

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 $100\sim\text{Tg}\sim\text{yr}^{-1}$ 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 $\text{PI}+1.5^\circ\text{C}$ temperature target” is provided, in line 332, at $\pm 0.2^\circ\text{C}$. 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* overshoot 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.