

Interactive author comment on “Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora Interactive Stratospheric Aerosol ensemble” by Margot Clyne et al.

Answers to reviewers on the ACPD paper (acp-2020-883): Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora Interactive Stratospheric Aerosol ensemble

We thank the three reviewers: Daniele Vioni (RC1), Anonymous (RC2), and Peter Colarco (RC3) for their helpful comments. We carefully considered the recommendations and made some changes to the manuscript. The questions and comments from the reviewers are in bold, our answers are in italic, and changes in the text are in blue. The line numbers in the referee comments have been updated to correspond to the revised manuscript.

After submitting the initial manuscript, we realized that the effective radius values we were using for UM-UKCA were for the dry effective radius instead of the wet effective radius. The effective radius values from that model have since been recalculated (to wet) and updated in the revised manuscript. We have verified that this problem was unique to the UM-UKCA data, and it has since been fixed in the text and figures (fig 3, 4, 6, 7, and 8 (previously 9)). This only changes the reported values about effective radius results. The internal model calculation of AOD is not involved here. The other changes to the manuscript are from comments raised by reviewers, which we list below.

Answers to reviewer 1

Thank you for your thorough comments. Your comments especially helped us improve the discussions about stratospheric dynamics.

Major comments:

Introduction: the introduction dives straight into discussing the VolMIP and ISA-MIP protocols, but gives little to no context. I understand that, for experts in the field, it’s irrelevant, but since it’s a good introductory paper for the issue of model differences in volcanic eruption simulations, I would suggest some paragraph explaining the motivation for running the experiment, and some further description for the Tambora event.

Thank you for the input. This was a popular suggestion. We have added to the beginning of the introduction: “Volcanic Eruptions impact climate by cooling temperatures (Robock, 2000). They inject sulfur dioxide gas (SO₂) into the atmosphere. This sulfur dioxide converts to sulfuric acid, and then to sulfate aerosol. The sulfate aerosol scatters sunlight, and causes an increase in aerosol optical depth, which is a key volcanic forcing parameter. The volcanic forcing cools Earth’s temperature. Depending on the size of the volcano, this may only have a small regional effect, or, for large explosive eruptions, the effect can be global. Interactive stratospheric aerosol models (or

ISA's) are used to calculate the aerosol optical depth. Volcanic eruptions are simulated in these ISA models by injecting SO₂ directly into the atmosphere. Basic information is needed about the injected SO₂, namely the mass, time, and altitude at which to inject it. There is uncertainty about the true values of these basic volcanic injection parameters due to limited availability of observational data for each eruption. Proxy estimates and model studies are also used to better constrain these input values. The variety in plausible injection parameters for a given eruption complicates volcano model intercomparison projects. Thus, the VolMIP-Tambora ISA experiment was created to assess intermodel differences by using a consistent set of volcanic injection parameters across models. The Tambora eruption was chosen as an example because it was large enough to have significantly altered the climate, but had no observations of the volcanic cloud so that modelers did not know the answer in advance."

243-246: I'm not really sure I agree with this conclusion (or maybe I misunderstood it). This is akin to saying that 15% of the SO₂ is not converted into stratospheric sulfate aerosol and is instead removed directly. Considering the height of injection, I find this a bit hard to believe, nor do I think the authors have proven this convincingly (and it would be hard to prove, since it would require looking carefully at the 3D strat-trop exchange of SO₂, or looking at direct SO₂ deposition). Notwithstanding the differences between impulsive injections like this case and sustained injections as in the case of geoengineering simulations, it is not even very consistent with previous findings (see Vioni et al. (2018) for a CCM and a CTM, and Vioni et al. (2020) for CESM(WACCM)) where all SO₂ was converted.

Overall, I just don't think the phrase is true, and the fact that (no) "more than 4 TgS is removed before the peak value of global stratospheric sulfate is reached" does not mean that, but rather that some SO₄ is removed before the peak (which is to be expected). Unless the authors can prove otherwise.

We agree with the arguments put forth by the reviewer, and have decided to remove the 2 sentences. Thank you very much. *(The last two sentences of paragraph 3 in section 3.2.1 were removed. The removed sentences were: "Although they vary widely on the timing and duration of elevated global stratospheric sulfate burden, none of the models predict that any more than 4 TgS is removed before the peak value of global stratospheric sulfate is reached. In other words, models find that more than 85% of the emitted sulfur in the volcanic SO₂ is converted into stratospheric sulfate aerosol.")*

256: A similar note to what I said above. By "Sulfur" what do the authors mean? If they mean aerosols, then it makes sense. If they mean only or also SO₂ (as the conclusion to the paragraph above implies) than I don't see any proof of that, but it is an important distinction!

Thank you for catching this! The corrected text now reads: "Unlike the other models, the mass of global stratospheric sulfur in LMDZ-S3A is not stable within the first several months following the injection of SO₂; sulfate sulfur is crossing from the stratosphere into the troposphere, where it is quickly removed."

This should make more sense now in context. The (unchanged) sentence that follows is: "The sum of the volcanic sulfur species stratospheric burden (SO₂ + H₂SO₄ + SO₄) exceeds 29 TgS for

the first two months in the LMDZ-S3A band injection experiment (April and May 1815), but then quickly drops (Figure S1).”

265: The phrase “CESM-WACCM and UM-UKCA produce the smallest $Reff$, with values never exceeding $0.5 \mu\text{m}$ ” is a bit in contrast with what previously stated in line 230, where the cause for the increased burden in CESM and UM-UKCA is attributed to the better representation of mid- atmospheric dynamics (assuming that means that the stronger upwelling and confinement in the tropical stratosphere are “correct”). But here it is stated that the two models have the lowest radii, which is a much more straightforward explanation for why the burden is higher. Smaller particles result in less gravitational settling, thus increasing the lifetime (see, for instance, Vioni et al. (2018) Fig. 5 and 7). If the confinement is stronger, particles tend to grow more and thus settle quicker. So overall, the increased burden is much more likely due to these microphysical differences than to dynamical ones (as is also discussed below in Section 4.1).

Agreed. We moved the previous text about the model high tops to section 4.3.2 where it makes more sense, and added further explanation. That part of section 4.3.2 now reads: “With the exception of EVA, part of why the models differ in stratospheric meridional circulation patterns in this study may be due to their different approaches in the treatment of the QBO (Table 2), and/or to differences in transport vertical diffusion associated with the various vertical model resolutions and amount number of vertical levels in the volcanic model (Table 1). For example, CESM-WACCM has the highest top of all of the models, well above the mesopause, allowing the most complete representation of the middle atmosphere circulation. UM-UKCA is the only other model to include the entire mesosphere. Both models have a long stratospheric lifetime, which we attribute mainly to their having smaller particles. As explained previously, having smaller particle sizes lowers the removal rate, and contributes to a longer lasting large burden. In addition, a small return cycle involving the mesosphere and stratosphere occurs in which sulfuric acid evaporates above about 3 hPa or 40 km, and then regenerates SO_2 at high altitude. When the air descends the sulfuric acid vapor can reform particles and the SO_2 can create additional sulfuric acid forming new particles at high latitudes. Simulation of this process could be affected by vertical model grids. The impact of this cycle to the stratospheric sulfate burden and AOD, and how it varies across models, is not quantified in this study but is expected to be minor.”

468: I find it a bit puzzling that there is no mention in the list of “Other model uncertainties” of the possible differences in self-lofting resulting from the stratospheric heating produce by the sulfate aerosols. Together with large-scale dynamical differences discussed in 4.3.2, I would assume that such a large burden would result in large heating rates and thus in an increase in w^* compared to unperturbed conditions. Are pre-eruption residual vertical velocities in the tropical pipe comparable between models? And what about heating rates/perturbed w^* ? It could be that not all models have calculated the transformed eulerian means quantities, which might make this comparison hard to do. If they did (or even if just a few of the models, like CESM and UM-UKCA, for which the authors can’t find a clear explanation for the differences) then it would be extremely interesting to show a profile of w^* in the tropical pipe. If not, it would be good to at least compare the temperature

differences (heating rates are probably more problematic than w^*) in the stratosphere and discuss the possibility that the differences are important.

This information is included in Sect. 4.3.2 in the same paragraph as the previous comment, directly following the new text. The text there reads: “All of the models except for EVA and LMDZ-S3A include aerosol influence on radiation, which warms the aerosol layer, which forces self-lofting and latitudinal spread (Young et al., 1994; Timmreck and Graf, 2006). Meridional transport may also simply be faster or slower depending on the internal model dynamics. For example, outside of this study, ECHAM5, the GCM used by both SOCOL-AER and MAECHAM5-HAM, has been documented to have a too-fast vertical ascent and/or mixing in the lower tropical stratosphere (Stenke et al., 2013) and too-fast poleward transport in the stratosphere from the tropics (Oman et al., 2006). Also, Niemeier et al. (2020) show that in the lower tropical stratosphere around 50 hPa, WACCM has 70% larger residual vertical velocity than ECHAM5. Simulations with ECHAM5 and WACCM in Niemeier et al. (2020) where the QBO is internally generated show that stronger residual vertical velocity strengths and subsequent vertical advection strengths can lead to different tropical sulfate altitudes, concentrations, and meridional stratospheric transport.”

Additionally, we added to the very end of the last paragraph in Sect. 4.3.2. that: “The model stratospheric circulation discrepancies, possibly arising from different treatments of the QBO, model grid resolutions, model tops, vertical advection strengths etc. may have the potential to impact the $Reff$ via changes in tropical confinement and concentration of the aerosols, but we do not have the ability to investigate this with the current VolMIP-Tambora ISA ensemble due to the larger conflicting simplifying parameterizations identified in this study which dominate AOD disagreement.”

We do not have the data required available to make these plots for comparison. However, we do not have anything to compare it with (a difficulty of an idealized experiment). We hope that the discussion in the text about these topics is sufficient for this study. This would certainly be worth looking further into in a model intercomparison project where possible consequences of differences in dynamics are not overshadowed by the other parameterizations we’ve identified here that dominate the AOD disagreement.

495 and following: I’m sorry to say this entire paragraph is utterly confusing to me. I will try to make sense of it, but I suggest a thorough checking by the authors. First, I assume the authors mean “southern” hemisphere. Second, I would say that the injection of SO_2 happens inside the tropical pipe, it doesn’t “go” there. But then, right after, the authors say that the aerosols move “towards the winter pole” which I honestly don’t know what it’s supposed to mean. And in what sense it “drains” the tropical pipe? I don’t understand the following phrase either, about the stratosphere “depth”. The tropopause is lower, sure, but I don’t get why that would be the explanation for why the AOD moves poleward. And then, see following comment.

We appreciate you admitting that this paragraph was confusing. We have made some changes in the text to clear this up. The beginning of the now reads as follows (and is continued in the next comment below in RC1 474):

“The VolMIP eruption goes ~~injected~~ **SO₂ directly** into the tropical pipe, which is a region ~~that has little poleward transport in the summer hemisphere.~~ **in which material is confined and prevented from poleward transport into the summer hemisphere.** Within the stratosphere aerosols are transported **during the Fall and Spring** meridionally towards the winter pole, which **then** drains the tropical pipe. ~~Additionally,~~ **As material is transported poleward,** the stratospheric optical depth maxima move poleward for the same reason that ozone columns are highest poleward. That is, the stratosphere is twice as deep at mid and high latitudes **and has less area so column quantities increase.**”

500: “Aerosols are removed from the high latitudes by tropopause folding” this is a very strong assertion, not proven by the authors for these simulations nor backed up by the literature, as far as I’m aware. Happy to be proven wrong. In general, both for ozone and for aerosols, tropopause folding is *one of the* possible mechanisms, surely not the only one, and I’m not even sure about how predominant it is compared to strat-trop exchange and, for aerosols, gravitational settling. The references I know of (again, mostly for ozone) do not indicate it as the predominant form of contribution to tropospheric air from the tropopause (see for instance Oltmans et al. (1989); Holton et al. (1995); Wimmers et al. (2003); Sprenger et al. 2003)

*We agree that this statement needed further information. We have added to the text (in a continuation of the paragraph in the comment above): “Aerosols are removed from the high latitudes by tropopause folding **and other stratosphere-troposphere exchange mechanisms.** Generally, aerosols smaller than about 0.5 μm radius are too small to fall out of the stratosphere before they are removed by dynamics. However, larger particles will fall out and thus have a shorter residence time than smaller particles. As transport next occurs towards the other pole during its winter, the tropical pipe is again depleted.”*

Figure D1-D2 and Appendix D: these figures are very complicated, but that is not the main problem to me. They could be really important, but it feels like the rationale behind using SOM is not explained satisfactorily. Why shouldn’t you do a similar analyses using a Dynamical Mode Decomposition instead, or a simple EOF? SOM lets the algorithm decide what the bases are, without a proper physical meaning. These bases are not necessarily orthogonal, so for instance it’s not immediate what the differences between the first and the second pattern are. The authors should explain a bit more (feel free to put it in the supplementary) what SOM is and why it was chosen. This is not something everybody would be familiar with in this field, so it might help people understand what the authors mean.

We added more explanation about how the SOMs were created, what the representative patterns had come from, etc. to a new section of the supplementary info (Sect. S2), as was also requested by reviewer 2.

Minor comments:

61: The ISA acronym needs to be explained here.

Done. The acronym explanation was added to the new first introduction paragraph.

63: I don't really follow what "to effect" means in this context, so the phrase seems a bit obscure to me.

The phrase "...to effect a common volcanic forcing in simulations..." has been replaced with "...to impose a common volcanic forcing in simulations..."

65: there are word missing. I guess the author wanted to say "since the experiment is designed..."

You are correct. This change has now been made.

69: this part is also a bit hard to understand, and very long. I am only able to understand what HErSEA is because I already know about it, but that "but for the HErSEA experiment" is otherwise a bit obscure (also, the acronym is not explained). I would suggest a stop after 20th century, and then a sequent phrase saying: "In most ISA-MIP experiments, the models run..."

Thank you for the suggestion. This change has now been made. The section now reads: "The VolMIP-Tambora ISA ensemble experiment is similar in approach to the ongoing Interactive Stratospheric Aerosol Model Intercomparison Project's (ISA-MIP)'s Historical Eruptions SO₂ Emission Assessment (HErSEA) experiment (Timmreck et al., 2018), which intercompares model simulations of the three largest major eruptions of the 20th century. In most ISA-MIP experiments, but for the HErSEA experiment, the models run different realizations of the volcanic aerosol cloud based on a small number of alternative specified SO₂ emission and injection heights for each eruption. In the VolMIP-Tambora ISA ensemble experiment, climatological variables and injection parameters were prescribed under a coordinated experimental protocol embedding historical information about the 1815 Mt. Tambora eruption to reduce intermodel differences due to from initial conditions."

71: I'd remove the word "ensemble" here

Done. (See previous response for context).

73: "dependence" from initial conditions, or "differences due to" initial conditions

Thank you. We changed this to "differences due to". (See Line 55 response above for context).

74: the ISA-MIP protocol prescribed for Pinatubo in Timmreck et al. (2018) a range of SO₂ emissions from 10 to 20 Tg-SO₂, with a "medium" injection of 14 Tg. The numbers given here result in a low of 14 and a high of 23 Tg. I'd suggest resolving this inconsistency.

Thank you for mentioning the inconsistency. We changed the text to: “The experimental protocol designated an emission of 60 Tg of sulfur dioxide (SO₂) into the stratosphere. For comparison, the emission estimate for the 1991 Mt. Pinatubo eruption used in the ISA-MIP H_{ER}SEA experiment is 10 to 20 Tg of SO₂, approximately 2.6 to 4.3 times the emission estimates for the 1991 Mt. Pinatubo eruption (Carn et al., 2016).”

138: proportional “to”

done

226: I’m not really sure what is intended by “elevated”, used here and elsewhere (at least three times in this paragraph). Is it intended as “in the stratosphere”? Or as “high”, “large”?

It is intended to mean “elevated” as in “larger quantities than normal”. Thank you for raising this discrepancy, and providing the opportunity for us to better explain the results throughout this paragraph.

*In this line, we replaced “shortest duration of elevated sulfate” with “shortest [perturbation of sulfate](#)”, and clarified this by adding information to the text about peaking earlier than the other models. The sentence now reads: “In Table 3 we see that MAECHAM5-HAM produces the shortest [perturbation of duration of elevated sulfate](#) in the stratosphere, with an *e*-folding time of 8 months for the point injection and 10 months for the band injection [after peaking early \(in August 1815\)](#).”*

We also changed the text a few sentences later to: “Figure 2 indicates that [large global stratospheric burden values of the perturbed volcanic sulfate are more stable within UM-UKCA and CESM-WACCM than in the other models](#). ~~UM-UKCA and CESM-WACCM have more stabilized elevated sulfate burdens.~~”

*Later in this same paragraph, we replaced “elevated” with “increased”, to read: “In addition to taking the longest time to reach its peak sulfate burden value, CESM-WACCM has the longest duration of [increased elevated sulfate burden](#), with an *e*-folding time twice that of MAECHAM5-HAM point.”*

At the end of the paragraph, we changed the text to remove the discrepancy of the word “elevated” via: “No sulfate deposition occurs in CESM-WACCM until 1816, when 35% of global sulfate deposition occurs followed by 46% in 1817 and 17% in 1818, with deposition [still occurring above background levels](#) ~~levels still elevated~~ at the end of the simulation (Marshall et al., 2018).”

We also changed “elevated” to “perturbed” in the second paragraph of the Conclusions section to read: “The rise and decay of sulfate aerosol burden in the stratosphere controls the timing of the onset and duration of [perturbed AOD](#).”

396: I don’t think “in number” is necessary.

Done. This was removed.

431: The version used in Mills et al. (2017) is the same as the one used here, so for consistency it should be called the same (CESM-WACCM)

done

506: please see comments about figure D1.

done

550: This is a very interesting observation, quite in line with a similar effect observed in Visioni et al. (2018).

Very interesting! Thank you for pointing this out. We have added the reference and this connection to the text.

651: “physical and chemical processes”

done

669: “aerosol layer” or “aerosols are spread”. “Rather” than a more...

done

Figure 4: hard to read, should be enlarged

Done. We re-stacked the subplots to make the figure fit.

Figure 7: it would be better if the scales were modified so as to include both lines in all panels. Otherwise the peak for the dashed line can't be evaluated.

*Extending the scale of the y-axis made the subplots and dashed vs solid lines visually harder to compare, so we left the figure as is and instead made a note in the caption. We added to the figure caption: **The peak values for the dashed line which are therefore slightly cut off from view by the y-axis of the subplots for the sulfate burden and effective radius are (b) 1.03 and (c) 1.16.***

References from RC1:

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L. (1995), Stratosphere-troposphere exchange, *Rev. Geophys.*, 33(4), 403– 439, doi:[10.1029/95RG02097](https://doi.org/10.1029/95RG02097).

Oltmans, S.J., Raatz, W.E. & Komhyr, W.D. On the transfer of stratospheric ozone into the troposphere near the north pole. *J Atmos Chem* 9, 245–253 (1989).
<https://doi.org/10.1007/BF00052835>

Sprenger, M., Croci Maspoli, M., and Wernli, H. (2003), Tropopause folds and cross-tropopause exchange: A global investigation based upon ECMWF analyses for the time period March 2000 to February 2001, *J. Geophys. Res.*, 108, 8518, doi:[10.1029/2002JD002587](https://doi.org/10.1029/2002JD002587), D12.

Visioni, D., Pitari, G., Tuccella, P., and Curci, G.: Sulfur deposition changes under sulfate geoengineering conditions: quasi-biennial oscillation effects on the transport and lifetime of stratospheric aerosols, *Atmos. Chem. Phys.*, 18, 2787–2808, <https://doi.org/10.5194/acp-18-2787-2018>, 2018.

Visioni, D., Slessarev, E., MacMartin, D., Mahowald, N. M., Goodale, C. L., and Xia, L.: What goes up must come down: impacts of deposition in a sulfate geoengineering scenario, *Environmental Research Letters*, 15(9), <http://iopscience.iop.org/10.1088/1748-9326/ab94eb>

Wimmers, A. J., Moody, J. L., Browell, E. V., Hair, J. W., Grant, W. B., Butler, C. F., Fenn, M. A., Schmidt, C. C., Li, J., and Ridley, B. A. (2003), Signatures of tropopause folding in satellite imagery, *J. Geophys. Res.*, 108, 8360, doi:[10.1029/2001JD001358](https://doi.org/10.1029/2001JD001358), D4.

Answers to reviewer 2

Thank you for reviewing our paper, and for your helpful comments. Your comments about the figures in Appendix D made us aware that we had not mentioned that appendix in the main text, and prompted us to add a new supplementary info section about the use of Self Organizing Maps to better visualize the meridional stratospheric circulation patterns.

Line 31: A “pre-study” experiment is unclear – does this refer to an experiment performed before VolMIP officially got underway? Is the experiment not an official VolMIP experiment?

Yes, this experiment was performed before the other VolMIP experiments. It is a “pre-study” experiment as per Table 5 in Zanchettin et al., (2016). This is further explained in the introduction.

Line 61: define “ISA”

Done. The acronym explanation was added to the new first introduction paragraph.

Lines 64-66: Sentence does not quite make sense; I think some words are missing.

*Indeed. The missing words were also pointed out by RCI. Thank you both! The sentence now reads: “The climate models running the VolMIP *volc-long-eq* experiment will not simulate the volcanic aerosol cloud interactively, [since](#) the experiment [is](#) designed to ensure all models specify the same reference aerosol optical properties for the volcanic forcing.”*

Line 134: Please also specify the latitude of Tambora.

done

Line 313: Section 4.1. Somewhere in this section it would be worth pointing to Appendix A so you can contrast how your ‘approximate’ AOD was calculated compared with how ‘real’ AOD was calculated by the models.

Thank you for the suggestion. We added to the text: “The purpose of Eqs. (1 and 2) is to develop a simple analysis method to understand why the various models differ so much in computed AOD, [which is output either directly or as extinction values at each level that are integrated to get AOD \(Appendix A\)](#). The climate models are very complex, but the underlying physics relating the computed parameters of mass, optical depth and effective radius is relatively simple.”

Line 721: are there disadvantages to using point eruptions? Otherwise why not recommend that they always be used?

We do recommend that point injections be used over band injections, but understand that some studies have already been performed with band injections, and the implications of that needs to be understood by the community when drawing references from literature. Also, some models

(not included in this study) will likely continue to have band injections until they are made more advanced.

Figures 1-3: I understand that the mean of five ensemble members is shown for each model. It would be good to get some measure of the internal model variation, perhaps by plotting the mean plus/minus one standard deviation in a lighter shade. Otherwise it's hard to assess just how different the models are from one another.

Although the plotted results in Figures 1-3 are from ensemble means, there actually is extremely little internal model variation. The most internal model variation from the different ensemble runs is shown in figure 9 (previously figure 8) for the CESM-WACCM runs, in which case the "red" and "blue" ensemble members of that model were run with different QBO strengths. As you can see on that plot, there is very little difference between the global stratospheric averages within the same QBO strength. The other models didn't vary their ensemble runs as drastically, and have even less internal disagreement for the global stratospheric averages (almost negligible), so we went with the ensemble means only for the plots for legibility.

Figure 4: mentions 'Figure 4' twice in the caption.

fixed

Figure 9: SOCOL-AER. The green lines are indistinguishable, which might be worth commenting on in the caption.

Thank you for the suggestion. We added to the figure 9 caption (now figure 8): "The light and dark green dashed lines (and shading) for the SOCOL-AER real (and reconstructed) AOD plot are indistinguishable from each other because the values from the point and band injections are overlapping."

Please also check the order of figures; it looks like figure 9 is discussed before figure 8.

Done. Figure 8 and 9 have been swapped.

Appendix D is not referred to in the text, so figures D1 and D2 caught me by surprise – I found myself wondering how the SOMs were created; where the representative patterns had come from, etc. Potentially figures D1 and D2 could be moved to the main body since they are central to the paper, but I leave that decision to the authors.

Thank you for pointing this out. We had previously only mentioned the figures in the main text, and not mentioned the appendix as a whole. We added to a reference to Appendix D into the text in section 4.3.2.

We have decided to keep figures D1 and D2 in the appendix where they can be explained in detail without distracting the reader from the discussion of their implications in the main text. As they are central to the paper, we put them in the appendix rather than in the supplementary information.

We added more explanation about how the SOMs were created, what the representative patterns had come from, etc. to a new section of the supplementary info (Sect. S2), as was also requested by reviewer 1.

Answers to reviewer 3

Thank you so much for the very thorough, well written, and accurate summary in your review! And thank you for your comments, which improved the paper.

I do agree though with one of the other reviewers that the introduction is a bit wonky going right into the VolMIP protocols and gets kind of technical. I agree that some further context on why any of this is done would help motivate the important results that follow.

Thank you for the recommendation. We have added a new first paragraph to the introduction. (See response to RCI Introduction).

Table 1: Please explain/distinguish what is meant for CCM versus AGCM. For example, LMDZ-S3A does not include interactive OH, per Table 2, so why is it a CCM?

Thank you for catching this. LMDZ-S3A should be labeled as a CTM (Chemistry Transport Model) in Table 1 instead of CCM (Chemistry Climate Model). This change has been made. MAECHAM5-HAM is still labeled as AGCM (Atmospheric General Circulation Model). LMDZ-S3A is a CTM in this study because it nudges winds and temperature towards ECMWF analysis and does not take the aerosol heating (and subsequent aerosol cloud lofting) into account.

Line 146: Requested wavelength of output is 525 nm, but no model provides that, per Table 1 all but SOCOL-AER provide 550 nm. Is there a typo in one place or the other?

Thank you for pointing this out. The requested wavelength had indeed been 525 nm, but all of the models except for SOCOL-AER provided 550 nm. This behind the scenes detail of requesting 525 nm and settling on 550 nm is not something that needs to be in the paper. Thus, the text for this paragraph has been changed to:

“The models provided AOD in the visible spectrum at the wavelength $\lambda = 550$ nm. The exception was SOCOL-AER, which calculated the AOD output over a wider band ($\lambda = 440$ to 690 nm), but still in the visible spectrum (Table 1). The requested wavelength from the experiment protocol for AOD in the visible was $\lambda = 525$ nm, but the actual wavelengths at which models provided their outputs are presented in Table 1. While different wavelengths were used within the visible band, they still fall within the Mie scattering regime for volcanic sulfate aerosols, because the optical size parameter of $\alpha = \frac{2\pi r}{\lambda}$ remains within the order of 1-10 for particles of radius 0.1-1 μm . Specifically, SOCOL-AER AOD reported values over a wide wavelength range, as shown in Table 1. However, SOCOL-AER produces relatively large particles, as discussed below, with size parameters around seven for 525 nm wavelength. In the large particle limit for optical scattering, AOD is not very wavelength dependent, so SOCOL-AER’s wide wavelength range is unimportant when comparing peak AOD magnitudes. SOCOL-AER and LMDZ-S3A use sectional size distribution schemes. The rest of the models use modal size distribution schemes. Further details about the size distribution schemes used by the models can be found in Table 2 and Appendix B.”

Additionally, we moved the comment about the wider wavelength range in SOCOL-AER being unimportant for comparing peak AOD magnitudes to the discussion section 4.1 We added to the text: “Although SOCOL-AER calculated AOD over the range $\lambda = 440$ nm to 690 nm, instead of at $\lambda = 550$ nm, the different wavelength is not very important for comparing AOD magnitudes across models because of the Mie scattering properties. The value of q , (and $q/Reff$), for a given effective radius at $\lambda = 550$ nm falls in the middle of the values between $\lambda = 440$ nm and $\lambda = 690$ nm.”

It is implied in a few places, but do all the models account for a full suite of tropospheric-sourced aerosols? Dust, sea salt, carbonaceous, surface SO₂ sources? Nitrates too? And to the extent it matters to the results, have you considered interactions of those aerosol sources with the volcanic plume, maybe especially for the modal models? Also, do all the models include an interactive Junge layer aerosol? How about meteoritic smoke?

As we are investigating the differences in the stratosphere, none of these tropospheric-sourced aerosols matter in our situation. (You can see them on the plots in figure 4 for CESM-WACCM and UM-UKCA in the lower troposphere due to their large effective radii), but they do not influence the stratosphere. Radiative forcings for CO₂, other greenhouse gases, tropospheric aerosols (and O₃ if specified in the model), were set to the values each model uses to define preindustrial (1850) climate conditions (Section 2 Methods). The modal models account for this in the size distributions (Appendix B). Since the aerosol cloud from Tambora is so large of a perturbation, background aerosols in the Junge layer and meteoric smoke should not be a cause for intermodel disagreement even if treated differently. None of the models have meteoritic smoke, which is episodic, in this experiment. For example, CESM-WACCM and SOCOL-AER have background sulfur in the forms of OCS and DMS, but their natural global burdens are negligible (several orders of magnitude lower) compared to the sulfur injected by the volcano.

Line 264 and Appendix A around eqn. (A3). This is pedantic, but I got a bit confused following the variable naming and subscripting. Maybe it is all very clear, but it wasn't to me. In A2 I understand $reff$ to the effective radius at a given grid point and moment in time. In A3 $Reff$ is the global averaged effective radius, but curiously $Reff$ appears on both sides of the equation. So I think this is just a typo and “ $reff$ ” is what is supposed to be on the right side. But A3 pertains to the vertical integration, and the horizontal integration is described somewhat strangely in lines 776-783. I'm not sure if this discussion about gaussian weights is illuminating or necessary consequence of the model output presented on zonal mean grids or what's going on. It might be more clear if A3 were written with the expression wrapped inside of some horizontal integral, is that what's going on?

We want this appendix to be clear for reproducibility purposes so that other groups could compare their results to ours in the future. This is why we go into detail about the methods used in the horizontal and vertical integrations.

The first paragraph in Appendix A now reads: “The participating models in this study provided monthly outputs of time-averaged data in different vertically and horizontally resolved grids, units and formats. Some values data were already pre-processed, for example provided in terms of

stratospheric values (i.e. as column sums or averages) and/or in zonal means-values (i.e. as longitudinal sums or averages) . ~~Some gave data for~~ Conversely, some data were output at all model levels ~~including the troposphere~~, and/or at all longitudes. Some models output pressure levels as vertical indices, some gave only altitude, and a few provided both. For consistency, the following post-processing methods were applied to obtain the monthly stratospheric (meaning vertically reduced) global (meaning horizontally reduced) mean values of AOD, sulfur species burdens, and effective radius. “

It is correct that $reff$ is the effective radius at a given grid point and moment in time. In A3, the right side of the equation has $Reff_tau$ (i.e. global average at each vertical level tau), and the left side has $Reff$ (i.e. global stratospheric average). So this was not a typo, but we have added more clarification of this in the text in the paragraph surrounding equation A3.

Line 297: It is interesting to find a statement like “The goal of this paper. . .” in Section 4 of the paper. I think such a clear statement belongs also in Section 1, right before the sentence on line 86 beginning “In this paper we go further. . .”

As this paper features a very long discussion section, we prefer to retain this statement where it is as an introduction to the discussion. In addition, we have added a similarly clear statement to Section 1 at the location suggested. The statement we added is: “Now, our goal is to identify and understand the causes of intermodel disagreement in the AOD itself.”

Line 308: The use of “omega” in a discussion of optical properties could be a bit confusing as it is frequently associated with single scatter albedo. I suggest another symbol. But it’s not a big deal, I follow just fine.

Since you are able to follow along with our use of “omega” without difficulty, we are opting to keep the symbol as-is after checking that it has been sufficiently defined in our text.

Line 321: Speaking of the particles picking up water I presume the hydrated properties are diagnosed and water is not transported on the particle by any of the models. Am I correct?

This is correct.

Figure 5: I was curious what this figure would look like if the circles were sized by effective radius instead of age. Did you make such a plot? I get this is what Figure 6 shows, but on different axes.

Thank you for the suggestion! We made this change to figure 5. The plot looks cleaner now and a hint of the AOD dependence on effective radius can be seen.

Line 383: I note that Figure 9 is introduced before Figure 8, which comes in several pages later.

Done. Figure 8 and 9 have been swapped.

Line 579: If LMDZ-S3A computed optics assuming $\omega = 0.75$ why does it's internal AOD track so closely with $\omega = 0.9$ curve in Figure 9?

(This is now Figure 8). ω still varies in the microphysics, even though a constant value is used for its refractive index calculation for q . Thus, the ω in the denominator of Eq. 2 is still changing in the model. These reconstructed AOD values are approximate, as explained in Sect. 4.1