

Response to reviewers of “A new Geoengineering Model Intercomparison Project (GeoMIP) experiment designed for climate and chemistry models”, by Tilmes et al.

We thank Referee #1 and #2 for helpful comments to the paper and address those below. Referee comments are shown in red and italic throughout the response.

Response to Referee #1:

This manuscript describes the G4SSA experiment of GeoMIP. It is a necessary document for the experiment. It is short, clear and deserves to be published in GMD. I have one minor comment. Abstract (L5 and L16): How will an intercomparison of model simulations with prescribed aerosol concentrations from a single model help us understand the composition of the atmosphere? The composition will be the same in every model. It seems like the earlier G4 GeoMIP experiment (described in L5-6) does more to understand the composition of the atmosphere than this new intercomparison. Oh, now I see when reading Section 3 that you're referring to ozone changes (not aerosol changes) when you are referring to composition. Are there many models that could not participate in the G3 and G4 experiments because they don't have aerosols that do have ozone chemistry? Regardless, good to be more specific about what you mean about composition.

We agree with the referee that the abstract was confusing and we changed the following sentences:

Line 5: “A new Geoengineering Model Intercomparison Project (GeoMIP) experiment “G4 specified stratospheric aerosols” (short name: G4SSA) is proposed to investigate the impact of stratospheric aerosol geoengineering on atmospheric **chemistry, dynamics**, climate, and the environment.”

Line 16: “The performance of this experiment using climate and chemistry models in a multi-model comparison framework will allow us to better understand the significance of the impact of geoengineering and its abrupt termination after 50 years in a changing environment.”

We also changed the second paragraph of Section 4:

“The following scientific questions may be addressed with the proposed geoengineering experiment, especially if performed in a multi-model framework like CCMI: What are the impacts of geoengineered stratospheric aerosols and the termination of geoengineering on **chemical composition, dynamics, and climate, in a changing future environment on...**”

To further answer the question of Referee #1, many chemistry-climate models (those who interactively simulate chemistry and therefore ozone) are currently not able to directly inject aerosols and therefore could not perform the proposed G3 and G4 experiments. However, even if they are, the very different setup of those models will not produce the same aerosol distribution. As shown by Pitari et al. (2014), this has a very large impact on the response of geoengineering on climate, as discussed in Section 3: “Furthermore, not many models have the ability to perform prognostic

aerosol experiments including detailed aerosol microphysics.” “The models that performed G3 and G4 experiments derived very different stratospheric aerosol distributions, due to different assumptions of aerosol properties and differences in stratospheric transport and heating rates of the models (Berdahl et al., 2014; Pitari et al., 2014).”

Response to Referee #2:

Having taken on the review I was perplexed after reading the article and thinking about the task given to me: What am I supposed to review? The experimental description at face value? Or, the scientific question, and possible implications that relate to the interpretation of the (expected) results? So I've decided to do both briefly:

Experimental description at face value:

The paper describes an experimental set-up for a simpler so-called G4 experiment that should be easier to implement in many models than an earlier version proposed previously under GeoMIP. Data is provided from a particular model system as a forcing to other models that cannot model the full process chain. There is nothing wrong with this approach, even though it always causes some inconsistencies between atmospheric circulation and forcing distribution. The description of how the data was generated is clear, and the same is true for the "how to use the data". I am sure that many modeling groups could do this experiment, if they choose to do so. In this respect the paper fulfills its purpose, being an adequate description of a proposed numerical experiment that can be repeated by other groups.

Scientific question and interpretation:

This is the area where everything becomes very difficult:

What is the actual science question this simple (but maybe already too complicated experiment) is supposed to answer?

In the paper, the science questions are summarized in Section 4.

Why does the question need many models? Given the accumulated uncertainties, are such experiments pushing the models too far and do we lose credibility by doing experiments we know have too many degrees of freedom and very large error bars (all this are important concerns that relate to the original G4 experiment as well)?

All the questions raised by referee are valid, but can be applied to any model intercomparison project that investigates future projections. Similar to other intercomparison projects, there are advantages and limitations of the proposed experiment, and one can learn a lot from the range of responses in different models. Model agreement on responses will show that the results are robust. As outlined in the paper, we do try to reduce the degrees of freedom from earlier model comparisons of G4, by prescribing an aerosol distribution, and this will reduce the spread of the response.

To clarify this point, we add the following sentence in the introduction after: “We propose a new GeoMIP experiment that uses a uniformly prescribed stratospheric aerosol distribution to address the dependence of the different parameterizations in fully-coupled chemistry and climate models and the impact of future climate change.”

“By constraining the prescribed stratospheric aerosol distribution, we reduce the degrees of freedom from earlier model comparisons of G4, which will reduce the spread of the responses and help identifying key sources of uncertainties in the chemical, dynamical, and radiative response to geoengineering with stratospheric sulphate aerosols.”

Should uncertainties be discussed more, already in the paper that suggests the experiment (I appreciate that the manuscript is already mentioning some issues, but is this enough)?

Uncertainties of current chemistry and climate models are large, since various processes in these models are parameterized, simplified, or not well known. But it is still useful to compare results from different models, even for conditions in the future, where no observations exist. In this way, we learn about uncertainty and variability of different models, acknowledging the limitations of the models used. Here, we are proposing to add a fixed aerosol distribution to the stratosphere and this constrain will help reduce the spread in the models. This is now mentioned in the paper, see comment above. Some uncertainties may be introduced, for example in converting the available pressure levels to the model grid. However besides that, we are adding a fixed forcing to all the models, in addition to prescribed greenhouse gases and other emissions. The main uncertainty of this experiment is in the prescribed aerosol distribution itself, but this cannot be addressed in this study, as outlined in the manuscript.

Why not start with even more basic experiments, like in the early CMIPs? We all appreciate that change in the atmosphere is transient. But if the aim is to diagnose robust features in modelled (circulation-chemistry) change due to (volcanic) aerosol changes, why not start with an even simpler design, e.g. a so-called time-slice experiment. Yes, those experiments are representing a quasi-equilibrium response, but they allow a good statistical evaluation (also in detecting robust feature across many models/experiments).

The idea of the referee to proposing time-slice experiments is certainly useful, and is another potential way the new aerosol distribution could be applied. We have changed Section 2, first paragraph to: “A different baseline scenario could be considered as well, for instance RCP4.5, which is used for the original GeoMIP G3 and G4 experiments and describes a very similar forcing in comparison to the RCP6.0 between 2020 and 2070. Even simpler experiments, like time-slice experiments for different climate and chemistry conditions, could be used to investigate the impact of changes to stratospheric aerosol loading. “

Scientifically, time-slice experiments make certainly sense, and an example of applying different chemistry and climate models to time-slice experiments was given in the Atmospheric Chemistry and Climate Model Intercomparison project (ACCMIP). However, the proposed experiment is meant to be part of the currently performed simulations designed by the Chemistry-Climate Model Initiative (CCMI). Proposed sensitivity experiments in CCMI are based on the REFC2 experiment, ranging from 1960-2100. In this experiment, the atmosphere and ocean are fully coupled and are based on the RCP6.0 scenario. The suggested G4SSA experiment is doing the same thing, and therefore will allow modelers to use the REFC2 experiment as the baseline. Therefore, just from the practical point of view, a transient experiment makes more sense, and we would not have to ask for additional baseline time-slice simulations.

In addition, transient experiments will allow us to answer questions that time-slice experiments cannot answer. For example, transient experiments will allow the investigation of the impact of geoengineering for a certain amount of halogen burden in the stratosphere, since different models will reach a certain burden at different points in time. Correlations between the changes in atmospheric circulation, stratospheric halogen loading, and greenhouse gases will be independent of the type of the experiment. Therefore, a lot can be gained in having CCMI models perform just one set of extra transient simulations, as outlined in detail in Section 3. Adding a fixed aerosol distribution over 50 years will provide results for a good statistical evaluation on the response to dynamics, since the additional forcing is constant.

To summarise: The proposed experiment seems still too complicated to provide a robust insight into model mechanisms (a problem encountered by Pitari et al., 2014) and is too simple as to be realistic (the authors note this problem themselves). However this implies a perception problem: People will think the result could be realistic.

It is certainly important to clarify that this experiment is not realistic, and we believe we have clearly pointed this out in the paper. We also believe that additional insights will be gained from this experiment in comparison to the original G3 and G4 experiments, as discussed by Pitari et al. (2014).

The above considerations lead to my problem: how do I answer the short questions I have to tick when submitting the review?

1) Scientific significance: Nothing new here, and of course not, it is a suggested experiment.

2) Scientific quality: This paper is a suggestion, it cites related work, but it does not provide a particular technical advance.

3) Scientific reproducibility: not applicable (or, alternatively, figure 2 will look different if produced with another model system, but the authors mention this)

We do think the paper provides a technical advance, since the derived distribution

will allow modelers to uniformly perform an experiment that otherwise could not be performed.

In summary, I would like to suggest that the listing of very generic science questions at the end reflects more on what can be expected from such an experimental set-up and what the obvious limitations are.

We are not really sure what the reviewer is suggesting here, but we think that the paper already lists specific science questions that will be addressed with the experiment, as well as its limitations.