Response to Prof. Hewitt comment RC1

We thank Prof. Hewitt for her supportive comments on our work and her suggestions for improvement. All comments were addressed as per her recommendations except for the first one (see below), and we briefly respond below to any points that required particular clarification. All editorial recommendations were implemented.

I would however like to request that all the code used to produce the forcing dataset archived in a github directory that has WCRP or CMIP ownership to ensure the long-term legacy of the dataset.

Although we think this is an excellent suggestion, following discussion with the CMIP Forcing Task team leadership and Prof. Hewitt, we decided to not implement the recommendation to put the dataset and codes on a WRCP or CMIP repository because:

i) Should the stratospheric aerosol forcing providers do this, this should likely be done consistently for all CMIP forcing providers. There is currently no established protocol for this and it is unclear all forcing providers would be in a position to do this.

ii) With zenodo being a permanent repository, the datasets and codes associated with the manuscript should be safe for long term storage.

However, we remain in agreement with Prof. Hewitt that long-term, establishing a protocol to publish CMIP forcing providers source datasets and codes on a WRCP or CMIP-owned repository would be a desirable practice.

L197: Can the authors comment on whether there are likely gaps in small to medium volcanoes in the Southern hemisphere in the pre satellite period given that the ice core is from Greenland?

We comment on the fact that D4i does not capture SH eruptions line 202. We revised lines 784-786 as: "However, the fact that the D4i dataset is based on a single Greenland core means it is characterized by very high uncertainties, that it does not provide any constraint on eruption latitude, and that it only records some tropical eruptions and no southern hemisphere eruptions."

L272: The relatively cheap EVA model is perhaps in contrast to the computational cost of the models used to produce the ozone datasets. This is perhaps worth commenting on in the future perspectives discussion

We added sentences mentioning that using interactive stratospheric aerosol models could be a better use of state-of-the-art tools and enable to produce ozone and volcanic forcing from the same simulations. However, large uncertainties persist in stratospheric aerosol modelling the choice of a model or a model ensemble is a difficult question (and the reason why EVA was used in VolMIP CMIP6, Zanchettin et al., 2016). From a practical point of view, the inexpensive nature of our model was also key to being responsive in implementing updates to the dataset.

L329-331: This sentence didn't make sense to me in particular 'deemed less risky to bias the model calibration'. Please reword

We reformulated in the revised manuscript as:

"Numerous other pyrocumulonimbus perturbations could have been introduced (e.g. 2017 Canadian pyrocumulonimbus, Kloss et al. 2019, 2021). However, owing to their smaller magnitudes, these events are less likely to have driven SAOD perturbations that would strongly bias the model calibration in the absence of corresponding SO_2 injection."

L355: Is the EVA_H v2 model in figures 4b-f the calibrated one? Please clarify

We clarified the model version everywhere in the revised manuscript.

Figure 4a: The calibrated model looks particularly smooth between ~2000 and 2005. Is it obvious why that might be the case?

We would argue that, given the logarithmic scale, EVA_H v2 looks smooth at any time compared to GloSSAC. This smoothness is due to a combination of: i) missing stratospheric aerosol (precursor) injections, including small eruptions, wildfire, and cross-tropopause transport; ii) the simplicity of aerosol formation, transport and removal processes in the model which are governed by a relatively small set of timescales; iii) the imperfect dependence of the model timescales on eruption parameters. These limitations are highlighted throughout this section and the wider manuscript.

Figure 6c: The comparison with UKESM1 suggests that the background SAOD is too high in 1850. Can you comment further on that?

We added a comment in the revised manuscript.

L394: Is it possible to quantify the impact of using a coarser resolution index?

The impact of using a coarser resolution index would be to have greater error associated with interpolation for a lower resolution. We will consider quantifying this error for our upcoming publication documenting an uncertainty product for the dataset.

L556: It would be worth noting that the impact of Hunga Tonga eruption can be assessed in CMIP7 if historical simulations are continued in parallel to the projections forced by scenarios (as described in Hewitt et al., PLOS Climate, 2025)

Agreed, we stated this explicitly in the discussion section in the revised version.

L639: I think that the conclusion here is that the two datasets could be used to explore the uncertainty in the forcing datasets and a future aim might be to provide some indication of uncertainty with the datasets. Is it possible to make an initial estimate of the order of magnitude of the uncertainty on SAOD given the comparison between CMIP6 and CMIP7?

L745: As above, I think it would be good to recommend providing uncertainties in a future update of the dataset

We added in section 6.1 that the development of an uncertainty product is important (and this product is being developed). Comparison between CMIP6 and CMIP7 can indeed give an idea of what uncertainties are expected, but we suggest that more careful analyses of the dataset and source of uncertainties are required, and these will be the focus of a paper on the upcoming uncertainty product.

L830: I was very pleased to see this list of key resources and suggest adding a comment that these need to be included in GCOS considerations

Excellent suggestion, we implemented it in the revised manuscript.

L860: It would be good to include an estimate of the person effort for extensions and updates

Excellent suggestion, we added it to the manuscript. We estimate a person time on the order of 1 month for extensions, and 12 months for updates.

Response to anonymous referee comment RC2

We thank the reviewer their supportive comments on our work and their suggestions which helped us improve the manuscript. We briefly respond below to any point that required a response. We did not respond to editorial/minor comments but implemented all the associated reviewer suggestions in our revised manuscript.

Specific comments:

Line 75; Section 1.2:

It is correct that there is no complete published CMIP6 dataset description, however, in the supplement to Jörimann et al. (2025) the SAGE-3λ record is documented, which was the latest iteration that went directly into the CMIP6 forcings. Using this source, you can confirm that for 1961-1978 pyrheliometer data from Stothers (2001) - which you cite - are used. Please also check in this entire subsection and Fig. 1a, if the supplement can complement your overview of the CMIP6 forcing. Since SAGE-3λ uses three wavelengths (wherever possible), it should also be specified that GloSSAC provides more than two wavelengths (line 83) on occasion.

We thank the reviewer for pointing to how we could better use Jörimann et al. (2025). Implemented the suggestion on GloSSAC wavelengths and pyrheliometer use in corresponding sections.

Line 211:

*You choose the VEI 4 SO_2 mass for GVP events to match the anomaly from 1998-2023 "... (defined **as** the deviation from its minimum) ...". I agree that the chosen time period is fit for this purpose, but it is not exactly clear what the deviation from the minimum is. Is the minimum the very lowest SAOD data point found in this time period? If so, is it representative for an "undisturbed" stratospheric aerosol or could it be an outlier? Or could an average such as 1999-2003 (quasi-quiescent state) be taken as the minimum to derive the deviation against? Please elaborate what is meant by the minimum and why it was chosen this way.*

The minimum indeed refers to the lowest value over the time period, and we believe the use of minimum is clear enough. However, to improve clarity and address the well justified reviewer comment, we added a sentence quoting the value of this minimum. We chose the actual minimum instead of the mean of a volcanically quiescent period because there is no time period which is perfectly volcanically quiescent, and taking the minimum thus likely already represents a conservative estimate of the true baseline.

We acknowledge that this approach is crude and, beyond the question of how to define a baseline, is subject to limitations such as non-volcanic influences on SAOD over 1998-2023 or a varying anthropogenic contribution to SAOD (in particular through cross-tropopause aerosol transport).

Line 232:

*Periods without SAGE coverage are supplemented, but periods with SAGE coverage are also partly supplemented, especially at high latitudes, can you confirm this? The sentence could simply be extended similar to: "Periods without SAGE coverage **and high latitude data not***

***captured during SAGE coverage** are supplemented by complementary spaceborne and ground-based observations …"*

Excellent suggestion implemented in revised manuscript.

Lines 294-297:

I find this sentence hard to understand due to its length. Consider splitting it into two sentences for the reader's convenience.

We reformulated the sentence to make it easier to understand and shorter.

Line 347:

*It says that after the initial model parameter search with eruption masses **higher** than 0.1 Tg SO_2, the refined search then respects the other eruptions "… with an upper end of stratospheric SO_2 mass **higher** than 0.1 Tg SO_2." Is it not supposed to be "lower" here, instead of "higher" again? Otherwise I do not understand the distinction between the initial and refined search.*

The distinction is between best estimate and upper-end estimate. We changed this sentence to:

"[…] with a best estimate of stratospheric SO2 mass higher than 0.1 Tg SO2. We then did a refined search allowing for adjustment of SO2 masses for the 88 eruptions with an upper-end estimate (as opposed to best estimate) of stratospheric SO2 mass higher than 0.1 Tg SO2."

Line 423:

M_{H2SO4} must be the molar weight, not the molar concentration of H_2SO_4. This follows from dimensional analysis of your Eq. (2), which yields mass per mole for M_{H2SO4}, not number per volume as it would be for molar concentration.

Thanks for catching that mistake, we corrected it in the manuscript.

Line 505:

From previous description it seems that the extinction coefficients are derived using version 2 (exclusively) of EVA_H. If so, specify "EVA_H v2", as the term "EVA_H" has been used before to distinctly talk about the EVA_H group (v1 and v2). The version number should also be used in the subplot titles of Fig. 7 & 8

This is correct and we specified v2 vs v1 throughout the manuscript where required, including where suggested by the reviewer.

Line 526:

In this section the wavelengths, on which you produced ext, ssa, asy are given, however, the radiative transfer models operate with wavelength bands. In the data you provide both "wavelength" and "wavelength_bnds" variables and it seems that the wavelength is the (linear!) center or midpoint of each wavelength band. The wavelength bands are continuous across the spectrum. Do the data (e.g. ext) you report on the 39 wavelengths, correspond to the data you computed for the respective wavelength band? If so, can you make it clear in the text, notably in this section (4.3)? Also specify what exactly is meant by "properties at arbitrary wavelengths" on line 545 in this context.

All data provided is for the corresponding wavelength, not wavelength band. However, the script briefly documented in section 4.5 calculates properties averaged over user-inputted wavelength bands. We clarified that distinction in section 4.3. “arbitrary wavelengths” refers to the fact we can produce these properties at any wavelength other than 525 and 1020 nm, although in the CMIP7 framework, we produced them at the list of 39 wavelengths provided.

Lines 574-575:

The colors you mention (red, green, and black) do not correspond to what is shown in Fig. 9 (different shades of red). Change the caption or the lines in the plot.

Thanks for catching that mistake, we corrected it in the manuscript.

Line 1130:

Please add the ETH research collection item Luo (2017), which was created as a more accessible and persistent item that contains the CMIP6 data and some documentation and also has a DOI (https://doi.org/10.3929/ethz-b-000715155).

Since the specific file on the FTP file server that you reference in Luo (2018) is not in the ETH research collection item, I suggest that you either keep the FTP link as the separate citation you already have, or add a note to the Luo (2017) citation to mention that some description documents are only available there, similar to (“with additional information accessible at …”).

Finally, update the last access.

Citations not already in the manuscript:

Luo, B.: SAGE-3λv4: Stratospheric aerosol data for use in CMIP6 models, ETH Research Collection [data set], https://doi.org/10.3929/ethz-b-000715155, previously distributed through ftp://iacftp.ethz.ch/pub_read/luo/CMIP6_SAD_radForcing_v4.0.0 (last access: 8 June 2025), 2017.

Thank you, we added this citation in the revised manuscript.