Reply to comments of RC1:

This manuscript presents a new setup of the EMAC model that allows to seamlessly simulate tropospheric and stratospheric aerosols. Compared to the two standard EMAC versions (one for the troposphere and one for the stratosphere), this new version includes an additional "small coarse" mode for particles between 800 and 3200 nm diameter. The authors evaluate extensively the new model setup against air-, ground-, and satellite-borne observations.

We thank the referee for the time to review our manuscript, and the constructive feedback. We report the comments (grey, bold) along with our replies (blue). Text changes are provided in red.

This is an very good model evaluation paper. The new setup is generally well described, the comparison with observations is detailed and rigorous, the text is clear, and the figures are meaningful and very well thought out (I really appreciate how they deliver a lot of information while remaining clearly legible). If this paper was submitted to GMD, I would only have minor comments to improve the readability of the manuscript (see below). However, I don't think this paper is within the scope of ACP ("Articles should have important and clearly argued implications for our understanding of the state and behavior of the atmosphere and climate or present substantial new insights into the atmosphere's role in other parts of the Earth system.").

We thank the referee for the generally positive assessment of the model description, evaluation strategy, clarity of presentation, and figures. We understand the concern that the manuscript contains a substantial model-development and evaluation component. However, we argue that the central contribution of this work extends beyond model configuration: the new, extensively evaluated framework is applied to address atmospheric-science questions concerning the global vertical structure, composition, and sources of aerosols from the boundary layer through the UTLS and into the stratosphere.

As noted by the referee, ACP articles should provide "important and clearly argued implications for our understanding of the state and behavior of the atmosphere" or "substantial new insights into the atmosphere's role in other parts of the Earth system."

To our knowledge, the manuscript presents the first consistent, global and size-resolved analysis that jointly examines aerosol mass, number concentration, composition, and size distribution across atmospheric regimes that are usually considered separately. Therefore, not only it improves the understanding of the interplay of chemical species and particle number concentrations on a global scale and in vertical transport, but also reveals new crucial aspects of aerosols transport in the UTLS, a generally difficult region to simulate by the model. Therefore, our work provides a benchmark for future model intercomparisons and observational campaigns targeting this atmospheric region. Specifically, the analysis results in several relevant atmospheric findings:

1. A global aerosol mass minimum in the 400–200 hPa layer, identifying an important transition region in aerosol sources and transport between the free troposphere and UTLS.
2. A pronounced enhancement of particle number concentrations in the UTLS, previously

documented only in tropical/subtropical regions (Curtius et al., 2024; Weigel et al., 2021; Williamson et al., 2019), is shown here to be a global feature across all latitudes.

3. A substantial organic aerosol burden in the lower stratosphere, suggesting a more complex chemical stratospheric composition than often assumed, and in line with recent observational evidence (Hodzic et al., 2020; Lyu et al., 2026; Murphy et al., 2014).

4. A lower contribution of boundary-layer-emitted primary particles, particularly dust, to the stratospheric burden, in closer agreement with recent observational constraints.

These findings have implications for the representation of aerosols in climate models, the interpretation of satellite and aircraft observations, and the assessment of aerosol–cloud interactions and radiative effects. To facilitate such future work, we will additionally make the global simulations publicly available as monthly time seried fields of aerosol composition and number concentration for 2016–2019.

We emphasize that the comprehensive evaluation is not merely a methodological exercise. Instead, it provides the basis for the presented atmospheric interpretation, identifies the strengths and limitations of the simulations, and thereby guides the application of the resulting global aerosol dataset. Although the manuscript includes substantial model development and evaluation, these aspects are not its sole scientific endpoint. The primary contribution is the observationally evaluated characterization of global aerosol structure, composition, and vertical distribution from the surface to the stratosphere.

In summary, we believe that the combination of model simulations with observational evidence in the way presented here is essential for a better understanding of aerosol processes in the troposphere and stratosphere, and how they are interconnected. To clarify this point, we applied some changes to Abstract, Introduction and the Conclusions, so that the scientific results are better highlighted.

The abstract was reformulated, such that the methodological part was slightly shortened (while maintaining the original content) and giving more emphasis to the results:

" Atmospheric aerosols play a crucial role in Earth's climate system, yet their distribution in the free troposphere (FT) and upper troposphere–lower stratosphere (UTLS) remains poorly constrained, a major source of uncertainty in aerosol radiative forcing estimates. Here, we present ECHAM/MESSy Atmospheric Chemistry (EMAC) model simulations with a newly developed setup, bridging the tropospheric and stratospheric regimes, evaluated against a comprehensive suite of aerosol mass, number concentrations, and optical property observations, with good agreement across vertical layers and most regions. The simulations provide a unified description of global aerosol composition, mass, and number concentrations from the Earth's surface to the stratosphere. Simulated aerosol mass exhibits a global minimum between 400 and 200 hPa, marking the transition between FT and UTLS. Particle numbers peak at similar or slightly higher altitudes, previously reported only regionally in the tropics. Primary material contributes less than 3.5% to aerosol mass in the stratospheric overworld up to 10 hPa, substantially less than suggested by previous modelling studies and in closer agreement with recent observations. Stratospheric aerosol mass is dominated by sulfate, with a notable contribution ($\sim$ 15%) from secondary organic aerosol throughout the global lower stratosphere, indicating a more chemically complex stratospheric aerosol than commonly assumed. This work resolves previously unconstrained features of aerosol structure in the FT and UTLS, with direct impli-

cations for aerosol representation in climate models, satellite and aircraft data interpretation, and the assessment of aerosol-cloud interactions. ”

We added the following paragraph in the introduction, before the last section roadmap paragraph:

” Beyond providing a unified modelling framework, the resulting dataset allows us to address several open questions in global aerosol research: how aerosol mass and number concentrations are structured across the transition between the FT and the UTLS, the extent to which secondary organic aerosol rather than sulfate alone shapes lower-stratospheric composition, and the magnitude of primary mass intrusion into the stratospheric overworld relative to previous simulations and recent in-situ constraints. The corresponding model output is made publicly available to support such applications (see Data availability). ”

We additionally reformulated the first paragraph of the conclusions:

” Using a newly developed and comprehensively evaluated EMAC model setup unifying previous tropospheric and stratospheric configurations, this study provides a globally consistent, observationally constrained characterization of aerosol distributions and composition from the Earth’s surface to the stratosphere, revealing several previously unresolved features of aerosol structure in the FT and UTLS. ”

Additionally, we added in the last paragraph of the conclusions:

” This study demonstrated the EMAC model’s capability to simulate atmospheric aerosols from the surface to the stratosphere, providing first consistent global estimates of key aerosol species across this entire domain, comprehensively evaluated against in-situ aircraft and stationary, as well as remote sensing observations. These simulations offer new insights, especially into the FT and UTLS, regions previously underrepresented in model studies. The newly developed model setup enables future process studies on critical phenomena such as new particle formation in the UT, troposphere–stratosphere exchange, and the transport of wildfire emissions, volcanic emissions, and spacecraft debris, all of which play key roles in shaping atmospheric composition and climate. To support such future studies, the underlying model output, i. e., monthly mean three-dimensional fields of aerosol composition and number concentration for 2016–2019, is made publicly available (see Data availability). ”

In the Code and Data availability, we now state:

” Model output from this study (TS setup), comprising monthly mean three-dimensional fields of aerosol mass concentration and composition as well as particle number concentration, is publicly available via `https://doi.org/10.17617/3.DOM8G4` (Kohl and Pozzer, 2026). Data are provided in NetCDF format on the native T63 model grid (1.875° × 1.875°), regridded to 20 and 4 pressure levels separately, in agreement with the figures and the analysis in the main text, as a monthly time series from January 2016 to December 2019. As the underlying meteorology is nudged towards ERA5 reanalysis and observed QBO winds (Sect. 2.2), individual months reflect the simulated aerosol state under the actual observed dynamical conditions of that period, rather than a long-term climatological mean. ”

General comment: Throughout the manuscript, the references to the relevant figures are generally at the beginning of the section that discuss them. Since each figure has many panels, it would be clearer if the panels were numbered and the

text contained the reference to the panels where each of them is discussed. I often had trouble understanding which figure I was supposed to look at while reading the text.

We thank the referee for this comment and agree, that the references to the figures are sometimes unclear, especially given the large number of panels in several figures. We labelled the panels in most figures (excluding the gridded ones), and more consistently referenced the discussed figures and panels.

L165: is half-year spin up enough for the stratosphere? If starting from a clear atmosphere, the spin up for the stratosphere is generally several years.

We apologize for the confusion here and agree that stratospheric equilibration typically requires several years. At the end of the paragraph, we additionally stated in the submitted manuscript: "Tracer initialization is based on ten-year simulations to ensure the establishment of stratospheric background aerosol equilibrium". So our spin-up substantially exceeds stratospheric equilibrium timescales. The spin-up is actually threefold:

1. We use OCS mixing ratios from 30 year-simulations by Schallock et al., 2023, which have been initialized with OCS mixing ratios from a multi-decadal two-dimensional (altitude-latitude) simulation. This ensures, that a stratospheric equilibrium state is reached for long-living OCS, important for background sulfate formation (e.g., Brühl et al., 2012).

2. All three simulations with the different setups were initialized from a clean-atmosphere state (with the exception of long-living gases such as OCS, which is are initialized from the Schallock et al., 2023 simulations), and we performed 10-year spin up-simulations, to ensure an aerosol equilibrium up to the stratosphere

3. The actual simulation from 2016 to June 2020 is then preceded by a half-year spin-up simulation (starting July 2015), with all tracers initialized from the 10-year spin-up simulation for the atmospheric state to adjust to the specific state of that period. The two other mentioned periods are also initialized from this 10-year simulation.

Taken together, these three initialization steps ensure that both short- and long-lived species reach stratospheric equilibrium well before the analysis period.

We adjusted the paragraph as follows (note that we extended the main simulation to end of June 2020):

" Short-lived gases and aerosols are initialized from ten-year spin-up simulations for all model setups to ensure the establishment of stratospheric background aerosol equilibrium, with long-lived trace gases initialized from previous multi-decadal simulations (Brühl et al., 2012; Schallock et al., 2023). The main simulation extends over the period from January 2016 to June 2020, following an additional half-year spin-up period (starting July 2015) to allow the atmospheric state to adjust to the conditions of that period. An additional shorter simulations was conducted for the period January 2022 to January 2023 (for comparison with the ACCLIP and CAFE-Brazil campaigns), initialized using the same ten-year spin-up simulations. "

L177: I am a bit confused by the description of the aerosol speciation. Are the concentrations of sulfate, ammonium, nitrate, sodium, chloride, potassium, calcium, magnesium, organic aerosols (separated in 26 species), black carbon, dust, and sea salt prognostically calculated for each of the modes (5 soluble and 5 in-

soluble modes)? Are soluble species allowed in the insoluble modes to simulate coating?

We apologize, if the modal setup was not outlined clearly enough here. The model contains in total 9 modes (soluble nucleation, Aitken, accumulation, small coarse, and coarse mode; and insolube Aitken, accumulation, small coarse, and coarse mode). There is no insoluble nucleation mode, as all nucleating species are soluble.

Soluble species are not allowed in the insoluble mode, as insoluble species can only be transferred to the soluble mode. As already stated in Sect. 2.2, this transfer occurs via coagulation with soluble particles or coating by condensing soluble species, following (Pringle et al., 2010). To make this clearer, we provide an additional table in the revised manuscript, describing which species are treated in which mode (Table 1 here). Additionally, the table provides the processes relevant for mass accumulation in the respective mode.

Table 1: (Table 2 of the revised manuscript) Species to mode mapping with the relevant processes for mass and number accumulation in the respective modes: primary emission; nucleation; growth- or evaporation-driven transfer from neighbouring modes; inter-modal coagulation; ageing via soluble coating (transfer to the corresponding soluble counterpart via coating or coagulation with soluble particles). A dash (–) indicates the pathway is not applicable to that mode. "Speciated" contains all speciated organic and inorganic aerosol.

Mode	**Species**	**Emiss.**	**Nucl.**	**Grow./Evap.**	**Coag.**	**Coat.**
Soluble						
Nucleation	speciated	–	x	–	–	–
Aitken	BC, speciated	POA, SO_4	–	x	x	x
Accumulation	SS, BC, speciated	POA, SO_4, SS	–	x	x	x
Small coarse	SS, BC, DU, speciated	SS	–	x	x	x
Coarse	SS, BC, DU, speciated	SS	–	x	x	x
Insoluble						
Aitken	BC	BC	–	–	–	–
Accumulation	BC	BC	–	–	x	–
Small coarse	DU	DU	–	–	x	–
Coarse	DU	DU	–	–	x	–

We added to the end of Sect. 2.2.1:

"Table 2 summarizes the species and pathways associated with each mode. All speciated soluble species can form in any soluble mode, while bulk SS, BC, POA, a small fraction of sulfate, and DU are introduced via direct emission. Freshly nucleated particles enter the nucleation mode, and particles can generally grow or shrink into neighbouring modes through condensational growth or evaporation; coagulation likewise places the resulting particle in the respective larger original mode. Upon coating with soluble material or coagulation with soluble particles, insoluble particles, i. e. dust and BC, transfer to the soluble mode. Within each mode, particles are internally mixed, while the modes themselves remain externally mixed with respect to one other (more details in Pringle et al., 2010). Fig. S1 and S2 present typical particle size distributions at the surface and in the stratosphere, respectively."

L230: I think it should be specified that the past studies are only the ones with

EMAC.

We agree with the referee and therefore the text was changed as follows:

L230: "Past EMAC studies ..."

L238: "Stratospheric EMAC studies ..."

L250: which parameterization is used for sedimentation? Is it the same for all modes (just varying the radius)?

Sedimentation is computed in two steps. First, a settling velocity is calculated for each mode using a Stokes–Cunningham approach as described by Kerkweg et al., 2006. Second, this velocity is used to vertically transport the tracer mass/number using the monotonic advection-type scheme of Walcek, 2000, adapted for sedimentation as described by Benduhn and Lawrence, 2013). This numerical transport scheme itself is identical across all modes and species, and only the input settling velocity (via mode diameter and species-dependent density) differs.

We added the following sentence:

" Sedimentation velocities are calculated for each mode based on a Stokes–Cunningham approach, accounting for mode diameter and density, with vertical transport computed using the monotonic advection scheme of Walcek, 2000, adapted for sedimentation (Benduhn and Lawrence, 2013; Kerkweg et al., 2006).

L265: I don't understand what "respectively" refers to. Maybe a word is missing?

We reformulated to:

"To ensure a direct and meaningful comparison between model results and observations, we integrate either the particle number size distribution (for number-concentration instruments) or the particle volume size distribution (for mass-concentration instruments) over the same lower and upper size limits used in the respective measurements (see also Kohl et al., 2023)."

L291: I am not sure I understand what is done to compare with aircraft measurements. Shouldn't the aircraft measurement (not the model) be binned on the model vertical levels? Is the vertical distance between consecutive aircraft measurements larger than one model level?

Regarding the first question: We chose not to bin observations onto the model's native vertical (hybrid-sigma) levels, since these vary in spacing with altitude and geographic location, and are not a stable reference grid across different regions. Instead, we extract the full model column at native vertical resolution at each model timestep via the S4D submodel, and subsequently aggregate both model and observational data onto common, fixed 1 km altitude bins. Because the full column is retained prior to binning, no vertical information is lost in this step, and the comparison treats model and observations symmetrically.

The vertical distance between consecutive aircraft measurements typically is not larger than one model level. However, the vertical distance the aircraft travelled during one model timestep (7.5 minutes) can be substantially larger than one model level. We also state that in the submitted manuscript in line 291: "Aircraft measurements, typically available at sub-minute temporal resolution and spanning a significant vertical range within a single model timestep, ..."

We adjusted the paragraph for better understanding:

" Aircraft measurements are typically available at sub-minute temporal resolution and often span a significant vertical range (covering several model levels) within a single model timestep. To account for this, the full vertical column of model output is sampled along the flight track at each model timestep using the MESSy S4D submodel (Jöckel et al., 2010). The resulting model data are then binned by altitude for each time step and geographical location, and subsequently spatio-temporally collocated with the corresponding averaged observations to enable a consistent comparison, while fully exploiting observational data from aircraft ascents and descents. "

L315: I am surprised that both TROPO and STRATO overestimate NO3, and TS does it better. Is there an explanation for why?

We thank the referee for this question, so that we can also clarify this in the manuscript. The NO_3 overestimation is consistent with (and likely driven by) the overestimation of sea salt (Na^+/Cl^-) in these setups. Excess sea salt mass shifts the ISORROPIA-II thermodynamic equilibrium (Fountoukis and Nenes, 2007) toward $NaNO_{3(aq)}$ formation from gas-phase HNO_3, releasing HCl to the gas phase. This process is also well documented observationally as sea-salt "chloride depletion" (Dasgupta et al., 2007; Gard et al., 1998; Guimbaud et al., 2002). This is supported by the matching ranking of overestimation severity across setups for Na^+, Cl^- and NO_3; while NH_4 which is not affected by sea-salt, shows comparatively little variation across setups, indicating a sea-salt-specific rather than a general aerosol-lifetime effect.

In particular, HNO_3 can heterogeneously react with sodium chloride (NaCl):

$$HNO_3(g) + NaCl_{(aq)} \rightleftharpoons NaNO_{3(aq)} + HCl(g) \tag{1}$$

We moved the original sentence after the discussion of the sea salt representation and added the following sentence:

" The overestimation in NO_3^- concentrations in the TROPO and STRATO simulations, also present in earlier evaluations using the TROPO modal setup (e.g., Pozzer et al., 2022), is significantly reduced in the TS setup, showing improved agreement with observations, although a residual overestimation remains. This behavior is closely linked to the sea salt sodium and chloride bias (see above): unlike sodium, which remains almost entirely in the particulate phase, chloride can be lost to the gas phase via acid displacement reactions with nitric and sulfuric acid (e. g., $HNO_3(g) + NaCl_{(aq)} \rightleftharpoons NaNO_{3(aq)} + HCl(g)$), shifting the ISORROPIA-II thermodynamic equilibrium toward particulate nitrate formation. This chloride depletion process is well documented observationally (Dasgupta et al., 2007; Gard et al., 1998; Guimbaud et al., 2002; Su et al., 2022) and typically produces a stronger marine-to-continental gradient for chloride than for sodium (Fig. S4). Our simulations reproduce this gradient only partially, consistent with the matching cross-setup ranking of sea salt and NO_3^- overestimation (Table 3), and indicating that the model underestimates the extent of chloride depletion during transport. The co-located overestimation of ammonium (Fig. 2) may additionally promote excess ammonium chloride formation, further contributing to the chloride bias. An additional complicating factor is that ISORROPIA-II assumes instantaneous equilibrium for the entire particle, which is less justified for the (comparatively coarse) sea-salt-dominated modes, where in reality only the particle surface interacts with the gas phase. "

L335: I don't think MERRA2 is a good benchmark to compare speciation. In MERRA2, the total (not speciated) AOD is assimilated, but the speciation is calculated based on the one of the underlying forecast model (GEOS). If GEOS gets the speciation incorrectly, this error will remain after the assimilation adjust the AOD.

We agree, that both MERRA-II and CAMS are only reanalysis datasets, and biases in the speciation will directly translate into the product. This comparison should not be a real evaluation with "Observations", but rather a comparison with reanalysis data, showing the agreement and disagreement. In any case, it shows that results from the TS simulation better agree with both reanalysis datasets (MERRA-II and CAMS) than the results from the STRATO and TROPO simulations.

We reformulated this to avoid confusion.

Main text changes (note that some additional changes included here result from comments of referee 2):

" The representation of sea salt sodium and chloride is significantly improved in the TS setup. However, sea salt concentrations remain overestimated relative to observations. A detailed comparison between modelled sea salt and reanalysis data from CAMS (Inness et al., 2019) and MERRA-II (Gelaro et al., 2017; Randles et al., 2017) indicates that the applied changes reduce the long-range transport of sea salt particles to continental regions, sharpening the sea salt gradients between oceanic and continental areas compared to simulations with previously used setups. This results in substantially improved agreement with reanalysis data across all size ranges (Appendix A and Fig. A1. Note that both MERRA-2 and CAMS assimilate total column AOD rather than speciated aerosol observations. The partitioning into individual components such as sea salt is therefore determined by the aerosol scheme of the respective underlying forecast model (GEOS-5 for MERRA-2; IFS-AER for CAMS) and is not directly constrained by the assimilated observations. Agreement with these products should thus be interpreted as consistency between independent aerosol model representations rather than validation against observationally constrained speciation (see also Appendix A). "

Appendix A changes:

" Importantly, these reanalysis datasets are not direct observations but represent pseudo-observations, as they are derived from model simulations and rely on specific assumptions and parameterizations during the assimilation process. In particular, since both products assimilate total column AOD (e.g., from MODIS) rather than speciated aerosol properties, the sea salt mass fraction in MERRA-2 and CAMS is not independently constrained by observations but instead reflects the aerosol speciation scheme of the respective host model (GEOS-5 and IFS-AER, respectively). Consequently, the comparison presented here assesses consistency with these models' speciation rather than providing independent validation of simulated sea salt concentrations. "

Table 2: generally, M/O is not bad, but R2 is quite low. Can the authors draw any conclusion from this? Could it be that compensating errors in different regions cancel each other, giving a good globally averaged M/O but making the R2 low? It would be interesting to see a map of M/O at each GHOST location: maybe it could show if there are systematic under/overestimation based on the aerosol species

dominating PM and give indications on which processes are better simulated.

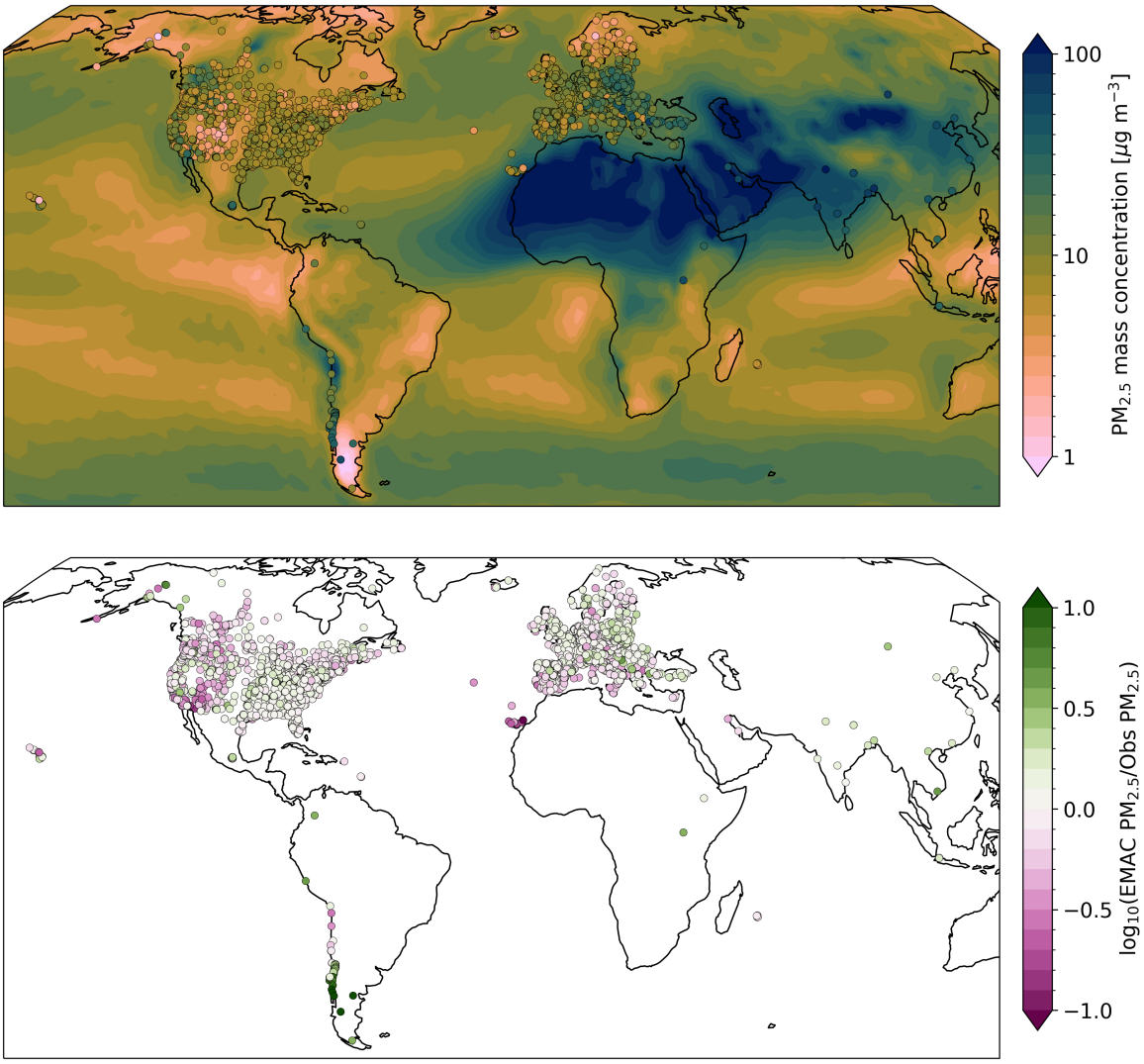


Figure 1: EMAC simulated global $PM_{2.5}$ mass concentrations for the year 2017 with layover GHOST observations (top), and global map of the model-to-observation ratio at each station (bottom).

We argue that R^2 values between 0.34 and 0.64 are not particular low values for such an evaluation, representing Pearson correlation coefficients between ~ 0.6 and 0.8. Higher R^2 are definitely hindered by the limited horizontal resolution of the model ($1.875° \times 1.875°$), as the local variability of $PM_{2.5}$ cannot be represented by such a model resolution. Figure 2 of the submitted manuscript shows that for the speciated compounds the good $\overline{M/O}$ agreement is not coming from compensating effects, but generally consistent agreement with some variability (with the exception of Cl^-).

We agree, that R^2_{log} of 0.25 for $PM_{2.5}$ is comparably low, and judging from Fig. 2 in the submitted manuscript, there are actually some compensating effects of negative and positive outliers here. However, we show in Table 2, that 90% of the observations lie within a factor of two of the simulated values. We additionally add a map of global model $PM_{2.5}$ concentrations

overlayed with observations and a model-to observation ratio map as supplementary material (Fig. 1 here).

The bottom panel shows the logarithmic ratio of model $PM_{2.5}$ to observations. Outliers are mostly confined to three regions: the western US and Canada, strongly influenced by wildfires in 2017 (Bernstein et al., 2021; Kloss et al., 2019); Chile, influenced by the 2017 mega-wildfires (Bowman et al., 2019; de la Barrera et al., 2018); and the northwest outflow region of the Sahara Desert, characterized by strong $PM_{2.5}$ gradients.

We added the following paragraph in the revised manuscript:

" The comparatively low R^2 for $PM_{2.5}$ across all setups is likely caused by strong outliers (Fig. 2a). A spatial comparison of simulated and observed $PM_{2.5}$ for 2017 (Fig. S3) shows that these outliers are mostly confined to three regions: the western US and Canada, strongly influenced by wildfires in 2017 (Bernstein et al., 2021; Kloss et al., 2019); Chile, with likely underestimated emissions from the 2017 mega-wildfires (Bowman et al., 2019; de la Barrera et al., 2018); and the northwest outflow region of the Sahara Desert, characterized by strong $PM_{2.5}$ gradients. "

Fig 3: is "log" in base 10 or e? For clarity, I'd use ln if base e and log_10 if base 10.

We apologize for the confusion. "log" is $\log_{10}$ in this case, and we adjust Fig. 3 and S13 in the revised manuscript and supplement accordingly.

L355: why is the pattern opposite in the comparison with MODIS vs AERONET? Is it because MODIS and AERONET are different even when they observe the same locations, or because they measure different locations?

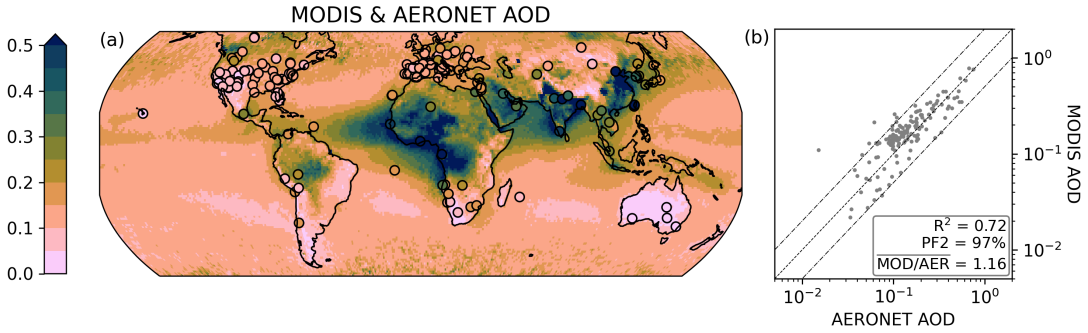


Figure 2: Comparison of annually averaged (2017) column Aerosol Optical Depth (AOD) from MODIS and AERONET observations. (a) MODIS AOD at 550 nm, overlayed with AERONET observations extrapolated to 550 nm. (b) Corresponding scatter plot, comparing MODIS and AERONET AOD at each AERONET station

We thank the referee for this question. To answer, we performed a set of diagnostic checks. First, a direct MODIS-vs-AERONET comparison (no model involved) shows the two products agree well (R^2=0.72), with only a 16% offset between AERONET and MODIS (Fig. 2 here). Second, when we resample MODIS only at AERONET station locations, the ranking of model setups already reverses to match the pattern seen in the AERONET comparison (both in $\overline{M/O}$ and in R^2), even though the reference dataset (MODIS) is unchanged (Fig. 3 here). Third, restricting the MODIS comparison to ocean-only grid cells reproduces almost exactly the global

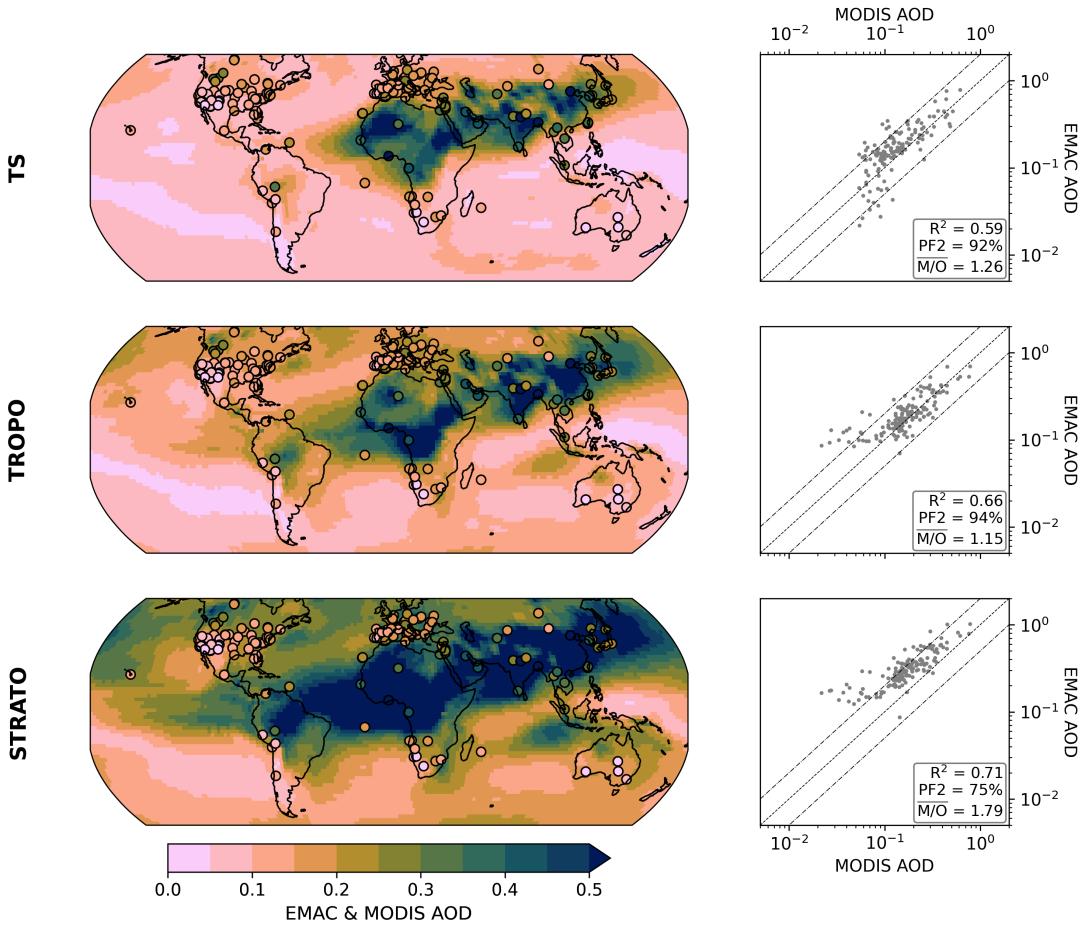


Figure 3: Comparison of annually averaged (2017) column Aerosol Optical Depth (AOD at 550 nm) from EMAC (TS, TROPO and STRATO setup) with MODIS observations sampled at AERONET stations. (a,c,e) Model AOD at 550 nm, overlayed with MODIS observations at each AERONET station. (b,d,f) Corresponding scatter plot, comparing model and MODIS AOD at each AERONET station.

MODIS statistics, confirming that the global comparison is effectively dominated by oceanic performance (Fig. 3 here).

These checks demonstrate that the mentioned "opposite pattern" is mostly driven by the differing spatial sampling of the two product (AERONET: sparse, land-based, skewed toward polluted/dust source regions; MODIS: near-global, ocean-dominated) rather than a fundamental disagreement between the MODIS and AERONET AOD products themselves. However, MODIS and AERONET still differ, and the overall statistical agreement between the TS simulation and AERONET is only moderately lower than that between MODIS and AERONET. This indicates that a substantial part of the residual model-observation discrepancy (especially over land) falls within the range of retrieval uncertainty in independent AOD observational products.

We added the following sentences:

" The reversal in model ranking between the MODIS and AERONET comparisons is primarily a sampling, not a retrieval, effect. MODIS and AERONET agree well at AERONET locations

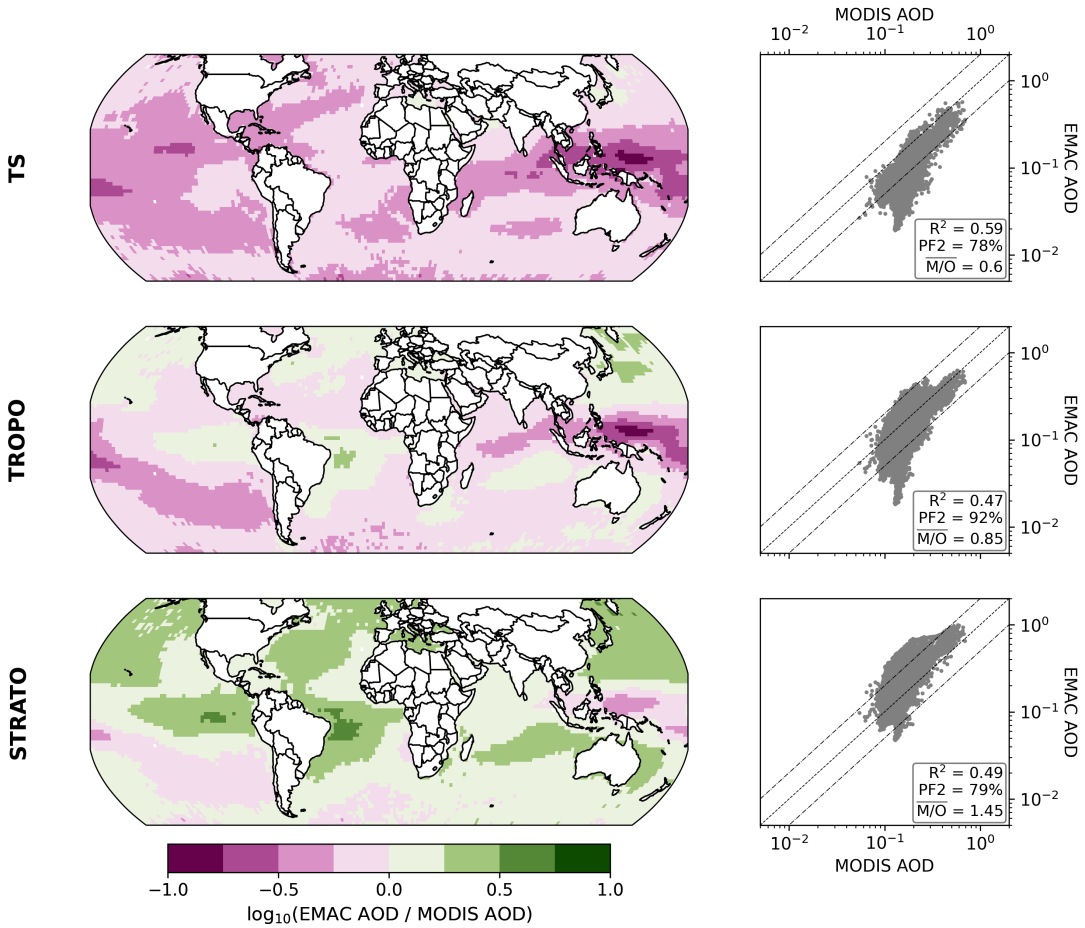


Figure 4: Comparison of annually averaged (2017) simulated column Aerosol Optical Depth (AOD) at 550 nm from EMAC (TS setup) with MODIS observations over the ocean only. (a) Global map of the logarithmic ratio of simulated to MODIS AOD; green indicates overestimation by the model, pink underestimation. (b) Corresponding scatter plot, with each dot representing one model grid box.

(R^2=0.72, PF2=97%), with only a 16% offset (Fig. S7). Restricting the MODIS comparison to AERONET station locations reproduces the AERONET-based model ranking (Fig. S8), while restricting it to ocean-only grid boxes reproduces the global MODIS statistics (Fig. S9), confirming that the global MODIS comparison is effectively dominated by ocean performance, where AERONET provides little constraint. Notably, the overall statistical agreement between the TS simulation and AERONET is only moderately lower than that between MODIS and AERONET, indicating that a substantial part of the model-observation discrepancy (especially over land) falls within the range of retrieval uncertainty in independent AOD observational products. ”

L377 “low M/O . . . datasets.”: I am not sure I understand this. If concentrations are below the detection limit, are they considered zero or equal to the detection limit? Couldn’t the same detection limit be applied to the model output?

We thank the referee for pointing out this issue, as it made us reflect our methodology, which we have now changed in the revised manuscript.

In the submitted manuscript, observations below the detection limit (including instances where measurement noise produced small negative values) were excluded (along with the respective simulated value) prior to computing logarithmic statistics, since the logarithm is undefined for non-positive values. On reflection, and in response to the referee's comment, we agree that this approach is unsatisfactory for two reasons: (i) it is not clearly specified how such values are treated, and (ii) excluding them selectively removes only those statistical events that happened to push the observation below zero/the detection limit, while retaining those pushed upward. This asymmetric exclusion introduces a bias toward an apparent model underestimation that is an artifact of the data handling rather than a model deficiency. Applying the detection limit to the simulated values would not remove this issue. Truncating these model values would discard sub-detection-limit model information and introduce a new one-sided artifact rather than removing the original one. In particular, low observational values are occasionally pushed above the detection limit by the instrument noise, while simulated values are not. Further, it is also crucial that the model reproduces near-zero observations, which would be omitted when the detection limit was applied to observations and model.

We have therefore revised the statistical analysis to retain these observations and use an inverse hyperbolic sine (arcsinh) transformation instead of a logarithmic transformation. The transformation is approximately linear around zero (below the detection limit/noise of the instrument), allowing zero and negative observations to be retained, while approaching logarithmic behaviour at concentrations well above the detection limit. We use the observational detection limit as the scale parameter of the transformation rather than applying the detection limit as a threshold to the model output. Thus, model values are not truncated or replaced by the detection limit. Note, that for comparisons where all values are well above the detection limit, the statistics do not change.

We have revised the Methods section accordingly and removed the previous statement that low $\overline{M/O}$ values were partly attributable to the detection limits. The revised approach provides a more consistent comparison across the full concentration range while explicitly accounting for the limited resolution of the observational datasets near their detection limits. We applied the revised metrics on all the model-observation comparison and slightly updated the discussion based on the updated statistical values. While the statistical agreement increased (mostly due to detection limit issues), the general conclusions remains the same.

Updated Methods paragraph:

" For statistical comparison with observations, the following metrics were calculated:

- $\overline{M/O}$: Geometric-mean-like ratio between model simulated (M) and observed (O) values
- PF2: Percentage of simulated values within a factor of two of the observations
- R^2: Coefficient of determination R^2, calculated as the squared Pearson correlation coefficient
- RMSE: Root Mean Squared Error

As model and observed values span several orders of magnitude, these metrics are computed using an inverse hyperbolic sine (arcsinh) transformation. This choice addresses the following issue: due to instrument noise, observations close to the detection limit are occasionally reported as zero or slightly negative, while modelled values remain strictly positive. Moreover,

observational uncertainty becomes particularly important at low concentrations. The inverse hyperbolic sine transformation is defined for positive, zero, and negative values and behaves approximately linearly close to zero while approaching logarithmic behavior at large positive concentrations. This allows observations close to or below the detection limit to be retained without assigning them a value of zero or replacing them by the detection limit.

The observational detection limit was used as a characteristic measurement scale in the arcsinh transformation. For pairwise error metrics, the transformation was scaled by the detection limit of each respective observation (see Appendix D). For the correlation analysis, a common representative scale was used (median of the detection limit) so that the same monotonic transformation was applied to all observations and the rank ordering of concentrations was not altered by observation-specific scaling. Further details of the transformations and metric definitions are provided in Appendix D.

We categorize evaluation results by model performance: higher performance (green in tables), intermediate performance (orange) and lower performance (red). The categories are defined as follows: $\overline{\mathrm{M/O}}$: green $(0.8-1.25)$, orange $(0.5-0.8$ or $1.25-2.0)$, and red $(<0.5$ or $>2.0)$; PF2: green $(>40\%)$, orange $(20-40\%)$, and red $(<20\%)$; R^2: green (>0.30), orange $(0.15-0.30)$, and red (<0.15); and arcsinh-based RMSE: green (<1.5), orange $(1.5-3.0)$, and red (>3.0). These thresholds are pragmatic, study-specific criteria intended to facilitate comparison among model configurations and should not be interpreted as universal standards for aerosol model performance. "

Additionally, we provide the explicit transformations in a new Appendix D ("Statistical Metrics and arcsinh transformation"):

" We define the detection-limit-scaled inverse hyperbolic sine transformation of the value x_i with the detection limit c_i as

$$T_i(x) = \operatorname{arcsinh}\left(\frac{x_i}{c_i}\right).$$

Unlike the logarithm, this transformation is defined for negative and zero values. For values close to zero, $\operatorname{arcsinh}(x_i/c_i) \approx x_i/c_i$, whereas for $x_i/c_i \gg 1$, $\operatorname{arcsinh}(x_i/c_i) \approx \ln\left(\frac{2x}{c_i}\right)$. Consequently, the transformation provides an approximately logarithmic scaling at concentrations well above the detection limit while avoiding the singularity of the logarithm near zero.

The difference between modelled (M_i) and observed (O_i) values on the transformed scale was calculated as

$$\Delta_i = T_i(M_i) - T_i(O_i).$$

For concentrations substantially above the detection limit, this difference approaches $\Delta_i \approx \ln(M_i/O_i)$, so that the transformed metrics retain an interpretation similar to conventional logarithmic model-observation statistics at higher concentrations.

Model-to-observation ratio

Because a conventional geometric mean of (M_i/O_i) cannot be calculated when observations are

zero or negative, we define an arcsinh-based multiplicative bias as

$$\exp\left(\frac{1}{n}\sum_{i=1}^{n}\Delta_i\right).$$

At concentrations well above the detection limit, this quantity approaches the conventional geometric mean of M_i/O_i. At concentrations close to or below the detection limit, it should instead be interpreted as a multiplicative bias derived from differences on the arcsinh scale, rather than as a literal geometric mean of the concentration ratios.

PF2

The conventional factor-of-two criterion is equivalent, for positive values, to

$$|\ln(M_i/O_i)| \leq \ln(2).$$

We therefore define the arcsinh-scale PF2 as the fraction of observations satisfying

$$|\Delta_i| \leq \ln(2).$$

At concentrations well above the detection limit this approaches the conventional factor-of-two criterion. Near the detection limit, the criterion transitions toward an absolute error criterion scaled by the observational measurement scale. Thus, the resulting metric is referred to as PF2 on the arcsinh scale.

RMSE

The root mean squared error on the transformed scale is

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}[T_i(M_i) - T_i(O_i)]^2}.$$

For concentrations well above the detection limit, this approaches the RMSE of the logarithmic model-observation differences. Close to the detection limit, the transformation instead scales differences relative to the observational measurement scale, preventing a small absolute error near zero from producing an arbitrarily large logarithmic error.

Correlation and R^2

For correlation analysis, a common scale parameter (c_0) was used for all observations,

$$T_0(x) = \text{arcsinh}(x/c_0),$$

and

$$R^2 = \left[\text{corr}\left(T_0(M), T_0(O)\right)\right]^2,$$

where corr denotes the Pearson correlation coefficient.

A common scale parameter c_0 (median of all c_i) is used for R^2 because applying a different scale parameter to each observation would effectively normalize each concentration by its individual detection limit. This can alter the ordering of observations and therefore change the quantity being measured by the correlation. The common c_0 ensures that the same monotonic transformation is applied to all observations. "

L386: how are biomass burning and volcanic emissions distributed in the vertical?

This is an important question, that is not sufficiently discussed in the submitted manuscript. The vertical distribution of biomass burning emissions is based on Table 4 of (Dentener et al., 2006), providing a fractional distribution of emission heights up to 6 km for different fire types. Obviously, in reality, the injection heights can substantially differ for individual fires. However, accurate estimation of injection heights from satellite observations is very challenging. These uncertainties translate into high uncertainties in long-range transport of biomass burning particles, especially when episodically compared to aircraft observations.

Tropospheric volcanic emissions are based on a climatology, also from Dentener et al., 2006, for the year 2000. This climatology has two main weaknesses. First, tropospheric degassing volcanic emissions show strong interannual variability, and emissions may have substantially changed since 2000. Second, emissions mostly only extend up to $\sim$ 6 km, similarly to the biomass buring emissions. Stratospheric volcanic emissions are applied based on altitudes provided by Schallock et al., 2023, Thus, there is a large gap between the tropospheric and statospheric emissions considered in the free and upper troposphere, that potentially leads to the underestimation of sulfate in the UTLS.

Both topics are subject to future improvements on the EMAC model.

We adjusted the text in the revised manuscript, starting from L386, as follows:

" This suggests that either biomass burning emissions or injection altitude is underestimated in the EMAC model. In fact, SO_2 emissions from biomass burning are highly uncertain, and strongly differ between emissions databases (Akagi et al., 2011; Andreae, 2019). Furthermore, accurate estimates of injection altitudes for individual fires remain challenging; here, a fractional vertical emission profile extending up to 6 km is assumed for each fire type, following Dentener et al., 2006, which may not capture the elevated injection heights of intense pyroconvective events (see Sect. 2.2). In other regions (e.g., South Pacific), the sulfate underestimation occurs independently of other aerosol components. We hypothesize that this may be due to missing emissions from degassing volcanoes or smaller eruptions that contribute to background sulfate in remote regions. Notably, tropospheric degassing volcanic emissions are prescribed from a static climatology representative of the year 2000 (Dentener et al., 2006) at the volcano elevation extending to similar altitudes ($\sim$ 6 km), and do not capture interannual variability in volcanic activity. Together with the altitude threshold for explosive stratospheric injections (Schallock et al., 2023), this leaves a vertical gap in prescribed emissions roughly between 6 km and the tropopause, which may contribute to underestimated sulfate in the UTLS. "

Additionally, we shortly mentioned the injection details in the model setup:

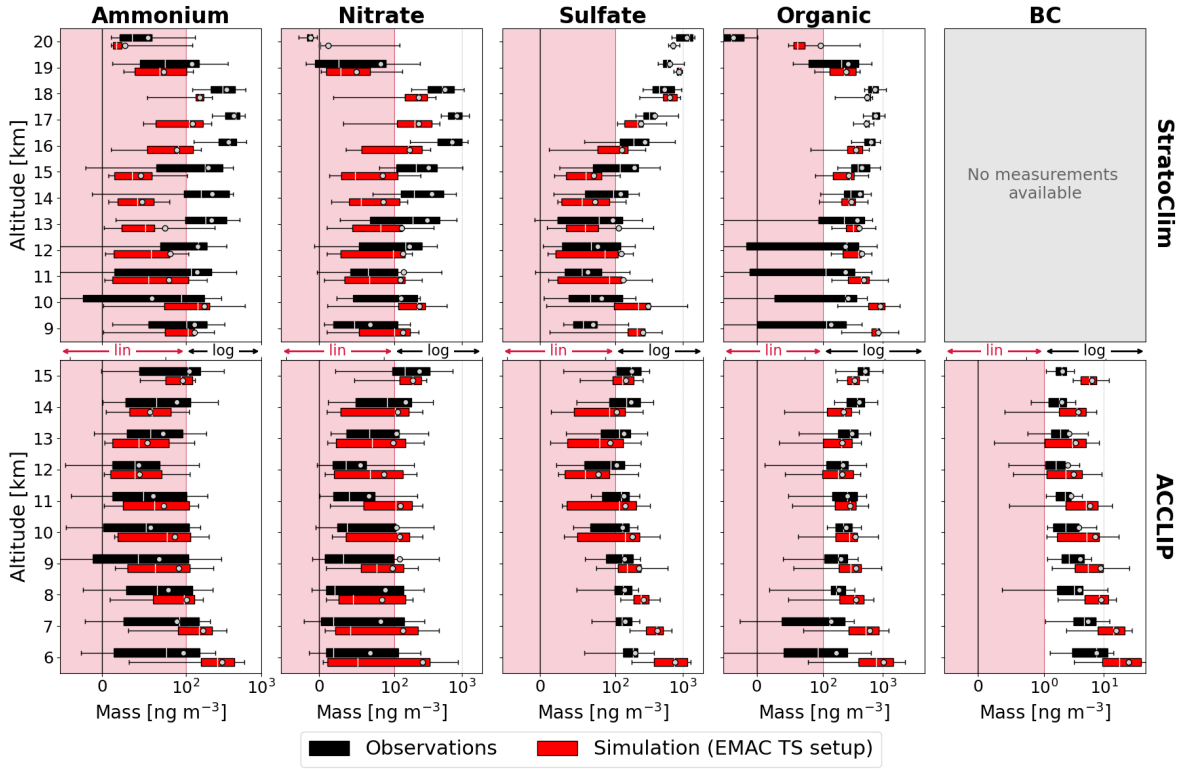


Figure 5: New Fig. 6 of the manuscript.

"Biomass burning and agricultural waste burning emissions are calculated based on observed dry matter burned and fire type, following Andreae (2019), using the BIOBURN submodel, and vertically distributed up to 6 km according to Dentener et al., 2006."

"Emissions from degassing tropospheric volcanoes are applied based on a volcanic climatology from AEROCOM (Dentener et al., 2006) for the year 2000 at the volcano elevation."

L486: BC is missing for StratoClim in Fig. 6. Was it not measured?

Yes, BC was unfortunately not measured during StratoClim. We had mentioned this in the caption ("..., and black carbon (BC) measurements from the SP2 instrument (only for ACCLIP as BC was not measured during StratoClim)"). Additionally, we now make it visually clearer, that the measurements are not available (see Fig. 5 here).

Fig A1: the vertical labels say PM, but only the sea salt component of PM is shown, correct? Also, I have the same comment as for L335. I don't think comparing speciation to MERRA2 is useful (and in any case the label on the color calendar should not say "observed" but rather "reanalysis" or "MERRA2"). If GEOS has a bias in sea salt, this bias will remain after assimilating the AOD.

We thank the referee for raising both issues.

Regarding the labeling: We adjusted the Figure, so that it says "SS PM". We also changed "observed" to "reanalysis" (see also above). Additionally, we changed the "log()" to "$\log_{10}(...)$" in the colorbar label. The new figure can be seen as Fig. 6 here.

Regarding the comparison with MERRA-2/CAMS: We agree with the referee's concern, that

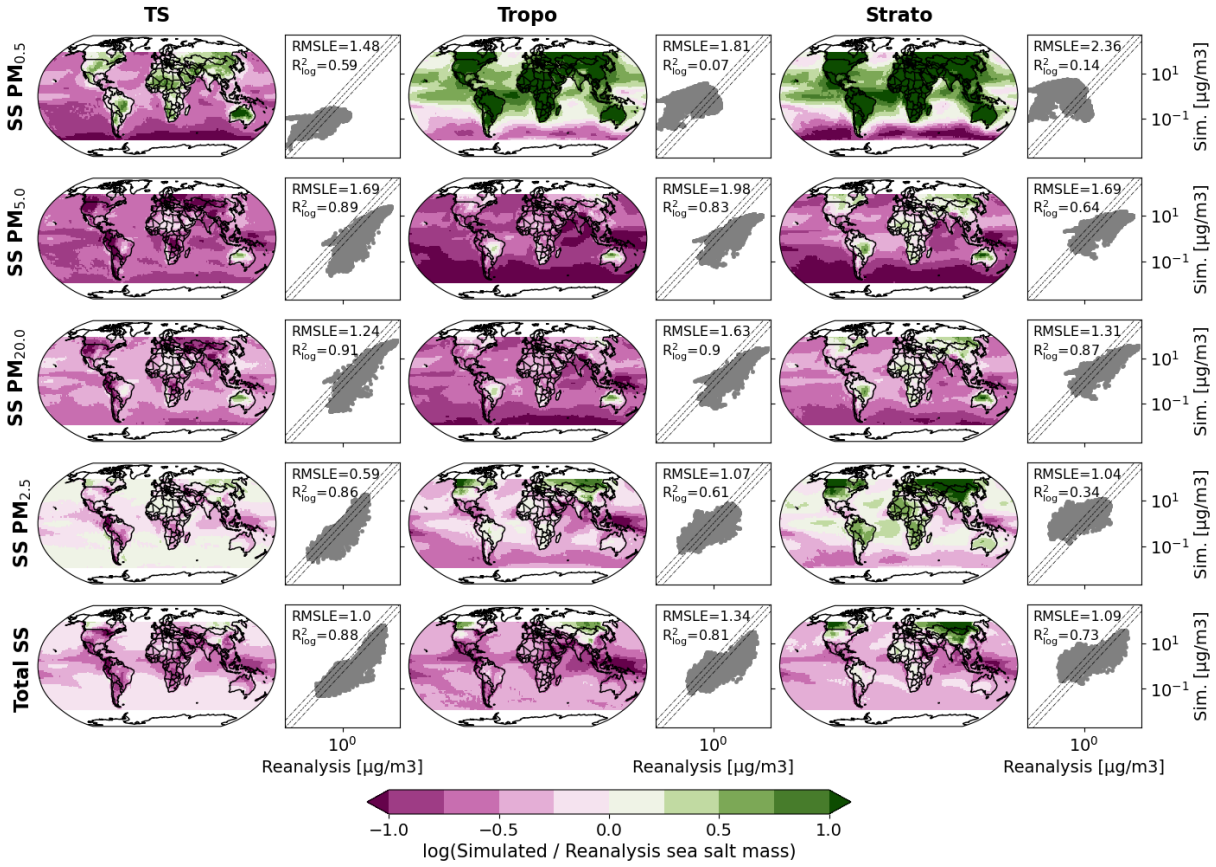


Figure 6: New Fig. A1 of the manuscript.

was already raised before. We have revised the manuscript (Sect. 3.1 and Appendix A; see our response to the L335 comment) to explicitly state the limitation of the speciation, and reframe the purpose of this comparison as a consistency check between independent aerosol modelling systems, rather than as a validation against independently constrained observations.

References

Akagi, S. K., R. J. Yokelson, C. Wiedinmyer, M. J. Alvarado, J. S. Reid, T. Karl, J. D. Crounse, and P. O. Wennberg (2011). "Emission factors for open and domestic biomass burning for use in atmospheric models". *Atmospheric Chemistry and Physics* 11.9, pp. 4039–4072. DOI: 10.5194/acp-11-4039-2011.

Andreae, M. O. (2019). "Emission of trace gases and aerosols from biomass burning – an updated assessment". *Atmospheric Chemistry and Physics* 19.13, pp. 8523–8546. DOI: 10.5194/acp-19-8523-2019.

Benduhn, F. and M. G. Lawrence (2013). "An investigation of the role of sedimentation for stratospheric solar radiation management". *Journal of Geophysical Research: Atmospheres* 118.14, pp. 7905–7921. DOI: doi.org/10.1002/jgrd.50622.

Bernstein, D. N., D. S. Hamilton, R. Krasnoff, N. M. Mahowald, D. S. Connelly, S. Tilmes, and P. G. M. Hess (2021). "Short-term impacts of 2017 western North American wildfires on meteorology, the atmosphere's energy budget, and premature mortality". *Environmental Research Letters* 16.6, p. 064065. DOI: 10.1088/1748-9326/ac02ee.

Bowman, D. M., A. Moreira-Muñoz, C. A. Kolden, R. O. Chávez, A. A. Muñoz, F. Salinas, Á. González-Reyes, R. Rocco, F. De la Barrera, G. J. Williamson, N. Borchers, L. A. Cifuentes, J. T. Abatzoglou, and F. H. Johnston (2019). "Human–environmental drivers and impacts of the globally extreme 2017 Chilean fires". *Ambio* 48.4, pp. 350–362. DOI: 10.1007/s13280-018-1084-1.

Brühl, C., J. Lelieveld, P. J. Crutzen, and H. Tost (2012). "The role of carbonyl sulphide as a source of stratospheric sulphate aerosol and its impact on climate". *Atmospheric Chemistry and Physics* 12, pp. 1239–1253. DOI: 10.5194/acp-12-1239-2012.

Curtius, J., M. Heinritzi, L. J. Beck, M. L. Pöhlker, N. Tripathi, B. E. Krumm, P. Holzbeck, C. M. Nussbaumer, L. Hernández Pardo, T. Klimach, K. Barmpounis, S. T. Andersen, R. Bardakov, B. Bohn, M. A. Cecchini, J.-P. Chaboureau, T. Dauhut, D. Dienhart, R. Dörich, A. Edtbauer, A. Giez, A. Hartmann, B. A. Holanda, P. Joppe, K. Kaiser, T. Keber, H. Klebach, O. O. Krüger, A. Kürten, C. Mallaun, D. Marno, M. Martinez, C. Monteiro, C. Nelson, L. Ort, S. S. Raj, S. Richter, A. Ringsdorf, F. Rocha, M. Simon, S. Sreekumar, A. Tsokankunku, G. R. Unfer, I. D. Valenti, N. Wang, A. Zahn, M. Zauner-Wieczorek, R. I. Albrecht, M. O. Andreae, P. Artaxo, J. N. Crowley, H. Fischer, H. Harder, D. L. Herdies, L. A. T. Machado, C. Pöhlker, U. Pöschl, A. Possner, A. Pozzer, J. Schneider, J. Williams, and J. Lelieveld (Dec. 2024). "Isoprene nitrates drive new particle formation in Amazon's upper troposphere". en. *Nature* 636.8041, pp. 124–130. ISSN: 1476-4687. DOI: 10.1038/s41586-024-08192-4.

Dasgupta, P. K., S. W. Campbell, R. S. Al-Horr, S. R. Ullah, J. Li, C. Amalfitano, and N. D. Poor (2007). "Conversion of sea salt aerosol to NaNO3 and the production of HCl: Analysis of temporal behavior of aerosol chloride/nitrate and gaseous HCl/HNO3 concentrations with AIM". *Atmospheric Environment* 41.20. (BRACE), pp. 4242–4257. ISSN: 1352-2310. DOI: j.atmosenv.2006.09.054.

de la Barrera, F., F. Barraza, P. Favier, V. Ruiz, and J. Quense (2018). "Megafires in Chile 2017: Monitoring multiscale environmental impacts of burned ecosystems". *Science of The Total Environment* 637-638, pp. 1526–1536. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2018.05.119.

Dentener, F., S. Kinne, T. Bond, O. Boucher, J. Cofala, S. Generoso, P. Ginoux, S. Gong, J. J. Hoelzemann, A. Ito, L. Marelli, J. E. Penner, J.-P. Putaud, C. Textor, M. Schulz, G. R.

van der Werf, and J. Wilson (2006). "Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom". *Atmospheric Chemistry and Physics* 6.12, pp. 4321–4344. DOI: 10.5194/acp-6-4321-2006.

Fountoukis, C. and A. Nenes (2007). "ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K^+–Ca^{2+}–Mg^{2+}–NH_4^+–Na^+–SO_4^-–NO_3^-–Cl^-–H_2O aerosols". *Atmospheric Chemistry and Physics* 7.17, pp. 4639–4659. DOI: 10.5194/acp-7-4639-2007.

Gard, E. E., M. J. Kleeman, D. S. Gross, L. S. Hughes, J. O. Allen, B. D. Morrical, D. P. Fergenson, T. Dienes, M. E. Gälli, R. J. Johnson, G. R. Cass, and K. A. Prather (1998). "Direct Observation of Heterogeneous Chemistry in the Atmosphere". *Science* 279.5354, pp. 1184–1187. DOI: 10.1126/science.279.5354.1184.

Gelaro, R., W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao (2017). "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)". *Journal of Climate* 30.14, pp. 5419–5454. DOI: 10.1175/JCLI-D-16-0758.1.

Guimbaud, C., F. Arens, L. Gutzwiller, H. W. Gäggeler, and M. Ammann (2002). "Uptake of HNO_3 to deliquescent sea-salt particles: a study using the short-lived radioactive isotope tracer ^{13}N". *Atmospheric Chemistry and Physics* 2.4, pp. 249–257. DOI: 10.5194/acp-2-249-2002.

Hodzic, A., P. Campuzano-Jost, H. Bian, M. Chin, P. R. Colarco, D. A. Day, K. D. Froyd, B. Heinold, D. S. Jo, J. M. Katich, J. K. Kodros, B. A. Nault, J. R. Pierce, E. Ray, J. Schacht, G. P. Schill, J. C. Schroder, J. P. Schwarz, D. T. Sueper, I. Tegen, S. Tilmes, K. Tsigaridis, P. Yu, and J. L. Jimenez (2020). "Characterization of organic aerosol across the global remote troposphere: a comparison of ATom measurements and global chemistry models". *Atmospheric Chemistry and Physics* 20.8, pp. 4607–4635. DOI: 10.5194/acp-20-4607-2020.

Inness, A., M. Ades, A. Agusti-Panareda, J. Barré, A. Benedictow, A.-M. Blechschmidt, J. J. Dominguez, R. Engelen, H. Eskes, J. Flemming, V. Huijnen, L. Jones, Z. Kipling, S. Massart, M. Parrington, V.-H. Peuch, M. Razinger, S. Remy, M. Schulz, and M. Suttie (2019). "The CAMS reanalysis of atmospheric composition". *Atmospheric Chemistry and Physics* 19.6, pp. 3515–3556. DOI: 10.5194/acp-19-3515-2019.

Jöckel, P., A. Kerkweg, A. Pozzer, R. Sander, H. Tost, H. Riede, A. Baumgaertner, S. Gromov, and B. Kern (2010). "Development cycle 2 of the Modular Earth Submodel System (MESSy2)". *Geoscientific Model Development* 3.2, pp. 717–752. DOI: 10.5194/gmd-3-717-2010.

Kerkweg, A., R. Sander, H. Tost, and P. Jöckel (2006). "Technical note: Implementation of prescribed (OFFLEM), calculated (ONLEM), and pseudo-emissions (TNUDGE) of chemical species in the Modular Earth Submodel System (MESSy)". *Atmospheric Chemistry and Physics* 6.11, pp. 3603–3609. DOI: 10.5194/acp-6-3603-2006.

Kloss, C., G. Berthet, P. Sellitto, F. Ploeger, S. Bucci, S. Khaykin, F. Jégou, G. Taha, L. W. Thomason, B. Barret, E. Le Flochmoen, M. von Hobe, A. Bossolasco, N. Bègue, and B. Legras (2019). "Transport of the 2017 Canadian wildfire plume to the tropics via the Asian monsoon circulation". *Atmospheric Chemistry and Physics* 19.21, pp. 13547–13567. DOI: 10.5194/acp-19-13547-2019.

Kohl, M., J. Lelieveld, S. Chowdhury, S. Ehrhart, D. Sharma, Y. Cheng, S. N. Tripathi, M. Sebastian, G. Pandithurai, H. Wang, and A. Pozzer (2023). "Numerical simulation and evaluation of global ultrafine particle concentrations at the Earth's surface". *Atmospheric Chemistry and Physics* 23.20, pp. 13191–13215. DOI: 10.5194/acp-23-13191-2023.

Kohl, M. and A. Pozzer (2026). *Global Monthly Mean Aerosol Mass and Number Concentrations from the Surface to the Stratosphere (2016–2019), EMAC Model Simulations*. Edmond [data set]. Version 1.0. DOI: 10.17617/3.DOM8G4.

Lyu, M., A. T. Ahern, G. P. Schill, M. J. Lawler, D. M. Murphy, S. J. Taylor, A. Fodel, M. Abou-Ghanem, C. Gurganus, Y. Zhu, S. Tilmes, E. Ray, T. D. Thornberry, R.-S. Gao, E. J. Hintsa, F. Moore, G. Dutton, D. Nance, B. Hall, A. W. Rollins, E. M. Waxman, K. Zuraski, G. S. Diskin, Y. Choi, R. B. Pierce, B. Weinzierl, F. Kuderna, M. Dollner, E. Jensen, and C. A. Brock (2026). "An unrecognized mode of small particles in the lower stratosphere". *Science* 392.6796, eadw8939. DOI: 10.1126/science.adw8939.

Murphy, D. M., K. D. Froyd, J. P. Schwarz, and J. C. Wilson (2014). "Observations of the chemical composition of stratospheric aerosol particles". *Quarterly Journal of the Royal Meteorological Society* 140.681, pp. 1269–1278. DOI: https://doi.org/10.1002/qj.2213.

Pozzer, A., S. F. Reifenberg, V. Kumar, B. Franco, M. Kohl, D. Taraborrelli, S. Gromov, S. Ehrhart, P. Jöckel, R. Sander, V. Fall, S. Rosanka, V. Karydis, D. Akritidis, T. Emmerichs, M. Crippa, D. Guizzardi, J. W. Kaiser, L. Clarisse, A. Kiendler-Scharr, H. Tost, and A. Tsimpidi (2022). "Simulation of organics in the atmosphere: evaluation of EMACv2.54 with the Mainz Organic Mechanism (MOM) coupled to the ORACLE (v1.0) submodel". *Geoscientific Model Development* 15.6, pp. 2673–2710. DOI: 10.5194/gmd-15-2673-2022.

Pringle, K. J., H. Tost, S. Metzger, B. Steil, D. Giannadaki, A. Nenes, C. Fountoukis, P. Stier, E. Vignati, and J. Lelieveld (2010). "Description and evaluation of GMXe: a new aerosol submodel for global simulations (v1)". *Geosci. Model Dev.* 3.2, pp. 391–412. DOI: 10.5194/gmd-3-391-2010.

Randles, C. A., A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, J. Hair, Y. Shinozuka, and C. J. Flynn (2017). "The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation". *Journal of Climate* 30.17, pp. 6823–6850. DOI: 10.1175/JCLI-D-16-0609.1.

Schallock, J., C. Brühl, C. Bingen, M. Höpfner, L. Rieger, and J. Lelieveld (2023). "Reconstructing volcanic radiative forcing since 1990, using a comprehensive emission inventory and spatially resolved sulfur injections from satellite data in a chemistry-climate model". *Atmospheric Chemistry and Physics* 23.2, pp. 1169–1207. DOI: 10.5194/acp-23-1169-2023.

Su, B., T. Wang, G. Zhang, Y. Liang, C. Lv, Y. Hu, L. Li, Z. Zhou, X. Wang, and X. Bi (2022). "A review of atmospheric aging of sea spray aerosols: Potential factors affecting chloride depletion". *Atmospheric Environment* 290, p. 119365. ISSN: 1352-2310. DOI: j.atmosenv.2022.119365.

Walcek, C. J. (2000). "Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection". *Journal of Geophysical Research: Atmospheres* 105.D7, pp. 9335–9348. DOI: https://doi.org/10.1029/1999JD901142.

Weigel, R., C. Mahnke, M. Baumgartner, A. Dragoneas, B. Vogel, F. Ploeger, S. Viciani, F. D'Amato, S. Bucci, B. Legras, B. Luo, and S. Borrmann (2021). "In situ observation of new particle formation (NPF) in the tropical tropopause layer of the 2017 Asian monsoon

anticyclone – Part 1: Summary of StratoClim results". *Atmospheric Chemistry and Physics* 21.15, pp. 11689–11722. DOI: 10.5194/acp-21-11689-2021.

Williamson, C. J., A. Kupc, D. Axisa, K. R. Bilsback, T. Bui, P. Campuzano-Jost, M. Dollner, K. D. Froyd, A. L. Hodshire, J. L. Jimenez, J. K. Kodros, G. Luo, D. M. Murphy, B. A. Nault, E. A. Ray, B. Weinzierl, J. C. Wilson, F. Yu, P. Yu, J. R. Pierce, and C. A. Brock (Oct. 2019). "A large source of cloud condensation nuclei from new particle formation in the tropics". en. *Nature* 574.7778, pp. 399–403. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1638-9.