Response to Reviewer 2

We thank the referee for the positive assessment of our work and the useful comments provided in the review. Below we provide the referee comments in black and our response in red and direct citations additionally in italics.

Review of “Assessing the stratospheric temperature response to volcanic sulfate injection by Mt. Pinatubo: insights from the Interactive Stratospheric Aerosol Model Intercomparison Project” by Perny et al.

The paper presents an evaluation of the stratospheric temperature response to volcanic aerosol forcing in a series of models that (mostly) adhered to the ISA-MIP HErSEA protocol. The models were mostly multi-member ensembles with interactive aerosols and chemistry run in an AMIP-style with forced SSTs and some care was taken to ensure proper QBO phasing out the outset. Of the simulations performed at set were focused on for analysis based on a 7 Tg sulfur injection to 22 km altitude as those simulations had generally the best agreement with the observed temperature anomalies. However, as is clear through the results and discussion the models differed widely in the simulated temperature response, in terms of magnitude, duration, and even altitude. The main conclusion of the paper is that “…by far the most important factor driving both magnitude and spread of the multi-model distribution in temperature response to volcanic aerosol forcing is model choice.” This appears to be correct, but doesn’t provide much clarity.

We understand the reviewer’s mild frustration with this broad conclusion. Given some limitations in experimental design and model outputs, which we discuss in the manuscript, we have kept the conclusions simple, to what can be cleanly observed and quantified. Given the limitations in attribution and separating process effects that remain in our study we provide recommendations for future research directions and considerations for experimental design.

That said, we consider the finding that model choice emerges as the most important factor in quantifying the stratospheric temperature response to volcanic forcing, highly important. We do so as, on the one hand, it highlights the need for more targeted sensitive experiments in MIPs to constrain and attribute model differences, and, on the other, argues for caution in interpreting outcomes of single model studies with focus on volcanic influence. The latter cautious note can be expanded to studies exploring the effects of stratospheric-aerosol injection as a climate intervention analogue.

Several suggestions on improved intercomparisons are put forward.

The paper is well organized and generally well written. My (hopefully) constructive suggestion is it would benefit from presenting more material on the particle size and optical properties. The Quaglia et al. (2023) paper seems to be doing a lot of heavy lifting as it is appealed to at all points where microphysics is brought up. But I have to think the different temperature responses in Figure 3 have some strong correlation with particle size and optical properties. I note that the Quaglia paper did not have MIROC or CARMA results, so it would be fair to in some way restate the microphysics.

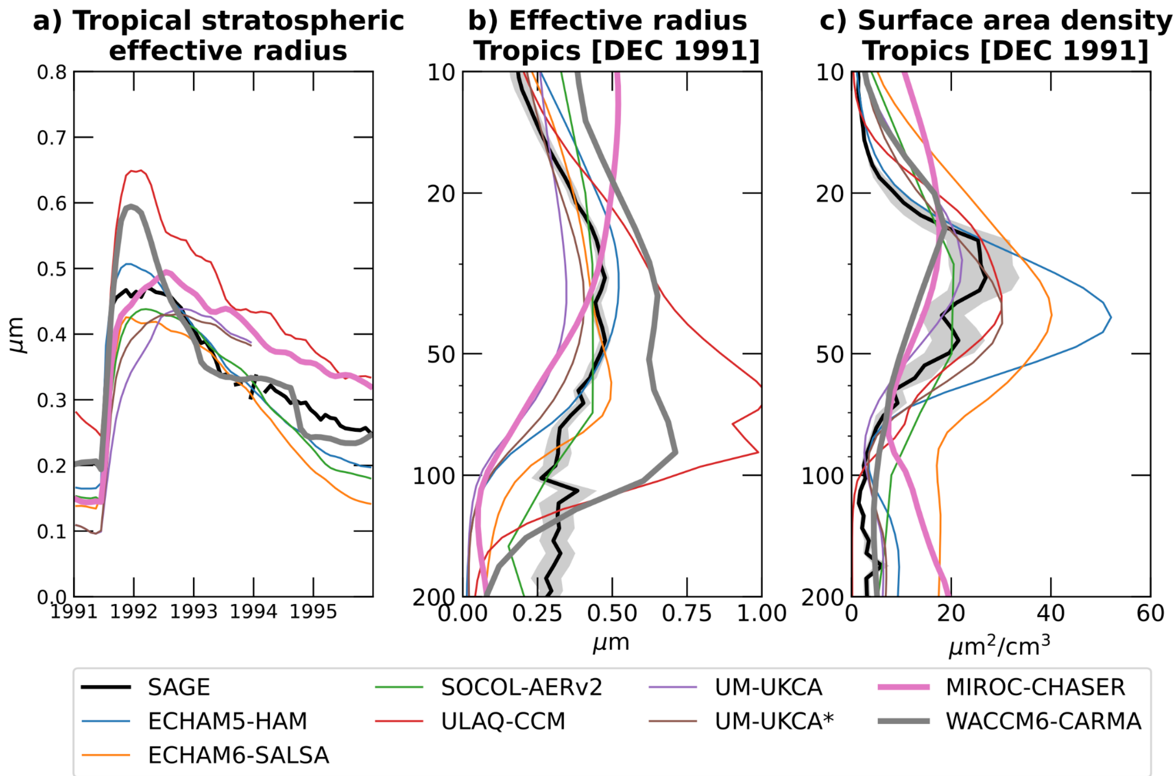


Figure R1: Vertical profile of effective radius a) and time evolution of stratospheric effective radius b) and surface area density c) in the tropics for Med-22km. Model results are compared to SAGE II following Quaglia et al., 2023.

We thank the reviewer for the overall positive assessment. Following the reviewer's suggestion, we analysed the time evolution of effective radius for the complete set of ISA-MIP models in extension of Quaglia et al., 2023. Figure R1 shows that the tropical effective radius increases following the volcanic eruption and remains substantially elevated over the same time period as a pronounced temperature signal is found. However, the effective radius remains somewhat elevated (while slowly declining) over a longer period than a significant temperature anomaly is found. In contrast, the time evolution of SO4 concentrations agrees much better with the temperature anomalies and thus we prefer to use SO4 concentration as tracer.

We note in passing that our additional analysis of effective radius expanding to the full set of models did not ease interpretations. In the initial draft we hypothesized that the lower altitude warming signal in ULAQ-CCM may stem from a larger effective radius. The updated analysis including WACCM6-CARMA and MIROC-CHASER shows that also WACCM6-CARMA has a larger effective radius at lower altitudes compared to the other models but no warming shift to these lower levels is found. Further model outputs regarding size distribution are not available and would be needed to further constrain.

Can you go beyond effective radius? Could you make a figure analogous to Figure 3 but showing some quantity more directly connected to the forcing the sulfate concentration? How about extinction? Or even mass-extinction efficiency?

Unfortunately, the needed output to look into a more detailed analysis of the forcing was not a mandatory deliverable within the ISA-MIP framework. Nevertheless, ECHAM6-SALSA, UM-UKCA, ULAQ-CCM, MIROC-CHASER and WACCM6-CARMA have provided some limited information, i.e. extinction coefficients at 525/550 nm and 1020 nm which are displayed in Fig. R2. No consistent pattern is visible when comparing the extinction coefficients at the two wavelengths, i.e. depending on the model the extinction coefficients were higher or lower at the short wavelength or vice versa. Absorption and scattering coefficients as well as information regarding the size distribution would be required to allow a more detailed analysis. However, as intensity, location and duration of the extinction signal agree well with SO4 mass concentrations across models, this points to the fact that it is the SO4 mass concentration driving the forcing.

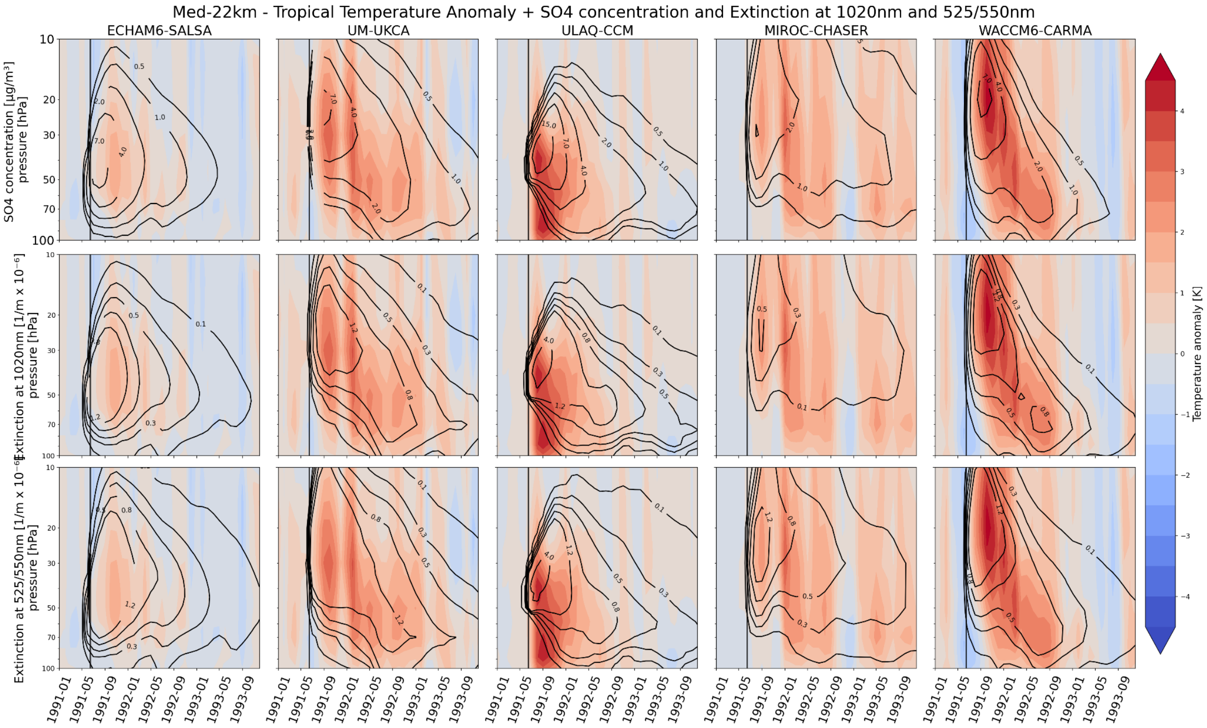


Figure R2: Time evolution of the vertical temperature anomalies in the tropics (20° S - 20° N) in experiment Med-22km for individual models. In the upper panel SO4 concentrations are shown as contours, the middle and lower panels show extinction coefficients at wavelengths 1020 nm and 525/550 nm respectively. The thick black vertical line represents the Mt. Pinatubo eruption time.

What are the optical property assumptions of the models? Can any be rejected as incompatible with what we know about sulfate aerosols? Do the models treat sedimentation sufficiently the same (i.e., given the same size particle it falls out at the same rate)? I think the paper will benefit from some further presentation in this space, so I recommend it for revision.

From the analysis presented above aerosol mass concentration and extinction appear consistent in those models where output has been available. Certainly, each model has its own aerosol scheme which might have its strengths and weaknesses. A structured comparison of the treatment of scattering, absorption, sedimentation and aerosol size distributions would certainly be of interest and valuable. ISA-MIP has been designed specifically for that purpose, but whilst such structured comparisons could be performed from the data on the DKRZ archive, this is outside the scope of the present study. We highlight in the Discussion section of our manuscript that further MIPs focusing on these aspects are recommended for future research.

Line 110 - I'm confused here about EMAC vs. EMAC*. You state a few lines earlier that you are only using some of the EMAC output because it did not adhere to the protocol. What is the distinction you are trying to make with EMAC*? Suggest you just call it EMAC.

While the asterisk might appear confusing, we wish to keep it in Figure 2 to highlight that EMAC is methodologically different to the other models. We explain what the asterisk means, for consistency, also in the figure caption now.

Lines 194-195 - there is a "(?)" where maybe a citation is intended?

Thank you! Missing references were added.

Line 273 - It would seem to me that these initial conditions would be in different phases of the QBO. I didn't really understand the "reset" to the QBO phase at the time of the injection. How can that be accomplished, and how do you reconcile with the WACCM-CARMA state?

We apologize that our wording regarding the handling of QBO was not clear enough. Following the experimental protocol, all models matched the observed post-Pinatubo QBO transition, either by nudging or from re-initialising from a prior control run. WACCM6-CARMA initialised with different conditions for 1979 (1979, 1980, 1981) to create the three ensemble members, from 1980 onwards all of these members nudged the tropical zonal winds to idealized observations between 4 and 86 hPa and 10°S to 10°N in accordance to Davis et al.,2023 and Matthes et al., 2010.

We revised the sections in the manuscript accordingly.

Line 437 - “Latest by November 1991…” reads oddly. Maybe omit “Latest” and just say “By November 1991…”

We changed the sentence following the reviewer's suggestions. We additionally modified another sentence with a similar wording from:

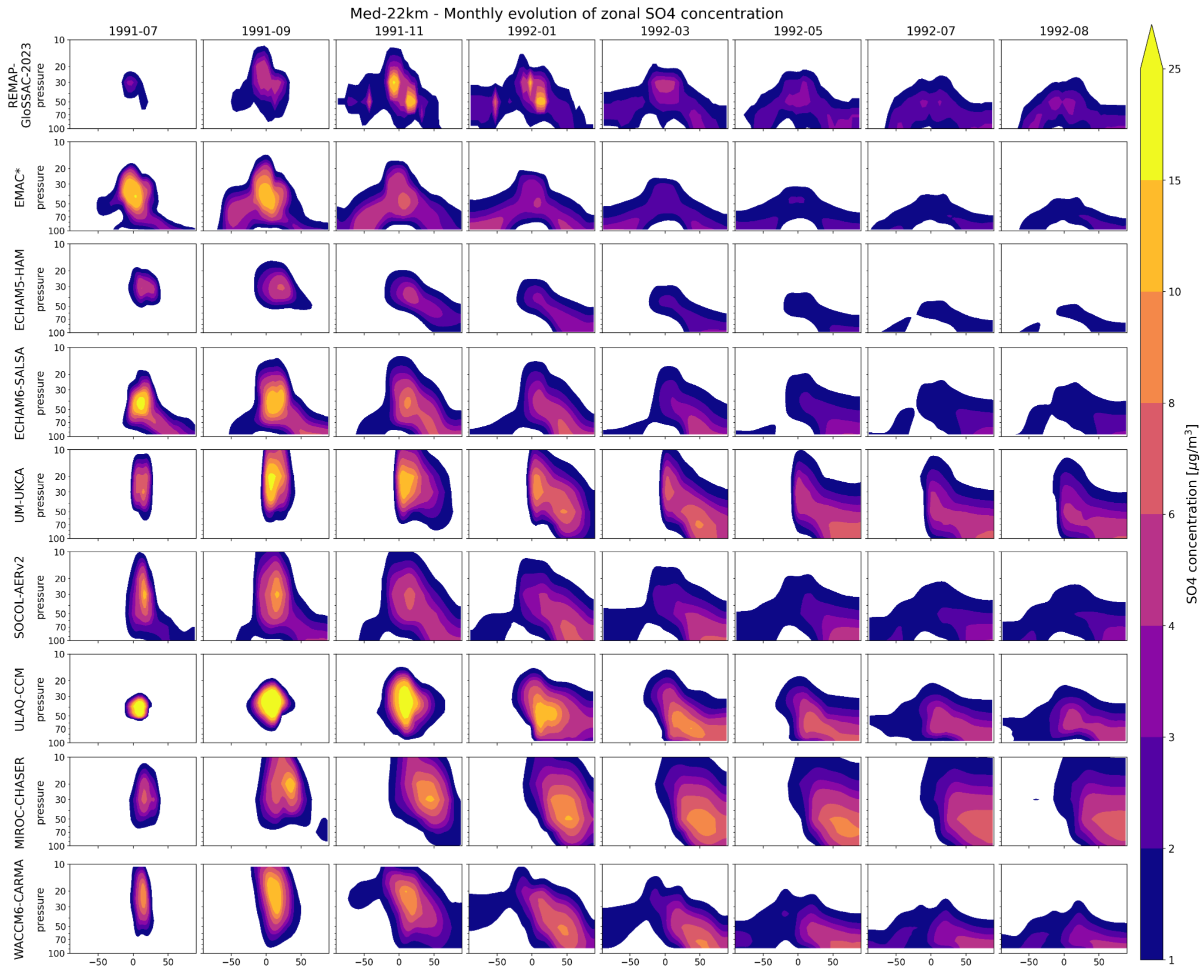
“*For all ISA-MIP model realizations, sedimentation of SO4 concentrations starts at the latest in November 1991, and downward propagation of the SO4 plume emerges between late winter and early spring 1992*.”

to:

“*For all ISA-MIP model realizations a net descent of SO4 concentrations begins in November 1991, likely reflecting the increased effective radius, and downward propagation of the SO4plume emerges between late winter and early spring 1992.*”

Line 448 - It is not clear from Figure 2 that there is good agreement between the models and GLOSSAC with respect to placement of highest aerosol loading in the Arctic. The models tend that way, but the GLOSSAC inferred concentrations appear pretty uniform at all longitudes. Maybe another contour interval would help if that’s really the case?

We thank the reviewer for raising this. We have updated the figure adding an additional contour interval (see Fig. R3).



R3: Vertical evolution of zonal mean SO4 concentrations in experiment Med-22km for individual models, from July 1991 to August 1992 (bi-monthly; columns) in comparison with REMAP-GloSSAC-2023 observations (first row).

Subsequently we also modified the accompanying text:

"...While this is in broad agreement with REMAP-GloSSAC-2023 observational composite retrievals, the latter shows stronger transport to the southern hemisphere than the models. Despite the northern hemispheric transport preference of the models, both low-resolution models such as SOCOL-AERv2 and intermediate resolution models such as ECHAM6-SALSA and EMAC, perform well in comparison to REMAP-GloSSAC-2023 observations regarding the magnitude and transport patterns of SO4 concentrations..."

Line 484 - "descent" instead of "descend"

Corrected.

Line 505 - "…enables us to assess…" would read better

Done.

Figure A5 - Caption doesn't make much sense to me. First line: how is this "models in the tropics" when it shows all latitudes? And I don't see the vertical lines indicated.

Thank you for noticing this. Corrected.

For completeness we note that small additional changes were made for the revised manuscript, i.e. sigma pressure level conversion of WACCM-CARMA was corrected using the US standard atmosphere, and an adjustment to reference latitudes of reanalyses in Fig. 5.