Review of Stratospheric aerosol forcing for CMIP7-Part3" by Man Mei Chim

Stratospheric aerosols play a pivotal role in the Earth's Climate system, especially after major volcanic eruptions when their concentrations can increase by several orders of magnitude for multiple years. For these reasons, their effects must be accounted in climate simulations and developing a database encompassing their impacts is essential for CMIP7. In their paper, Chim et al., 2026 provides a comparison between the stratospheric aerosol forcing dataset used in CMIP6 and the new CMIP7. This is a very challenging problem since both datasets were developed by different groups involved in the CMIPs exercises which did not fully account for the potential intrinsic issues raised by such an evolution. I believe that a large fraction of the discrepancies described in this paper is somehow related to the discontinuity in this evolution.

I consider this effort a milestone toward improving the representation of stratospheric aerosol forcing, which is nevertheless still at a very premature level in the interpretations of the pre-satellite and post-satellite differences. A significant number of datasets have been omitted in this evaluation exercise, especially in situ airborne measurements. I found that the overall error analysis of stellar extinction, lunar eclipse and pyrheliometers are largely underestimated or missing, which naturally makes the comparison and the validation of the two CMIPs very speculative. The authors do not try to make use of the large density of lidar or in situ measurements to verify results from CMIP7 after moderate volcanic eruptions which took place such as Raikoke and Hunga. For all these reasons, I do not recommend the publication of this article before major revisions. Below are additional comments that support this decision.

L70. I found this sentence a little misleading. CMIP6 after 1979 relies on the Retrieval Method for optical Physical and Aerosol Properties (Remap, Jorimann et al., 2025) using mostly multi-wavelength approach from SAGE II/ISS data except between 2005 and 2017.

L105. Could you further explain the use of a linear trend between 1850 and 1978. Is seems rather speculative and not supported by the limited observations available.

L123. I think airborne measurements should also be used in this evaluation exercise.

Figure 1. This figure is missing critical datasets to study stratospheric aerosols such as lidar and airborne measurements. Several satellite observations are also missing and it is not clear why certain satellite observations are preferred (SCIAMACHY, OMPS) compared to others (MIPAS, GOMOS, CALIPSO).

Section 2.3.1. I think this section is very weak since there is no comparison/validation of the lunar eclipse technique with more robust observations.

Figure 2. The quality of the figure could be improved with a grid. It is true for most of the figures in this paper. An important dataset such as Mauna Loa transmission is missing. https://gml.noaa.gov/grad/mloapt.html. Balloon data available from the 70's are also missing. The lunar eclipse data show almost no correlation with simulations prior to 1960. The post-

Pinatubo period shows that both eclipse and stellar extinction strongly overestimate SAOD. To me the error bar of the eclipse data is largely underestimated. Error bar for the stellar occultation measurements should be included.

L292. I do not see a quantitative statement which could help stating that CMIP7 yields of a more coherent time series.

Figure 3. This figure could be improved by including % difference, the position of the tropopause and maybe isentropic surfaces.

In addition, there is an apparent problem with CMIP7 simulations at high latitudes (Fig. 3c) where SAOD above 60N seems to be largely overestimated compared to CMIP6. The authors should explain where this comes from.

L320. 0.08 Tg seems to be a very small injection to be considered as a VEI-4. I would say that most of the injections below 0.5Tg would probably be VEI 3.

Figure 4b indicates that the pre-satellite SAOD anomalies has a very different distribution compared to the post-satellite period with many injections below SAOD=0.003. On the contrary, the satellite period indicates that most of the small to medium eruption SAOD peaks between 10^{-3} and 10^{-2}.

L369. What does this sentence mean? EVA_H was not introduced before.

L379. It is a little hard to trust lunar eclipse data. Most of the SAOD derived is overestimated compared to the CMIP6 and CMIP7 datasets, especially from 1750 to 1960.

The SO2 estimates that are provided across the manuscript do not carry any error estimates.

Figure 6. A grid would be helpful on this figure. What is the meaning of showing global SAOD after Krakatoa with only one Pyrheliometer site ?

Figure 8. Add a grid. What is the uncertainty of the searchlight SAOD?

Figure10b. The comparison of datasets in the Nh includes both tropical and NH and polar latitudes. It does not make much sense to compare it with lidar data at 20N from Mauna Loa. CMIP6 and CMIP7 should probably be averaged between +/-10 degrees from Mauna Loa for the comparison to be done.

Figure 10c. SAOD is peaking in July near the North Pole while no eruptions are present. The maximum appears to be between 10-13 km in Fig.10d. What is the origin of this feature ?

On the same figure, the tropopause level should be included as the current figure mixes tropospheric and stratospheric air masses.

L420-440. The comparison of the datasets available after Mt Pinatubo and CMIP6/CMIP7 contains only a few datasets while airborne such as in situ balloon measurements are available from Laramie.

Figure 11. Some lidar data have been used to fill gaps in SAGE II data post-Pinatubo implying relatively close agreements. The evaluation of CMIP6/7 should be done with independent data.

L460. There is no justification about the choice of OMPS-LP product from the University of Bremen compared to the official product provided by NASA. OMPS being a NASA mission, it would seem more logical to utilize the official product even with its current limitations. I would recommend either justifying the choice of the product or including other products available.

Also, I do not believe that OMPS provides either an evaluation of CMIP6 and CMIP7 since most of data used in the two forcings are using the most accurate stratospheric aerosol extinction products available, namely SAGE data. This part is most likely unnecessary and introduce limitations inherent to limb scatter measurements. This is an unnecessary analysis for evaluating CMIP6 and CMIP7 in my opinion.

Figure 12. I'm not sure that the evaluation of CMIP6 and CMIP7 from CREST is necessary and make lots of sense since CREST uses all the data available including SAGE which are already in GloSSAC and therefore CMIP6/CMIP7. This is again a step which in my opinion is not very convincing and should be removed. I do not dispute that a good evaluation between CREST and GloSSAC should be done but likely in a separate paper.

L473. The seasonal cycle in the extratropic is likely associated with the main assumption of limb scatter measurements about particle size distribution. The seasonal variation of the Single Scattering Angle of OMPS likely creates an artificial cycle when higher SSA is associated with lower extinction and vice versa.

Section 3 and section 4 are very poorly written and provide almost no serious evaluation within situ measurements which can be used to retrieved effective radius or number and volume density. In addition, satellite products from SAGE III include a PSD product that could have been used. There are NASA airborne measurements with the ER-2 which could have also being used almost since Agung in 1963.

Figure 16. I'm also very doubtful about the quality of the effective radius derived from SCHIAMACHY (~3km). In addition, the limited vertical resolution of limb scatter makes them very sensitive of ice clouds near the tropopause and the effective radius shown in Fig. 16 could be influenced.

Figure 17. The number density is most likely underestimated given the lack of sensitivity of satellite observations to aerosols with diameter below 0.2-0.3 μm.

Section 4.2. The first sentence of this section should be rephrased. CMIP7 optical properties have to be more accurate during the satellite era as it is constrained by observations.

L625-626. Since limb-scatter rely on assumed PSD, the 525nm-to-1020nm will not carry information other than the assumed PSD. This sentence should be removed.

L632. What does it mean? Why is the eruption delayed? Do you mean that the low spatial sampling of SAGE II created resulted in a delay in the maximum SAOD observed by SAGE III ?

L652-653. What is the point of running them if they carry the same results?

Section 4.4.2. The improvement of non-volcanic background should also include the Asia Summer Monsoon as a source of UTLS aerosols.