Reply to referee comments on egusphere-2025-6257

(Referee original comments are in bulleted italic, and our replies are under each bullet. The line numbers in bracket [ ] are text locations in the revised manuscript.)

Reply to RC1

- *In subsection 2.1, the authors mention the use of several emission datasets to cover the full simulation period of 2000 to 2018. Specifically, anthropogenic emissions from 2015–2018 are taken to be constant based on 2014 data, and volcanic emissions switch from EP-TOMS to Aura-OMI in 2004. While I recognize that this is a standard and necessary strategy given data availability, it is good practice to briefly discuss how these discontinuities might influence the results or cite references that discuss this. Given that a key finding of the manuscript is a statistically significant trend in AEC, could the authors clarify how holding anthropogenic emissions constant for the final four years might influence that trend? A brief comment on the potential artifacts introduced by switching satellite instruments for volcanic data would also be beneficial.*

Reply:

We have addressed the reviewer's comments in the revised manuscript as follows:

Regarding using two different datasets for volcanic emission (TOMS 2000-2003, OMI 2004-2018), we have added the sentence in section 2.1 [line 121–124]:

"The two datasets are considered consistent for long-term records of sporadic eruptions, as OMI was designed to continue the TOMS data record (Thomas and Watson, 2010), although OMI offers superior detection of small-scale degassing due to higher spatial and spectral resolution. Continuous volcanic degassing emissions follow OMI-based estimates for 2005–2015, with the 2005–2015 climatological means of each volcano applied outside of that range."

To test if the "flat" anthropogenic emissions for 2015-2018 may artifact our trend results, we recalculated the trends and p-values using the Mann-Kendall (MK) method while excluding the 2015–2018 period to confirm the trend. We have added the following text in Sect. 3.4 [line 360-368] after presenting the trends and significance in Fig. 10:

"As described in Sect. 2.1, the model simulations for 2015–2018 used CMIP6 anthropogenic emissions in 2014, due to the lack of emission estimates after 2014. To examine the effects of using the constant anthropogenic emissions for the last 4 years in model-estimated trends, we calculate p-values and Sen's slopes with the MK method by removing the 2015-2018 simulations. The resultant statistics remains similar for all variables, e.g., p-value for AEC increases slightly to 0.023 and, as expected, the trend increases to 0.018 10^3 km^{-1} yr^{-1}. At the same time, changes of correlation coefficients between AEC and anthropogenic emissions in Asia are within ± 0.01, and AEC still does not correlate with the transport and removal tracers. These results indicate that the increase in non-volcanic AEC in the ASM upper troposphere in recent decades has been determined primarily by the growth of anthropogenic emissions in Asia, and the future changes of such trend, either increase or decrease, can be projected with the forecast of future anthropogenic emissions."

- *On lines 304–307, the authors state that R^2≈0.50 is "consistent with a simple summation of the CV values… which is roughly half of the 79% CV." This comparison is mathematically* *difficult to justify. Coefficients of variation, which are derived from standard deviations, do not sum linearly. For independent variables, variances add linearly, while standard deviations add as the square root of the sum. R^2 represents the fraction of explained variance, whereas CV represents a fraction of standard deviation. An R^2 of 0.50 implies that ~71% of the standard deviation is explained, which contradicts that the summed CVs are only half of the total. I suggest removing this specific sentence as the result of R^2=0.505 stands strong enough on its own without this non-rigorous comparison.*

Reply:

We appreciate the reviewer's correction regarding the linear summation of CVs. We have removed the non-rigorous comparison sentence from the manuscript as suggested.

- *There is no clear description of which statistical test was used to determine the statistical significance of the increasing AEC trend in subsection 3.4. Based on the caption of Fig. 10, I conclude that a Student's t-test was used. The standard Student t-test relies on assumptions that may not hold for this data, particularly the assumption of independence. Time series of aerosol extinction coefficients often have autocorrelation. Could the authors clarify exactly which type of t-test was used and if any adjustments were made to account for autocorrelation? If a standard Student t-test was used, I suggest performing a more robust test, such as the Mann-Kendall test, to confirm the results.*

Reply:

We thank the reviewer for this excellent suggestion. To ensure our results are not biased by autocorrelation, we have replaced the Student's t-test with the Mann-Kendall test for all trend analyses in the main text [line 342–347] and in Figure 10. Our results remain consistent across both methods, confirming the statistical significance of the increasing AEC trends.

- *The authors correctly note in subsection 3.1 that satellite retrievals below the tropopause are limited by cloud presence. However, deep convection in the ASMA region can also reach the tropopause (100 hPa). I would thus expect that even at the studied height of Fig. 2 the satellite data misses some times when clouds are present, whereas the models likely span the full period. Was any filtering applied to the model data to match the satellite data availability? Otherwise, this discrepancy in data availability could bias the comparison.*

Reply:

We have added clarifying text to Sect. 3.1 on spatiotemporal match between satellite data and model [line 184-188]:

"Considering that the L3 satellite data represent the monthly composite of all available measurements falling within a particular month in the locations bounded by the spatial grids while the model results are monthly means of spatially and temporally continuous simulations without coverage gaps, the comparisons shown in Fig. 2 and 3 should be considered in a broad sense, such as magnitudes, seasonal cycles, and interannual variations, rather than absolute numerical agreement."

Reply to RC2

- *I found some sections of the introduction a bit disjointed, and the authors could do a better job offering an overview of the topic, expanding on them a bit. For instance, the phrase "Aerosols in this region affect radiative forcing, cloud microphysics, and chemical composition" is a very generic and bland statement. It's true, but it's also true for aerosols everywhere! A better argument would highlight the difference compared to lower tropospheric aerosols better (for instance, discussing lifetime and sources better), as well as to stratospheric aerosols.*

Reply:

We have made substantial modifications in the Introduction section. For example, here is the new first paragraph [line 46–53]:

"The upper troposphere is a crucial region of the Earth's atmosphere, acting as an efficient pathway for material originating in the lower troposphere to spread across hemispheric scales and even into the lower stratosphere. Although aerosol concentrations in the upper troposphere are generally much lower than in the boundary layer, their atmospheric lifetimes are longer, and their radiative and chemical influences can be disproportionately large (Kärcher, 2012; Boucher et al., 2013). One key role of upper tropospheric aerosols is serving as ice-nucleating particles that affect the formation and microphysical properties of cirrus clouds, which exert a net warming effect on the climate system by trapping outgoing longwave radiation (Kärcher and Lohmann, 2002; Kärcher, 2017). Aerosols also provide surfaces for heterogeneous chemical reactions leading to ozone destruction in the lower stratosphere (e.g., Solomon, 1999)."

- *Similarly, the section about AeroCom could be better explained, with some references and examples to the finding it produced [as well as, for instance, including some references and discussion to similar model intercomparison projects. For instance, the ISAMIP background sulfur budget analyses in Brodowsky et al. 2024 might be particularly relevant for this work, see the comment in line 385 and 467].*

Reply:

Regarding AeroCom: We have expanded the description of AeroCom in Introduction by adding the following text and refer the results from many previous AeroCom studies to the publication list on the AeroCom website in the revised paragraph [line 88–93]:

"Since its inception in 2002, AeroCom has organized numerous multi-model experiments, targeting various topics including aerosol composition, chemical, physical, and optical properties, trends, radiative forcing, aerosol-cloud interaction, impacts on environment and climate, etc. (see https://aerocom.met.no/publications/ for published results from previous AeroCom studies). AeroCom is also in alliance with another international initiative, the International Satellite Aerosol Science Network (a.k.a. AeroSat, https://aero-sat.org/), since 2013 to foster a close collaboration between the global modelling and observation communities."

Considering ISAMIP: We carefully considered including the findings of Brodowsky et al. (2024). However, as ISAMIP focuses on the stratospheric background sulfate mass budget while our study focuses on upper-tropospheric aerosol extinction within the Asian summer monsoon anticyclone, the overlap is limited. To maintain a concise narrative focused on ASM dynamics, we have elected not to discuss ISAMIP in detail, though we have included a broader reference to the AeroCom publication list to provide readers with more intercomparison context.

- *Line 121: "The prescribed sources for the tracers were used repeatedly for all simulated years (2000-2018)" Not sure what this means?*

Reply:

We have modified the description of tracers to clarify the tracer sources. The sentence has been changed to [line 131–134]:

"Large-scale transport is evaluated using a CO-like tracer, TR_{CO50}, with prescribed monthly CO anthropogenic and biomass burning emissions from 2010 CMIP6 dataset and specified secondary CO production sources from methane and non-methane VOC oxidations (see Supplement section S1). The same annual cycle is repeated each year during 2000–2018."

We have also added tracer description in Table 1. In addition, we provide more detailed description and implementation of tracers in Supplement Sect. S1.

- *Line 136: non-permanent links such as the two in this section should not be used in a paper that will be permanent. What if the link changes 5 years from now? If you want a link to external info, it will either have to be permanently archived somewhere, or included as supplementary here. Or just repeat the information somehow.*

Reply:

We have expanded the model experiment description in section 2.1 and added information on emission in Table 1. In addition, we provide the Supplement to describe in more detail about the transport and removal tracers [Sect. S1]. These changes will remove the dependence of this information from external links and make them a part of permanent record in the paper.

- *Table 2: I was surprised that there is no mention in the methodology or in the discussion about the modeling differences in the meteorology. A replay is very different from a nudged simulation, and those are very different in terms of "realistic" conditions from fixed SST, AMIP style simulations, especially when comparing to observations. The authors should make sure to acknowledge these differences and integrate them in the discussion phase of the manuscript (see for instance Orbe et al., 2016). On a similar note, it would be useful to have information in the table about the aerosol microphysics scheme used, since that is then discussed in 318-320.*

Reply:

We have made several modifications to address the referee's comments, as follows.

We have updated Table 2 to include a dedicated column for the aerosol microphysical scheme (Bulk vs. Modal/Sectional). We have also added a description of meteorological fields and aerosol microphysics in Section 2.2 and Section 3.3.3 regarding these configurations:

In Sect. 2.2, model configurations, the following text is added [line 152–160]:

"The experimental ensemble represents two distinct approaches to meteorological forcing: Reanalysis-forced and General Circulation Model (GCM)-driven. While three models (CIESM-MAM7, GFDL-fSST, and ECHAM6-HAMMOZ) are free-running GCMs forced by observed sea surface temperatures (SST), the majority of the models (six) are driven by reanalysis-based meteorology, including MERRA-2 (used by GEOS-i33p2, CISS-OMA, GISS-MATRIX, CAM5-ATRAS), National Center for Environmental Prediction (NCEP) (used by GFDL-nSST), and ECMWF Atmospheric Reanalysis Version 5 (ERA5) (used by MIROC-SPRINTARS).

Aerosol microphysics schemes also differ. Five models use bulk aerosol schemes simulating mass concentrations with fixed size bins and considering particle growth as a function of ambient relative humidity, whereas four models employ more advanced microphysical schemes (either modal or sectional) for treatment of aerosol mixing states and particle size distributions."

In Sect. 3.3.3, we have leveraged the "paired" model configurations to infer the effects of meteorology and aerosol microphysical schemes in reference to Fig. 9, although we cannot disentangle these effects across the entire ensemble. Specifically, we use the pair of GISS models to show the effect of microphysics scheme and the pair of GFDL models to discuss the influence of nudged vs. free-running GCM meteorology, described in the following new paragraphs [line 314–326]:

"Figure 9 also illustrates the impacts of different aerosol microphysics schemes and meteorological forcing on model-simulated AEC, using results from two GISS and two GFDL model configurations. As detailed in Table 2, GISS-OMA and GISS-MATRIX share the same meteorological forcing (MERRA-2) but differ in aerosol microphysics schemes: GISS-OMA uses a bulk scheme assuming external mixtures of aerosol species, while GISS-MATRIX utilizes a modal microphysics scheme that tracks both aerosol mass and number concentrations and explicitly simulates the aerosol mixing state (Bauer et al., 2022). Consequently, both models yield similar TR_{CO50} and $TR_{Pb/Rn}$ values, as these are primarily meteorology driven. However, GISS-MATRIX produces nearly three times lower AEC than GISS-OMA, a discrepancy arising largely from their different representations of aerosols microphysical properties and associated processes, and $TR_{Pb/Rn}$ may not be suitable representing the aerosol removal for GISS-MATRIX. Meanwhile, the pair of the GFDL models adopts the same bulk aerosol scheme but differs in meteorological forcing: GFDL-fSST uses meteorology generated by the underlying GCM forced by SST, whereas GFDL-nSST incorporates additional "nudging" toward NCEP reanalysis (Table 2). Although their overall results are comparable, AEC from GFDL-nSST is 22% lower than GFDL-fSST, affected by higher wet scavenging efficiency (9% higher $TR_{Pb/Rn}$) and slightly lower transport efficiency (2% lower TR_{CO50}) in the nudged configuration."

In Sect. 4 (Discussion), we have added a paragraph in discussion about separating the effects of meteorology and model parameterization [line 422–430]:

"Broadly, two primary factors govern the simulated transport and removal processes: the representation of parameterized physical processes (e.g., convective transport, wet scavenging) and the dynamic meteorological fields (e.g., winds, precipitation) driving these processes. The current experimental design does not allow separation of the inter-model differences arising from model parameterizations versus meteorology; rather, it diagnoses their compounded effects. Future model experiments, such as prescribing a unified meteorological forcing across all participating models, would better differentiate the variance associated with large-scale atmospheric flow from that stemming from sub-grid scale parameterizations. Previous work has shown that even with identical meteorological forcing, inter-model differences in parameterized convection can substantially alter the transport characteristics (Orbe et al., 2017). Adopting similar "constrained-meteorology" framework in future AeroCom experiments would allow more in-depth diagnostics and lead to improvements of parameterizations among models."

- *Figure 2-3-4-6: These figures are all too small, and make it very hard to distinguish the results, especially in a quantitative way. Some of them (especially the maps) could be more zoomed in, whereas the full map could stay in the supplementary.*

Reply:

We have remade Figures 2 through 8, transitioning from a three-column to a two-column layout to significantly enlarge the individual panels. While we appreciate the suggestion to zoom in, we have elected to retain the global (60S-60N) maps in the main text to clearly illustrate the broad-scale impact of ASMA transport.

- *I strongly agree with reviewer 1 about lines 304–307.*

Reply: We have removed the sentence as suggested by reviewer #1.

- *Section 3.4: Considering some models in Figures 2 and 3 do not capture aerosols trends and contributions at all, I wonder if the authors could show also, and comment on, how the trends look like if such models are de-weighted or removed altogether from the multi-model average. Some comments also in the conclusions about how to leverage the model-observation evaluation to differentially weight the results could be of interest to the community (even in the presence of observational uncertainties).*

Reply:

We thank the reviewer for this sophisticated suggestion. Although we agree that a weighted multi-model ensemble is a desirable goal for the community, we have decided to maintain an unweighted ensemble in this study because we are focusing on the non-volcanic aerosols that is not directly observable. That means we don't have a reliable standard to decide which model(s) should be excluded. In addition, as discussed in the manuscript, satellite retrievals in the upper troposphere carry large uncertainties due to cirrus cloud interference, making them an imperfect "calibration standard" for model weighting.

We have added a paragraph in the Discussion section [line 404–408]:

"The large model diversity in AEC may hamper the robustness of the two-decadal multi-model ensemble AEC trend in the upper tropospheric ASMA region presented in Fig. 10. While differentially weighting models based on observational agreement (e.g., Brunner et al., 2019) could potentially refine the multi-model ensemble means by excluding outliers, our focus on the tropospheric-originating, non-volcanic aerosols that are not directly observable, thus making an unweighted ensemble approach more reasonable for this study."

- *Line 405: I found this part quite vague. Yes, the tropopause height can be diagnosed in different ways, but are you saying that different models have different ways to output it? In the Chemistry-Climate Model Intercomparison (CCMI), this was often solved by re-calculating the tropopause in a coherent way (see for instance Section 2.2 in Pisoft et al., 2021).*

Reply:

We've significantly revised the "Tropopause height" discussion. Yes, models can have different ways to define the tropopause height, which is usually determined by the underlying meteorological model. Regarding recalculating the tropopause height, we are unable to implement a unified recalculation in this study due to limitations of the available fields. We have explicitly cited Pisoft et al. 2021 to acknowledge the methodology gap. Below is the revised text in Section 4 (discussion) on that topic [line 434–450]:

"**Tropopause height:** Figures 5, 7, and 8 have revealed substantial inter-model variability in tropopause height during August, with differences of several tens of hPa. For instance, the GISS models consistently show the highest tropopause altitudes, whereas the GFDL models sharply drop the tropopause to lower altitudes over the ASM core (~90°E). These disparities reflect differences in underlying meteorology as well as variations in tropopause diagnosis methods across modeling groups. Tropopause height is typically identified using several distinct criteria including thermal definition based on the temperature lapse rate

(World Meteorological Organization, 1957), dynamical definition using specific potential vorticity thresholds (e.g., Hoinka, 1998), and composition-based definition derived from the sharpness of vertical gradients in trace gases such as ozone or water vapor (e.g., Shepherd, 2002). In the current study, we are unable to resolve these inter-model differences by imposing a unified definition or recalculating a coherent tropopause for each model (e.g., Pisoft et al., 2021) from the model outputs.

Notably, the tropopause height is a diagnostic quantity that does not influence the prognostic physical processes evaluated here. The inter-model spread in August mean tropopause height does not appear to correlate with the AEC differences in the upper troposphere, confirming that the large variability in ASMA aerosol loading is primarily driven by internal physical processes rather than differences in the geometrical definition of the tropopause boundary. However, the vertical placement of tropopause has directly impact on stratosphere-related quantitative assessments including stratosphere-troposphere exchange, composition in the lower stratosphere, and integrated stratospheric column quantities such as stratospheric aerosol optical depth (e.g., Millán et al., 2024). Consequently, users engaged in those studies must exercise caution, as the perceived "stratospheric" or "tropospheric" character of a species is highly sensitive to the tropopause altitude."

- *Data Availability statement: "The AeroCom UTLS model output is stored in the AeroCom repository, which can be accessed on request, as described at https://aerocom.met.no/data/" I couldn't find this when reading the page linked, and considering the last update to that page is from 2021, I would suggest updating it to make it more usable. I would also note that, as is, this kind of statement is not in line with ACP editorial policies (https://www.atmospheric-chemistry-and-physics.net/policies/data_policy.html). At the very list, the authors should make the final output as analyzed here available in a repository with a DOI.*

Reply:

We acknowledge the importance of stable, citable data access. The challenge in this specific multi-model study is that the datasets used (emissions, model outputs, and satellite retrievals) are from over 10 different international organizations, each with its own institutional data-sharing protocols. At this stage, we are unable to aggregate all data into a single public archive like Zenodo without multi-institutional approval.

Meanwhile, we have updated the Data Availability Statement to provide direct, stable links for data access, and have confirmed with the individual PIs that these datasets will remain accessible for the foreseeable future. We believe this approach meets the intent of the ACP policy given the collaborative nature of the AeroCom project. In addition, we include in Table S2 in Supplement the multi-model ensemble means of model-simulated AEC, TR_{CO50}, and $TR_{Pb/Rn}$ over the ASMA region at 150 hPa for each August during 2000-2018, together with the corresponding standard deviation and CV, to make them directly accessible through egusphere or ACP.