Geosci. Model Dev. Discuss., 7, C1637–C1639, 2014
www.geosci-model-dev-discuss.net/7/C1637/2014/

Geoscientific Model Development Discussions Open Access

Interactive comment on "A new Geoengineering Model Intercomparison Project (GeoMIP) experiment designed for climate and chemistry models" *by* S. Tilmes et al.

Anonymous Referee #2

Received and published: 9 September 2014

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 modelling groups could do this experiment, if they choose to do so. In this respect the paper fulfils 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 were everything becomes very difficult: What is the actual science question this simple (but maybe already too complicated experiment) is supposed to answer? 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)? 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)? 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). 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 . . .

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)

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.

Interactive comment on Geosci. Model Dev. Discuss., 7, 5447, 2014.