Author comments

Reviewer comments are written in black and the author's response is in blue.

We thank the anonymous reviewers for their valuable feedback and additions to the text. All comments have been addressed as stated below.

The following changes not in connection with the Reviewer's comments were also made:

A bug was found in the processing script for the MIPAS SO2 stratospheric burdens, which led to the lowermost levels of the stratosphere being excluded. Since most SO2 resides in the lower stratosphere, this changes the burden considerably from 5.7 Gg(S) to 10.6 Gg(S). The higher extent of the dataset is not expected to increase the burden much, as the values there are orders of magnitude smaller.

Additionally, we now use an extended dataset for SO2 (MIPAS) and introduced a new dataset for OCS (MIPAS). These data were provided by Michael Höpfner, who has been added to the author list.

Another bug was identified in the calculation of SAD in ECHAM6-SALSA, which resulted in low values, not in accordance with SAOD and sulfate burdens. This has now been corrected. Further, the mass density of sulfate for ECHAM6-SALSA has now also been added to Fig. 4.

Reviewer 1

This is a long paper, and it seems to border somewhere between comprehensive and tedious. It would be nice for the science issues to be separated from the listing of values for each and every model for each and every variable. Perhaps the paper could be made quite a bit shorter, with many of the details moved to the appendix, for those interested in them.

The manuscript is indeed long and sometimes, the listing of values disrupts the text flow. However, we believe this is necessary to present the model differences of the most important variables directly with numbers in the text to not disperse the information too much throughout the manuscript.

The conclusions are that models are different from each other and different from the available observations. But what now? Why is this?

We have kept the conclusions simple and don't make any hard claims on why the models are different. This is on purpose since many more simulations are needed to conclusively separate and attribute the processes involved. We mention that more process-oriented experiments are needed, as also stated in Quaglia et al. (2023). Both publications are meant as benchmarks (background and volcanically perturbed), which will be referenced in the future, to develop targeted sensitivity studies.

What specifically has to be addressed to improve them?

As stated above, this paper makes only very soft claims on what could be addressed to improve models. We address possible underlying issues but an exact attribution will require

additional experiments, which are already planned for the future or are already performed individually by the modeling groups. Such experiments should focus on specific components of the models, e.g. only the microphysical modules to help understand the interplay of nucleation and condensation and their influence on the aerosol size distribution.

And how important are these climatologies for non-volcanic periods?

Whenever there are no major volcanic eruptions, the background aerosol layer still contributes to the global radiative forcing. This is shown by Solomon et al. (2011), where the background contributes with a forcing of about -0.1 to -0.2 W/m2, although there are considerable variations also in this background.

Additionally, most models are usually only tested for a perturbed aerosol layer. Under background conditions, their behavior could be different and should be tested to provide a reference for future studies.

How much do these differences affect their responses to volcanic eruptions or potentially to climate intervention?

This is specific to the case at hand. Biases in e.g. stratospheric transport are always important, as well as differences in microphysics and chemistry.

Other factors, such as the OCS burden, transport, and photochemical conversion are less important in volcanically perturbed conditions.

There are problems with several of the plots. The complete seasonal cycle is not plotted, with the variations from December to January missing. Italics should not be used for units. Many of the contour labels cannot be read with a background very similar to the text. Units for pressure should be hPa, and not Pa.

All of the above suggestions for changes to the plots have been implemented now in all figures.

In many places, “notably” or “It should be noted” are used, and all should be deleted. Every sentence in the paper should be noted or it should not be in the paper.

The corresponding sentences have now been changed everywhere in the manuscript, according to the document provided as a supplement to this comment.

Several acronyms, including SAOD and UTLS are not defined. And for SAOD, at what wavelength? And SAD is defined multiple times.

The acronyms have now been defined and the information on the wavelength has been added to the text.

Northern Hemisphere and Southern Hemisphere are names and should be capitalized. And it would be better to use NH and SH as acronyms for them throughout the paper rather than just occasionally.

We now use the acronyms NH and SH throughout the paper manuscript.

The 56 comments in the attached manuscript all have to be addressed.

All comments marked in the attached document were addressed. The most important ones have been pasted and answered below:

line 7: I don't understand. How can you identify discrepancies that are masked

We have deleted the sentence in question to avoid confusion and to make the abstract more concise. What we mean to say is that volcanic eruptions have a very large signal. Therefore the processes assessed in such studies are mostly connected to the changes due to this large signal. Our question here is are there processes in the models affecting the background burden, which are overlooked when dealing with large perturbations.

line 8: I don't understand. What may still matter?

See answer above.

line 11: But these are model-based. How do you separate the model biases of the reanalysis models from those you are trying to study?

The reanalysis was now deleted from the sentence, since we don't use it directly to compare models to but only to calculate the tropopause to then get the stratospheric burden for aerosol, SO2, and OCS.

line 14: But that's because the majority of the emissions are in the NH

The sentence has been extended to include this statement.

line 17: Delete "We investigate the areas of greatest variability in the sulfur species burdens and find that"

Deleted

line 18: some rewordings

Corrected

line 21: Delete "This study highlights that"

Deleted

lines 24-25: Delete the last sentence, this is always true

It is always true that more research should be conducted on a topic. However, with this sentence we provide an outlook and specify, where this research will be the most useful, that is for studies on stratospheric aerosol injections (SAI).

line 64: Add Kravitz et al. (2011) and Visioni et al. (2023). The ones you chose are not model intercomparison programs.

We refer here to model intercomparison studies, not to model intercomparison programs, Weisenstein et al. (2022) and Franke et al (2021) both show results of geoengineering scenarios studies with global stratospheric aerosol models and therefore comparable to the volcanic studies by Marshall et al. (2018), Clyne et al. (2021), Quaglia et al. (2023). We therefore stick to our original choice

line 81: what does this mean?

The CMIP models use a semi-background aerosol state, which includes volcanic activity, as it is just a time average over 1850-2014. This study can help get a better estimate of the background state.

line 83: define acronym. And what does this mean? IPCC assesses research. Why does it need radiative forcing estimates?

The sentence where we state that we could provide better forcing estimates for the IPCC has now been deleted, since we don't discuss radiative forcing in this paper. However, our results may still be helpful for related research, but not directly for the IPCC.

Caption Figure 1: This is confusing. How can the reader tell the difference?

The caption has now been changed, so as to clarify the meaning of the observational and previous model estimates provided.

lines 423-427: This is a lot of detail that can be in a table or supplemental info.

See comment above

line 453: ???

In this sentence, we state that the relationship between SAOD and stratospheric sulfate burden is linear in the tropics and less so in the extratropical region. What we meant to say is that the relationship between SAOD and sulfate burden is always linear in this background state, however, the scatter around this linear relationship increases towards the poles. This has been changed in the text.

Caption Figure 3: I don't understand the x-axis. Are the left-hand side Jan. 15 or Jan. average, and the right-hand side Dec. average? If so, you have only plotted 11/12 of the seasonal cycle. Please re-plot these with the same date on the left and right sides, either Jan. 1 or Jan. average, so that you plot the missing Dec. to Jan. data. The left and right sides of the plots need to match.

This has now been corrected as the Reviewer suggested. The values are monthly means and each time step is on the 16th of each month. The values from December have been added to the beginning and the values from January to the end of the time-axis, in order to depict the full seasonal cycle.

Caption Figure 7: Why do the contours not match the shading increments?

The shading and contours have now been set to values that match.

line 1064: Unacceptable reference

The paper in question (Tilmes et al., 2023) has been formally reviewed and published and the reference in this manuscript was updated.

References:

Clyne, M., Lamarque, J.-F., Mills, M. J., Khodri, M., Ball, W., Bekki, S., Dhomse, S. S., Lebas, N., Mann, G., Marshall, L., Niemeier, U., Poulain, V., Robock, A., Rozanov, E., Schmidt, A., Stenke, A., Sukhodolov, T., Timmreck, C., Toohey, M., Tummon, F., Zanchettin, D., Zhu, Y., and Toon, O. B.: Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora Interactive Stratospheric Aerosol ensemble, Atmos. Chem. Phys., 21, 3317–3343, https://doi.org/10.5194/acp-21-3317-2021, 2021.

Franke, H., Niemeier, U., and Visioni, D.: Differences in the quasi-biennial oscillation response to stratospheric aerosol modification depending on injection strategy and species, Atmos. Chem. Phys., 21, 8615–8635, https://doi.org/10.5194/acp-21-8615-2021, 2021.

Marshall, L., Schmidt, A., Toohey, M., Carslaw, K. S., Mann, G. W., Sigl, M., Khodri, M., Timmreck, C., Zanchettin, D., Ball, W. T., Bekki, S., Brooke, J. S. A., Dhomse, S., Johnson, C., Lamarque, J.-F., LeGrande, A. N., Mills, M. J., Niemeier, U., Pope, J. O., Poulain, V., Robock, A., Rozanov, E., Stenke, A., Sukhodolov, T., Tilmes, S., Tsigaridis, K., and Tummon, F.: Multi-model comparison of the volcanic sulfate deposition from the 1815 eruption of Mt. Tambora, Atmos. Chem. Phys., 18, 2307–2328, https://doi.org/10.5194/acp-18-2307-2018, 2018.

Quaglia, I., Timmreck, C., Niemeier, U., Visioni, D., Pitari, G., Brodowsky, C., Brühl, C., Dhomse, S. S., Franke, H., Laakso, A., Mann, G. W., Rozanov, E., and Sukhodolov, T.: Interactive stratospheric aerosol models' response to different amounts and altitudes of SO2 injection during the 1991 Pinatubo eruption, Atmos. Chem. Phys., 23, 921–948, https://doi.org/10.5194/acp-23-921-2023, 2023.

Tilmes, S., Mills, M. J., Zhu, Y., Bardeen, C. G., Vitt, F., Yu, P., Fillmore, D., Liu, X., Toon, B., and Deshler, T.: Description and performance of a sectional aerosol microphysical model in the Community Earth System Model (CESM2), Geosci. Model Dev., 16, 6087–6125,1095 https://doi.org/10.5194/gmd-16-6087-2023, 2023.

Visioni, D., Kravitz, B., Robock, A., Tilmes, S., Haywood, J., Boucher, O., Lawrence, M., Irvine, P., Niemeier, U., Xia, L., Chiodo, G., Lennard, C., Watanabe, S., Moore, J. C., and Muri, H.: Opinion: The scientific and community-building roles of the Geoengineering Model Intercomparison Project (GeoMIP) - past, present, and future, Atmos. Chem. Phys., 23, 5149–5176, https://doi.org/10.5194/acp-23-5149-2023, 2023.

Weisenstein, D. K., Visioni, D., Franke, H., Niemeier, U., Vattioni, S., Chiodo, G., Peter, T., and Keith, D. W.: An interactive stratospheric aerosol model intercomparison of solar geoengineering by stratospheric injection of SO2 or accumulation-mode sulfuric acid aerosols, Atmos. Chem. Phys., 22, 2955–2973, https://doi.org/10.5194/acp-22-2955-2022, 2022.

Reviewer 2

The paper presents a comparison of time-slice simulations by 9 chemistry climate models for a volcanically quiet situation. The idea goes back to a SSIRC (Stratospheric Sulfur and its Role in Climate) meeting in 2013 and a paper in 2018. The results show that the outliers are models with inappropriate horizontal resolution (e.g. Lofverstrom and Liakka, 2018) or with lack of important parts of sulfur chemistry. This should be addressed also in the conclusions.

The differences arising from low horizontal resolution or incomplete chemistry schemes have been addressed in the conclusions. However, we want to refrain from making hard statements here, as the present study with only one simulation per model does not allow for a definitive attribution of causes.

A problem in the generally well written paper is that available satellite observations of OCS and middle atmospheric SO_2 and the reanalysis data (ERA5 or ERA-I) are not used in the comparisons but only multi-model averages instead (except for the tropopause height). This should be improved before publication.

We now include both an extended dataset for SO2 and an additional dataset for OCS, provided by Michael Höpfner (who is now Co-authoring this publication).

While the MIPAS, SAGE-3l and OPC measurements provide important insights, ERA-I was intentionally not used in the comparisons, since we want to first and foremost intercompare the models. Further, the comparison of time-sliced model runs with observations or reanalysis, which are by nature transient, comes with many caveats. In particular, temperature would need further processing, such as detrending the data. For OCS, this is unproblematic, as there is no trend and seasonal cycles are very stable. SO2 is mostly influenced by volcanic eruptions, which has been addressed by excluding the most affected months.

Specific comments

Line 30: A reference to Watts (2000) would be useful here too (also in line 324).

The reference has been added in line 30.

Line 243ff: Which sulfur species are included? This model is the only one where this information is missing.

The information has now been included in the UKCA model description as follows: The chemical scheme accounts for OCS, SO$_2$, DMS, H$_2$SO$_4$, SO$_3$, and sulfate aerosol (Dhomse et al., 2014).

Line 263 or 282: Two important datasets are missing: MIPAS OCS by Glatthor et al. (2017) and MIPAS SO_2 up to 45 km (Höpfner et al., 2013). This should be also included in Figure 1 and later.

The datasets in question have now been added in the text and in all Figures.

Line 392: Add: "and missing OCS, see later".

In both studies cited here, no OCS was present, therefore it was not specifically mentioned.

Line 550: This is in contradiction to MIPAS data.

A second estimate for the emissions and plume height by Höpfner et al. (2015) has now been included.

Figure 6: Please include MIPAS data (caution with averaging kernel effect, Glatthor et al., 2017).

The data has now been included. We mention the averaging kernel effect, but do not modify the model data. Since we do not see the smeared out patterns as in Glatthor et al. (2017) but also since we want to see how the models directly represent the distribution of OCS.

Line 584: Replace "polar" by "Antarctic".

Corrected

Figure 7: Fill in MIPAS data in panel k. Here it might be useful to display only the ones of Höpfner et al. (2013) which show the secondary maximum due to OCS photolysis and also the increase in the middle stratosphere simulated e.g. by WACCM.

The MIPAS data provided by Michael Höpfner has been added in panel k.

Line 628ff: Modify corresponding text.

The text has been modified to accommodate the information on the added data by MIPAS.

Figure A1: Please comment on outgassing volcanoes in caption. Compared to standard references some important ones appear to be missing in scenario NAT.

Volcanic outgassing emissions are included in both scenarios, the information is now in the caption.

Figure A3: Why is there no comparison with ERA-I?

We do not compare it to ERA-I due to two main reasons. First, we want to set the focus on inter-model differences, hence we compare each model to the multi-model mean. Secondly, temperature observations would require additional processing to account for trends and inter-annual variability to enable a meaningful comparison to our time-sliced experiments.

Technical corrections

Figure 1: Some fonts too small.

Checked and corrected

Figure 3: Please include more ticks at the abscissa in panels a-e. List the peak values of ULAQ in panels s,t in the caption. Include a remark there that panel u-y include only a subset of models (?).

More ticks have been included in the first row of subfigures. Peak values of ULAQ (for stratospheric SO2) and the WACCM6 models (for stratospheric OH) have been included in the caption. OH was not provided in the output of all models and is therefore only included for those who did.

Line 348: Include "e.g.".

Included

Line 548: Typo in volcano name.

Corrected

Figure 5 and 6: Is there a gap between Dec 15 and Jan 15? This looks odd, please improve that.

Since we have monthly mean data, corresponding to the 16th of each month, this initially meant that the values between Dec 15 and Jan 15 appear to be cut off. This has now been fixed and the whole seasonal cycle is depicted in all figures.

Line 656: Typo or grammar problem.

Corrected

Figure A2-A4: Consider superscripts for the units in the captions.

Corrected

Please correct several typos in author names and missing subscripts in the reference list. There are also wrong links (lines 1065 and 1114). Don't repeat identifiers several times (line 816f).

All references were now scanned for typos and wrong links.

References:

Dhomse, S. S., Emmerson, K. M., Mann, G. W., Bellouin, N., Carslaw, K. S., Chipperfield, M. P., Hommel, R., Abraham, N. L., Telford, P., Braesicke, P., Dalvi, M., Johnson, C. E., O'Connor, F., Morgenstern, O., Pyle, J. A., Deshler, T., Zawodny, J. M., and Thomason, L. W.: Aerosol microphysics simulations of the Mt. Pinatubo eruption with the UM-UKCA composition-climate model, Atmos. Chem. Phys., 14, 11 221–11 246, https://doi.org/10.5194/acp-14-11221-2014, 2014.

Glatthor, N., Höpfner, M., Leyser, A., Stiller, G. P., von Clarmann, T., Grabowski, U., Kellmann, S., Linden, A., Sinnhuber, B.-M., Krysztofiak, G., and Walker, K. A.: Global carbonyl sulfide (OCS) measured by MIPAS/Envisat during 2002–2012, Atmos. Chem. Phys.17, 2631–2652, https://doi.org/10.5194/acp-17-2631-2017, 2017.

Höpfner, M., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Orphal, J., Stiller, G., von Clarmann, T., Funke, B., and Boone, C. D.: Sulfur dioxide (SO2) as observed by MIPAS/Envisat: temporal development and spatial distribution at 15–45 km altitude,13, 10 405–10 423, https://doi.org/10.5194/acp-13-10405-2013, 2013

Lofverstrom, M. and J. Liakka: The influence of atmospheric grid resolution in a climate model-forced ice sheet simulation, The Cryosphere, 12, 1499–1510, doi:10.5194/tc-12-1499-2018, 2018.

Watts, S.F.: The mass budgets of carbonyl sulfide, dimethyl sulfide, carbon disulfide and hydrogen sulfide, Atmos. Environ., 34, 761–779, 2000.