

We thank the reviewers for their kind comments and productive suggestions. We have updated the manuscript following their recommendations. In addition, we have made two substantial changes to the nomenclature used in the paper following discussion at a recent GeoMIP meeting and the reviewer comments.

- Following discussions with other GeoMIP modelers, we propose an updated naming convention for the calibration simulations (G4SST- seasalt-midlat -> Inj-seasalt-midlat-SST and G4-seasalt-midlat -> Inj-seasalt-midlat). This change is meant to reduce confusion around the meaning of the GX numbering convention in previous GeoMIP simulations and because there will be a shift to numbering GeoMIP simulations by CMIP generation.
- Following the recommendation of reviewer 2, we have revised the manuscript to use consistent nomenclature for aerosol-cloud interaction forcing ( $F_{aci}$ ) and aerosol-radiation interaction forcing ( $F_{ari}$ ).

#### Reviewer 1:

This is a nice paper. It does a good job of setting up a new protocol for GeoMIP and provides substantial details on how to do it, what to expect, and some preliminary results from other models.

I have one general comment: some more context is needed. What exactly is this paper doing? It leans very heavily on previous papers to describe aspects of the protocol, but it also provides major aspects of the protocol. It comes across as both suggesting that this experiment be included in GeoMIP as well as implying that it was practically commissioned by GeoMIP. I'd appreciate some clarity of purpose and delineating what this paper adds.

- We have expanded the discussion of previous GeoMIP simulations as well as how this paper fits with two previous studies from our research group to provide clarity on how this work fits in with the wider GeoMIP effort (Line 100-112; Line 124-129). Specifically, this paper describes a set of testbed simulations which will be used to inform the next set of GeoMIP7 simulations and was developed through discussions at recent GeoMIP meetings. The need for a new set of MCB protocols for GeoMIP has been recognized for some time, so in a sense it is true this experiment was commissioned by GeoMIP insofar as the authors are participants of the MIP.
- Line 39: its
  - o Corrected
- Line 40: Ahlm et al. (2017) is relevant here. And elsewhere – see below.
  - o We have elaborated the point on MSB
- Lines 73 and 75: I question your claims about impacts analysis. Given the uncertainties in aerosol-cloud interactions in models, as well as the compromises

you make to allow broad model participation (section 2.4), these are still somewhat idealized.

- We have edited this sentence to make it clearer that we are comparing “step” perturbation simulations versus scenario simulations. The common methodologies for impact analyses tend to be designed for scenario-style simulations, which is why we make the claim that scenario simulations are more suitable for impact analyses.
- Lines 146: What about this simulation setup is specific to the midlatitudes? Did you mean to specify a latitude range? (Or refer to Section 2.2.)
  - Added information on emission regions and reference to section 2.2
- Line 154: Can you be more specific about what “early” means? If this is to be a protocol, either a year range or a forcing level would be useful.
  - Added description that matches the 2010 climatology conditions used for CESM2.1 and E3SMv2.0
- Line 159: It would be useful to talk about Ahlm et al. (2017) here. Simulations can vary widely depending upon how (un)sophisticated the model representations are.
  - We have expanded the discussion comparing these results to the simulations in Alterskjaer et al., 2013 and Ahlm et al., 2017 in Section 2.2
- Line 217: Plus at such high rates of injection, the direct effect becomes important.
  - Added
- Line 236: enhancements
  - Corrected
- Line 236: Is it feasible to reproduce the equations here? Then the paper is more self-contained.
  - Equations added
- At the end of section 2.4, you might want to talk about the benefits you would gain if models choose to do both cdnc and sea salt.
  - We have added a few sentences on how performing both CDNC and sea salt perturbations could provide insight into inter-model differences.
- Lines 309-312: I don't think your results support these sentences. It's true for CESM and E3SM (which have a lot of overlap) and not true for UKESM.
  - We have adjusted the wording to reflect the mix of MSB and MCB in UKESM1 and to clarify that we are contrasting the (relatively) strong MCB effect to previous modeling that showed the forcing is almost entirely MSB (specifically Ahlm et al., 2017 and Mahfouz et al., 2023).
- Lines 327-328: I promise you, it does. But it may not be enough to affect your results.
  - We have softened our claim that there is no effect on the variability.
- Line 367: There's a typo in here somewhere.
  - Corrected
- Lines 375: See Kravitz et al. (2016), who talked about a land vs ocean timescale. The point being, there are lots of processes that your system identification simulations are likely failing to pick up. AMOC changes, for example.
  - Added a sentence on longer timescale responses

- Line 476: I strongly advise you to avoid language about policy relevance unless you're explicit about what policies and how it would be relevant. Trust me – that's a hard-won lesson. I think your language about cooperative MCB is more than enough justification.
  - o Point taken! We have removed the claim about policy relevance of the simulations.

Reviewer 2:

**Review of Hirasawa et al., G6-1.5K-MCB: Marine Cloud Brightening Scenario design for the Geoengineering Model Intercomparison Project (GeoMIP) in CESM2.1, E3SMv2.0, and UKESM1.1**

This article's aim is to provide the protocol for the submission of a mid-latitude marine cloud brightening experiment within the GeoMIP project. The target intervention simulation consists in sea salt aerosol injections in the lower boundary layer of 5 midlatitude oceanic regions, with programmatic emission adjustments to keep 21<sup>st</sup> century global mean surface temperatures within  $+1.5\pm 0.2\text{K}$  compared to pre-industrial levels against a SSP2-4.5 scenario background. The required set of simulation also includes two benchmarking simulations aimed at adapting the emissions to each model's radiative sensitivity to sea salt perturbations. The requested output variables are described. The protocol is demonstrated with three test earth system models, CESM2.1, E3SMv2.0, and UKESM1.1. Results with these 3 models show that midlatitude MCB effectively counteracts SSP2-4.5 annual warming with minimal anomalies in surface temperature and precipitation, despite some differences in intra- and inter-model regional and seasonal responses. This article is clear and well-written, and I recommend accepting it after minor reviews.

**General comments:**

- The introduction could benefit from an earlier description of the past and present GeoMIP aerosol experiments, and how the proposed experiments complement the already existing ones: many mentions of G6-1.5K-SAI are already made, but the past G4seasalt MCB experiment is only mentioned briefly on Ln 115, while the mention of the G4cdnc experiment only comes in the conclusion. A sentence about the scope and aim of GeoMIP at the beginning of the introduction could also be informative.
  - We have expanded our discussion of previous GeoMIP MCB simulations to provide clearer context on how this work fits with previous intercomparisons (line 100-112)
- Mid-latitude sea salt injections seem to work better than tropical injections without adverse side effects. At least in the three models studied here and in H2025, but this is limited evidence. How confident are you that other models would reproduce the same behavior?
  - While we cannot be certain other models will reproduce similar behaviour, studies of other midlatitude radiative perturbations indicate that the climate dynamics underlying the response are common across models. Examples include Kang et al., 2018, which tested applying heat flux anomalies in a slab ocean model version of GFDL-AM4, and Liu et al., 2018, which found hemisphere-wide cooling in response to Asian and European sulphate emissions. (Line 623-624)

- Kang, Sarah M., et al. "Contrasting tropical climate response pattern to localized thermal forcing over different ocean basins." *Geophysical Research Letters* 45.22 (2018): 12-544.
- Liu, L., et al. "A PDRMIP multimodel study on the impacts of regional aerosol forcings on global and regional precipitation." *Journal of climate* 31.11 (2018): 4429-4447.
- On Ln 360, it is said that “these simulations are not suitable for estimating the iSSA mass or number emission required for actual deployment but are instead aimed at identifying and understanding the sources of inter-model differences in cloud and climate response to MCB”. In my opinion this is a critical point, and even though this is re-iterated on Ln 532 and 550, I think the aims of this GeoMIP project should be addressed more clearly early in the introduction: 1) evaluate climate impacts of midlatitude MCB as a geoengineering approach, while 2) assessing inter-model variability in interactions between sea-salt aerosols and clouds, but 3) not providing any guidelines for actual deployment due to these uncertainties.
  - We have added a couple sentences on the suitability of ESMs for estimating MCB emission requirements in the introduction (Line 165-167).

#### Specific comments:

- Ln 12: suggestion to add the following precision for clarity: “maintain 21<sup>st</sup> century GMST at 2020-2039 temperatures”
  - Corrected
- Ln 30: “among other considerations” sounds purposely vague, I would remove
  - Corrected
- Ln35:
  - Cite the Twomey effect paper
  - Corrected
  - You should also explain that there are cloud adjustments, especially since you mention the cloud lifetime effect several times in the article.
    - Corrected, including references to papers on different cloud adjustment processes.
- Ln 38/39: References missing here. Which “evidence”/“studies”?
  - We have added a number of references on opportunistic experiment studies that analyse the effect of volcanic eruptions and ship tracks on clouds.
- Ln 39: “it’s its”
  - Corrected
- Ln 54/55: I would include the spatial extents of the 5 regions here, instead of having this information in the caption of Fig.1.
  - We have added the region latitude/longitude ranges

- Ln 56: You could describe the mechanism behind the “stronger positive radiative feedbacks”
  - We have added references that discuss spatial variations in feedbacks.
- Ln 116: Why is the degree of comparison between studies and models “limited”?
  - This is due to differences in the pattern of emissions between the protocols. We have added a sentence to clarify this point.
- Ln 126: Calling this MCB scenario a “deployment scenario” somehow contradicts the later statement (Ln 360) that this initiative should not be considered informative about actual deployment.
  - We have clarified that the scenario is meant to serve as an example of a hypothetical cooperative deployment.
- Ln 144: missing word “based [...] the”
  - Corrected
- Ln 148: I don’t think the NH and SH have been defined before
  - Corrected
- Ln 160-163: This sentence is unclear to me, maybe consider rephrasing
  - We have rewritten this paragraph to make the intention of the discussion clearer
- Ln 162: Did you mean “computeD”?
  - Corrected
- Ln 189: It’s not clear to me which portions of the SH are not covered?
  - There are some gaps between the SH midlatitude regions. We have clarified this discussion.
- Ln 190: I assume the emission is uniform in all of the gridboxes inside of each region, but maybe it’s worth specifying.
  - Added
- Ln 283: Parentheses missing around the citation
  - Corrected
- Ln 293: It sounds like only UKESM1.1 represents the cloud albedo effect, but I think all 3 models do here. As per my previous comment, “cloud lifetime effect” hasn’t been defined before (nor “cloud albedo”, even though I assume you use it as a synonym for the Twomey effect)
  - We added details in the CESM2.1 and E3SMv2.0 descriptions to clarify that all three models include both the cloud lifetime and cloud albedo effect. We have also defined the two terms in the introduction
- Ln 314: I’m not sure what the “initial analysis” refers to. Fig. 1 to 8? Or previous analyses in H2025? Or Stage 1 and 2 simulations only?

- We have rewritten the sentence to clarify that we are using simulations from H2026, which differ slightly from the Inj-seasalt-midlat simulations outlined in this paper.
- Ln 315: I don't understand why some simulations have balanced hemispheric emissions, in disagreement with the proposed protocol. If it's just because you've improved the protocol since the first simulations have been run and you don't want to re-run them for computation resources reasons, I think it's ok to say so
  - We are referring previously completed simulations from H2026. We have updated the discussion to clarify that the simulations we use to calibrate G6-1.5K-MCB are from that previous work and are not the same protocol as we recommend for Inj-seasalt-midlat-SST and Inj-seasalt-midlat here.
- Ln 316: For Fig.1, It bothers me to have “TOA Global Average” or “cloud”, “non-cloud” as labels. I would use less ambiguous y-axis and colorscale labels, for instance: Global ERF (bottom row) ERF\_direct (middle) , ERF\_aerosol-cloud (top).
  - Labels have been updated. We have also made an effort to standardize the nomenclature used throughout the paper to refer to  $F_{aci}$  for the aerosol-cloud interaction and  $F_{ari}$  for the aerosol-radiation interaction, following the convention of Zelinka et al., 2023.
- Ln 356: Fig. 2
  - could you add in the y-axis labels that the plotted variables are *anomalies*? And specify in the caption what is the baseline for computing these anomalies? I assume it's with respect to the 2020-2039 average in the SSP2-4.5 scenario. Same comment for Fig. 3.
    - Labels have been updated
  - Note that using red and green is not color blind-friendly
    - We have adjusted the shade of green and purple to make the lines friendly to red-green colorblindness
- Ln 370: Does the fact that the cloud lifetime effect could modulate the Twomey effect also contribute to the non-linearity?
  - The cloud lifetime effect could potentially modulate the Twomey effect, but the net effect of cloud adjustments on the Twomey effect is unclear. A simple initial assessment suggests that the LWP increase from the Albrecht effect would also increase the Twomey effect susceptibility. However, other processes such as entrainment changes buffer this effect. We have added a sentence discussing the uncertainty in the interaction between the two processes.
- Ln 388: What does a “hydrologic increase vs. decrease” mean?

- We have modified the sentence to clarify that we mean the global mean precipitation
- Ln 393: is there a missing “no” here: “show [...] change”? Because the tendencies look rather undistinguishable from the constant line in Fig. 3d-f
  - fixed
- Ln 403: Is there a missing preposition here? “flux [...] equator”
  - The energy flux equator is a term for the latitude of highest energy flux, which shifts with season and radiative forcing. We have added a short descriptor for the energy flux equator.
- Ln 410: did you mean southward for CESM2.1 and northward for E3SMv2 ? In which case it does not oppose the SSP2-4.5 scenario effect
  - The effect is quite small. However, for CESM2 the G6-1.5K-MCB sees a weaker ITCZ shift than SSP2-4.5, so we infer that MCB drives a northward shift in the ITCZ. We clarified the wording
- Ln 453: The cooling pattern does not oppose the SSP2-4.5 warming pattern at high latitudes though, does it?
  - Here, we are comparing the MCB response (panels 4 g,h,i) compared to the SSP2-4.5 response (panels 4 a,b,c). Since the MCB response sees cooling throughout the high northern latitudes it has a similar response to the GHG effect. We added a note that the Southern Ocean response differs between the MCB and GHG response.
- Ln 462-466: Maybe the hatching in the precipitation plots is too coarse, or I’m not looking at the right place, but I don’t see the pointed out regions as having significant changes (Central Africa for E3SM, or the increases in the subtropical regions). Same comment for the seasonal maps described Ln 467-473,. For example, I see significant changes over Central Africa in JJA for UKESM, but not the changes described on line 473.
  - We have updated the hatching
- Ln 492: Could you remind the reader what the references for the recent ESM studies mentioned here are?
  - We have added references to the recent work that has highlighted the importance of shifting to higher latitude MCB.
- Ln 498: “comparison of the intervention strategies” – here, do you suggest that this protocol’s simulations could be compared to previous simulations from other GeoMIP projects such as G4seasalt and G4cdnc?
  - Modified to clarify that we are discussing comparisons between MCB and stratospheric aerosol injection simulations.

- Ln 509: whether the differences are “subtle” between cloud droplet number and aerosol perturbations could be questioned. I would remove the adverb.
  - Modified following your suggestion
- Ln 516: “with the majority of the forcing occurring via aerosol-cloud interactions”: that is only true in 2 models out of 3. To come back to my general comment about generalization to more models, do you think you can make a general statement like this before other models are added to the comparison?
  - We have corrected the wording to reflect that UKESM1 does not show a majority of forcing is from aerosol-cloud interactions. Considering past results from Ahlm et al., 2017 and Mahfouz et al., 2023 indicate MCB forcing is predominantly the direct aerosol forcing, it is difficult to say what a larger inter-comparison of modern ESMs would show. We are currently finalizing the analysis for an intercomparison with six ESMs that shows the aerosol-cloud effect dominates the overall forcing in four of the six models.
- Ln 519: You could be more specific than “substantially lower” (20 and 2.5 times lower, resp.)
  - We have added specific numerical values to compare against Alterskjaer et al., 2013.