open as modellers will likely use the AMIP reference simulations which are available, and starting them in 2035 may be problematic. The main objective of these simulations is to get an estimate of the forcing and temperature responses to inform the injection amounts in G6-1.5K-MCB.

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 lnj-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.