Response to Reviewer 1

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.

This paper is an impressive amount of work resulting in a really nice study and a pleasant read. Particularly nice is the massive comparison between models, reanalysis, and observational products. The paper is quite well done. I really like Section 3.1 where you justify the selection of a core experiment. I only have a few minor comments, mostly about clarifying details and improving replicability.

Thank you. We hope our comments below clarify your open points.

The abstract basically says, "different models have similarities and differences." Can you be more specific?

Given the limitations in the Abstract word count and the multitude of differences between models, this would require to single out specifics. In the 'Summary and Discussion' section, we explain that the differences between the models are greater than the differences within individual models across the various experiments, and that the sensitivity of the stratospheric response depends heavily on the choice of model – this should appear most prominent given the word limit available.

Lines 47-54: While I like talking about SAI, this paragraph just sort of seems like a non sequitur. Can you better tie it into the points you're making?

We have rearranged the paragraphs in the introduction following the reviewer's comment.

Section 2.1: You need more details in your experiment descriptions. What year(s) do you use to create the SSTs? Why are you specifying the injection amounts as Tg S if you're only injecting SO2? Are you injecting throughout the entire altitude range equally, and how do you handle models with different vertical resolutions?

The HErSEA experiment protocols (Timmreck et al., 2018) involve AMIP-type time-varying SSTs and used the HadISST climatology between 1991 - 1995 (Rayner et al .2003). We specify this now explicitly in the manuscript.

The HErSEA protocol also specified injection amounts in TgS, thus we are using TgS for injection amounts throughout our manuscript. TgS converts to SO2 by a factor of 2.

The injection is specified as a uniform vertical profile of emission.

Line 116: What do you mean by reinitialize? To a particular state? And where do you get that state?

We apologize that our wording 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. This ensures a consistent QBO state across different models, minimizing the effect of model-specific QBO dynamics on model spread on aerosols.

We have revised the manuscript accordingly to:

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. This ensures a consistent QBO state across different models, minimizing the effect of model-specific QBO dynamics on model spread on aerosols.

Line 145: Was the diffusion changed permanently starting on January 1? Or was it perturbed for a day and then changed back? And if permanently, how might this have affected your results?

The diffusion coefficient was changed permanently. This leads to a slight perturbation for an individual ensemble member, but is not causing broad divergence or even opposing directions within the ensemble. Thus, it does not affect the overall results.

Lines 194-195: Some missing references. Also, is ozone prescribed or interactive? You talked about this in the previous subsections.

Thank you for noticing this, the references are updated. SOCOL-AERv2 calculates ozone interactively. We adapted the text to:

Chemistry is represented in the module MEZON that calculates the interactions of 89 species (e.g. oxygen, hydrogen, nitrogen,...) using 234 reactions, including the detailed sulfur cycle (Brodowsky et al., 2024) and heterogeneous reactions on sulfuric, as well as polar stratospheric cloud (PSC) aerosol particles.

Lines 222-224: So this is a bulk scheme?

No, ULAQ-CCM uses a sectional microphysics scheme. Information is passed to the radiative transfer module for the whole size distribution, not a bulk (i.e. Reff) value.

We modified the respective paragraph to:

"Aerosol optical properties derived from the ULAQ-CCM sectional aerosol module are passed daily to the radiative transfer module, with appropriate wavelength-dependent values of Q-ext, g, and single scattering albedo, given the calculated size distribution of the particles (Pitari et al., 2016a)"

Table 1: It might be useful if you had a column talking about the aerosol scheme, at least talking about bulk, modal, sectional, etc.

We have added a corresponding column to table 1.

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.